Indian J Microbiol (Jan–Mar 2016) 56(1):46–58
DOI 10.1007/s12088-015-0561-5

ORIGINAL ARTICLE

A Genome-Wide Profiling Strategy as an Aid for Searching Unique Identification Biomarkers for *Streptococcus*

Vipin Chandra Kalia[1,2] · Ravi Kumar[1] · Prasun Kumar[1] · Shikha Koul[1,2]

Received: 27 October 2015 / Accepted: 13 November 2015 / Published online: 20 November 2015

Abstract The use of *rrs* (16S rRNA) gene is widely regarded as the "gold standard" for identifying bacteria and determining their phylogenetic relationships. Nevertheless, multiple copies of this gene in a genome is likely to give an overestimation of the bacterial diversity. In each of the 50 *Streptococcus* genomes (16 species, 50 strains), 4–7 copies of *rrs* are present. The nucleotide sequences of these *rrs* genes show high similarity within and among genomes, which did not allow unambiguous identification. A genome-wide search revealed the presence of 27 gene sequences common to all the *Streptococcus* species. Digestion of these 27 gene sequences with 10 type II restriction endonucleases (REs) showed that unique RE digestion in *purH* gene is sufficient for clear cut identification of 30 genomes belonging to 16 species. Additional gene-RE combinations allowed identification of another 15 strains belonging to *S. pneumoniae*, *S. pyogenes*, and *S. suis*. For the rest 5 strains, a combination of 2 genes was required for identifying them. The proposed strategy is likely to prove helpful in proper detection of pathogens like *Streptococcus.*

Keywords Biomarkers · Diagnosis · Genome · In silico · Restriction endonuclease · *Streptococcus*

Introduction

Streptococcus species cause severe diseases such as bacteremia, pneumonia, meningitis, and otitis media. Since these are responsible for high morbidity and mortality rates, clinicians and microbiologists have been constantly struggling to develop assays to identify them [1, 2]. The diseases caused by these pathogens may assume an epidemic dimension, if their growth is unchecked. The process of identification has progressed dramatically from conventional biochemical assays to molecular techniques [3]. Yet, in spite of a large number of identification strategies being developed in the past few decades, there seems to be no universal procedure and characteristic, which can be used to target all the bacteria. Evaluation of pathogens belonging to the genus *Streptococcus* has been highly problematic and hampered due to: (1) the close phylogenetic relationship among its species, and (2) sharing of phenotypic traits through horizontal gene transfer [4]. These events blur the boundaries and lead to difficulties in unequivocal identification of bacterial isolates [5]. There is a lack of sensitivity and specificity of the identification assays. Efforts to improve the quality of these assays have been made using diverse characteristics. The tube bile solubility test is widely employed for the identification of *Streptococcus pneumoniae* [6]. The major limitations are time consuming and difficulties in interpreting the results. Matrix-assisted laser desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) is another tool, which has also been proving quite effective in bacterial identification [3, 7–9]. This procedure has been reported to

Electronic supplementary material The online version of this article (doi:10.1007/s12088-015-0561-5) contains supplementary material, which is available to authorized users.

✉ Vipin Chandra Kalia
vckalia@igib.res.in; vc_kalia@yahoo.co.in

[1] Microbial Biotechnology and Genomics, CSIR - Institute of Genomics and Integrative Biology (IGIB), Delhi University Campus, Mall Road, Delhi 110007, India

[2] Academy of Scientific and Innovative Research (AcSIR), 2, Rafi Marg, Anusandhan Bhawan, New Delhi 110001, India

be not very effective in closely related *Streptococcus* species.

The reliability and reproducibility of molecular techniques, especially amplification of genes has taken precedence over culture based methods. The gene most widely used for identifying bacteria is the house keeping gene (HKG) −16S rRNA (*rrs*), which is highly conserved among all prokaryotes. The full length sequence of *rrs* of *Streptococcus* spp.: *S. pneumoniae*, *S. pseudopneumoniae*, *S. mitis*, and *S. oralis* show more than 99 % similarity leading to unsuccessful or misleading results [10]. Recent studies have elucidated certain latent but unique characteristics in *rrs*: (1) 30–50 nucleotides (nts) long signatures, and (2) restriction endonuclease (RE) digestion patterns [12–16]. These features enable easy identification of organisms up to the species level. However, *rrs* gene does not prove helpful in the following scenarios: (1) in phylogenetically very closely related strains, which differ in less than 1 % nts along their length, and (2) in strains with multiple copies of this gene in each genome [17–20]. An obvious solution to circumvent this problem is to use other HKGs [13, 21]. *Streptococcus* poses a big challenge as most pathogenic strains are genetically very close to each other and have multiple copies of *rrs* per genome. A few genes which are frequently used for the identification of *Streptococcus* include genes: *cpsA*, *gdh*, *groESL*, *lytA*, *ply*, *psaA*, *pspA*, *recA*, *recN*, *rpoA*, *rpoB*, *sodA*, *tuf*, *wzg*, 16S-23S ribosomal DNA spacer region, and the DNA fragment Spn9802 [3, 4, 11, 22–31]. Optochin susceptibility testing (CO_2 atmosphere) in combination with amplification of Spn9802 fragment and the autolysin genes: *lytA*, *rpoB*, and *tuf*, have proved effective in identifying *S. pneumoniae* [32]. It must be realized that the most idyllic *rrs* gene is offering no escape as human beings are struggling to reach a consensus, while pathogenic bacteria are threatening our long-term survival. Although, multilocus sequence analysis (MLSA) has been relatively quite successful in the identification of a wide range of bacterial species [33], however, using sequences of 7 HKGs: *guaA*, *map*, *pfl*, *ppaC*, *pyk*, *rpoB*, and *sodA*, also did not prove suitable for rapid identification of *Streptococcus* species [21]. MLSA using *gdh*, *ddl* along with *guaA*, *map*, *pfl*, *ppaC*, *pyk*, *rnpB*, *rpoB*, *sodA*, and *tuf* have been useful in identifying *Streptococcus* species [32, 34, 35].

Although many molecular techniques are available, however, they display variable specificity and reliability, and thus continue to be inconvenient for routine diagnostic assays. There is an urgent necessity to look for a procedure, which can discriminate *Streptococcus* species with higher precision than the existing ones. It must be based on certain genes with unique and easily identifiable characteristics. It should have the strength to be implemented in routine diagnostic procedures; otherwise the rapid escalation in expenditures on health may lead towards an otherwise avoidable economic collapse.

We have selected genes common to all the species of *Streptococcus* and systematically identified unique RE digestion patterns. These gene-RE combinations have the potential for being used in diagnosis of *Streptococcus* even among a large population of distantly or closely related organisms.

Materials and Methods

Sequence Data and Comparative Genome Analysis

Completely sequenced genomes of the 50 strains of 16 species belonging to genus *Streptococcus* were retrieved (http://www.ncbi.nlm.nih.gov/): *S. agalactiae*, *S. dysgalactiae*, *S. gordonii*, *S. intermedius*, *S. macedonicus*, *S. mitis*, *S. mutans* (2 strains), *S. oralis*, *S. parasanguinis*, *S. pasteurianus*, *S. pneumoniae* (20 strains), *S. pyogenes* (7 strains), *S. salivarius*, *S. sanguinis*, *S. suis* (9 strains), and *S. uberis* (Table S1). Characteristics of the *Streptococcus* genomes such as Accession number, GC percentage, size, and number of genes has been presented (Table S1). Based on GenBank (Full) data of *Streptococcus* genomes, we could trace 27 common genes, varying in size from 471 to 2514 nucleotides (nts; Tables S1 and S2). Gene, *rrs* was also taken into consideration. Sequence analysis and their orientation were done using BioEdit [36].

Restriction Endonuclease Analysis of Common Genes

A total of 10 Type II REs were considered for digestion on the basis of our previous works [18–20]. The following REs were used: (1) 4 base cutters *AluI* (AG'CT), *BfaI* (C'TA_G), *BfuCI* (_GATC'), *CviAII* (C_AT'G), *HpyCH4V* (TG'CA), *RsaI* (GT'AC), *TaqI* (T_CG'A), *Tru9I* (T_TA'A), and (2) 6 base cutters *HaeI* (WGG'CCW), *Hin1I* (GR_CG'YC). Cleaver (http://cleaver.sourceforge.net/) was used to get RE digestion patterns of the 27 common gene sequences (Table S2). Data matrices of REs producing 5–15 fragments were taken into consideration for further analysis [18–20].

Results

In Silico RE Digestion Analysis of *rrs* Gene

The genomes of *Streptococcus* strains have 4–7 copies of *rrs* gene. The different copies of *rrs* within a genome are exactly similar among themselves in 44 out of 50

Table 1 In silico restriction endonuclease (RE) digestion patterns (5′–3′) of *rrs* genes of *Streptococcus* strains

Streptococcus spp.	GenBank ID	Copies of *rrs*	Unique RE digestion pattern
RE-*AluI*			
S. gordonii str. Challis substr. CH1	CP000725.1	4	74•90•86•186•429•207•207•231
S. pyogenes MGAS9429	CP000259.1	6	122•429•207•207•266
S. pyogenes MGAS15252	CP003116.1	5	149•86•186•429•207•207•328•5•232
S. pyogenes NZ131	CP000829.1	6	63•86•186•429•207•207•157
S. salivarius 57.I	CP002888.1	6	74•86•86•186•56•373•207•207•261
S. uberis 0140 J	AM946015.1	4	160•86•186•429•161•46•207•269
RE-*BfaI*			
S. pasteurianus ATCC 43144	AP012054.1	5	79•52•89•27•578•186•137•195•194
S. pneumoniae CGSP14	CP001033.1	4	206•577•186•137•195•111
S. pyogenes MGAS9429	CP000259.1	2/6	515•186•137•195•198
		4/6[a]	236•578•186•137•195•314•67•116
RE-*BfuCI*			
S. parasanguinis FW213	CP003122.1	4	213•119•931•166
S. pneumoniae 670-6B	CP002176.1	4	246•119•932•174•8
S. pneumoniae R6	AE007317.1	4	4•294•119•932•165
S. pneumoniae TCH8431/19A	CP001993.1	4	7•294•119•932•174•8
S. pyogenes MGAS9429	CP000259.1	2/6	112•932•175•12
		4/6[a]	292•119•932•175•120•191
S. salivarius 57.I	CP002888.1	6	7•296•119•412•520•174•8
RE-*CviAII*			
S. dysgalactiae subsp. equisimilis RE378	AP011114.1	5	53•140•289•203•269•106•148•125•217
S. parasanguinis FW213	CP003122.1	4	97•492•268•106•148•125•193
S. pneumoniae 670-6B	CP002176.1	4	130•492•269•106•148•125•209
S. pneumoniae SPN034156	FQ312045.1	1/4	29•138•492•269•106•150•33•90•127
		3/4[a]	29•138•492•269•106•148•125•127
S. pyogenes MGAS9429	CP000259.1	6	369•269•106•148•125•214
S. pyogenes MGAS15252	CP003116.1	5	36•140•492•269•106•148•125•436•77
*S. suis*TL15	CP006246.1	4	47•142•492•269•106•148•125•210
S. uberis 0140 J	AM946015.1	4	47•123•17•492•269•106•148•125•217
RE-*HpyCH4 V*			
S. dysgalactiae subsp. equisimilis RE378	AP011114.1	5	56•28•130•6•443•396•262•229
S. gordonii str. Challis substr. CH1	CP000725.1	1/4	46•144•28•443•396•262•191
		3/4[a]	46•144•22•6•443•396•262•191
S. pneumoniae CGSP14	CP001033.1	1/4	11•134•22•6•443•88•308•262•139
		3/4[a]	11•134•22•6•443•395•262•139
S. pyogenes MGAS9429	CP000259.1	2/6	347•396•262•226
		4/6[a]	39•28•579•396•262•313•125•87
S. pyogenes NZ131	CP000829.1	2/6	560•396•262•117
		4/6[a]	51•28•579•396•262•117
S. suis D12	CP002644.1	1/4	78•579•396•262•222
		3/4[a]	50•28•579•396•262•222
S. suis TL13	CP003993.1	4	47•28•108•471•396•262•230
RE-*RasI*			
S. intermedius JTH08	AP010969.1	4	389•405•143•212•146•40
S. parasanguinis FW213	CP003122.1	4	539•262•355•146•128
S. pyogenes MGAS9429	CP000259.1	2/6	319•262•143•212•146•149
		4/6[a]	618•262•143•212•146•448

Table 1 continued

Streptococcus spp.	GenBank ID	Copies of *rrs*	Unique RE digestion pattern
S. pyogenes NZ131	CP000829.1	2/6	532•262•143•212•146•40
		4/6[a]	631•262•143•212•146•40
S. uberis 0140 J	AM946015.1	4	183•8•700•143•212•146•152
RE-*TaqI*			
S. parasanguinis FW213	CP003122.1	1/4	672•199•559
		3/4	88•584•198•559
S. pneumoniae 670-6B	CP002176.1	4	705•199•575
S. pyogenes MGAS9429	CP000259.1	2/6	452•199•580
		4/6[a]	751•199•687•192
S. pyogenes NZ131	CP000829.1	1/6	8•658•199•471
		5/6[a]	764•199•583
S. suis D12	CP002644.1	1/4	82•196•484•199•576
		3/4[a]	278•484•199•576
RE-*Tru9I*			
S. agalactiae 2603 V/R	AE009948.1	7	181•18•394•12•14•251•86•134•44•150•223
S. dysgalactiae subsp. equisimilis RE378	AP011114.1	5	1•186•10•8•394•12•265•86•134•44•150•260
S. intermedius JTH08	AP010969.1	4	102•242•48•116•41•224•86•134•342
S. parasanguinis FW213	CP003122.1	4	515•14•27•223•86•134•194•236
S. pasteurianus ATCC 43144	AP012054.1	5	181•409•15•14•26•225•86•134•194•253
S. salivarius 57.I	CP002888.1	6	605•265•86•134•194•252
S. sanguinis SK36	CP000387.1	1/4	2•200•410•14•27•224•86•134•454
		3/4	2•610•14•27•224•86•134•454
S. uberis 0140 J	AM946015.1	4	199•406•14•251•86•134•44•150•260

Symbol (•) indicates RE site in the gene sequences

[a] This pattern is not unique. It has been presented to indicate the RE digestion pattern of the rest of the *rrs* copies

sequenced genomes. Multiple sequence alignments of 217 copies of *rrs* belonging to 50 *Streptococcus* genomes resulted in segregating them into 40 different groups. The number of *rrs* copies in each group varied from 1 to 20 copies such that 150 copies cannot be distinguished from each other. On the other hand, in silico digestion of *rrs* gene sequences with 10 REs, allowed us to identify certain unique RE digestion patterns. The RE digestion patterns of all the *rrs* copies were unique to a strain (but exactly similar to each other), in the cases of: (1) *S. agalactiae* 2603V/R, (2) *S. dysgalactiae* subsp. equisimilis RE378, (3) *S. intermedius* JTH08, (4) *S. pasteurianus* ATCC 43144, (5) *S. pneumoniae* R6, (6) *S. pneumoniae* TCH8431/19A, (7) *S. pneumoniae* 670-6B, (8) *S. pyogenes* MGAS15252, (9) *S. salivarius* 57.I, (10) *S. suis* TL13, (11) *S. suis* TL15, and (12) *S. uberis* 0140J. This analysis revealed that 12 genomes belonging to 9 *Streptococcus* species can be easily distinguished because of the unique digestion pattern of their *rrs* gene copies with REs—*AluI*, *BfaI*, *BfuCI*, *CviAII*, *HpyCH4V*, *RsaI*, *TaqI*, and *Tru9I* (Table 1). In the cases of *S. gordonii* str. Challis substr. CH1, *S. parasanguinis* FW213, *S. pneumoniae* CGSP14, *S. pneumoniae* SPN034156, *S. pyogenes* MGAS9429, *S. pyogenes* NZ131, *S. sanguinis* SK36, and *S. suis* D12, the 4–6 copies of *rrs* could be segregated into two groups. However, only one of the two groups in the latter set of genomes had unique RE digestion patterns (Table 1). It is concluded that *rrs* may not prove as a very good candidate gene for identifying *Streptococcus* species in an unambiguous manner. It indicates that there is a need to look for other genes, which may be highly conserved and should have certain latent features like unique RE digestion patterns and prove helpful in deriving useful information.

In Silico RE Digestion Analysis of Common Genes

In this analysis, we have identified 27 genes (in addition to *rrs*), common to all the 50 sequenced genomes of *Streptococcus*. These 27 genes varied from 471 to 2514 nts, in size (Tables S2). In silico digestion of 27 genes with all the

Table 2 Unique in silico restriction endonuclease digestion patterns (5′–3′) of *purH* gene of *Streptococcus* genomes

Streptococcus spp.	Restriction endonucleases			
	AluI	*BfaI*	*BfuCI*	*CviAII*
(a)				
S. agalactiae 2603 V/R	400•42•75•57•138•8•148•21•63•66•60•18•452	416•63•536•360•173	891•384•158•81•34	–
S. dysgalactiae subsp. equisimilis RE378	223•33•20•352•30•291•426•170	1012•360•173	282•150•360•96•579•78	409•421•112•556•47
S. gordonii str. Challis	403•72•99•29•28•89•130•18•84•46•20•144•72•255•59	–[a]	83•151•1119•195	412•869•231•36
S. intermedius JTH08	403•138•141•186•84•66•63•345•15•107	–	719•183•531•115	211•36•21•144•413•8•448•207•60
S. macedonicus ACA-DC 198	–	–	285•150•906•207	211•201•413•8•448•231•36
S. mitis B6	259•20•295•146•232•46•215•21•114•141•59	–	1275•184•89	412•221•192•8•131•302•282
S. mutans LJ23	406•138•179•148•84•66•63•345•15•107	–	905•451•195	214•36•21•144•413•8•448•267
S. mutans UA159	418•72•66•179•148•84•66•63•153•114•78•15•107	–	917•451•195	15•211•36•21•144•413•8•448•267
S. oralis Uo5	541•33•294•84•46•164•386	–	285•974•289	633•192•8•131•317•231•36
S. parasanguinis FW213	406•279•168•102•46•61•289•200	253•166•666•466	557•19•45•158•63•75•528•72•34	828•8•131•317•231•36
S. pasteurianus ATCC 43144	–	–	285•150•1035•78	211•201•413•456•231•36
S. pneumoniae CGSP14	–	–	456•456•134•328•195	433•413•8•448•231•36
S. pneumoniae G54	279•295•146•232•46•20•144•386	–	–	412•221•192•8•448•267
S. pneumoniae Hungary19A-6	–	–	435•590•328•17•178	–
S. pneumoniae INV200	–	–	435•456•134•328•195	–
S. pneumoniae TCH8431/19A	–	–	456•590•328•195	268•119•438•390•273•60
S. pyogenes A20	279•238•203•169•129•60•18•452	158•258•599•360•173	–	–
S. pyogenes MGAS5005	309•238•203•169•129•60•18•422	7•181•258•599•360•143	–	855•8•448•220•11•6
S. salivarius 57.I	403•135•36•138•8•130•18•21•63•46•164•51•21•114•141•59	416•666•466	1459•55•34	211•614•8•448•220•47
S. sanguinis SK36	78•7•372•75•24•62•18•99•148•84•46•20•63•81•267•119	130•900•533	32•418•138•222•96•462•161•34	15•412•221•879•36
S. suis D12	259•450•26•263•20•351•179	–	435•699•414	268•119•828•273•24•36
S. suis ST1	72•184•235•50•168•26•154•129•195•213•122	–	435•360•339•219•195	268•557•687•36
S. suis T15	491•50•168•180•129•530	–	117•318•438•261•414	268•119•438•52•338•273•60
S. suis TL13	491•50•168•26•133•150•408•122	–	–	–
S. uberis 0140 J	355•48•105•9•24•120•207•84•46•20•60•81•54•108•27•81•12	289•127•44•516•39•178•167•19•169	435•918•195	211•201•413•676•47

Table 2 continued

Streptococcus spp.	Restriction endonucleases			
	HpyCH4 V	*RsaI*	*TaqI*	*Tru9I*
(b)				
S. agalactiae 2603 V/R	–[a]	–	146•222•972•208	16•172•306•129•303•321•298•3
S. dysgalactiae subsp. equisimilis RE378	10•108•131•148•159•201•203•585	–	750•185•610	61•124•126•408•21•48•27•429•189•109•3
S. gordonii str. Challis	198•361•147•257•168•417	1243•130•175	91•25•123•36•93•609•507•64	743•183•622
S. intermedius JTH08	943•20•93•75•417	–	–	–
S. macedonicus ACA-DC 198	13•246•300•210•194•28•411•24•122	–	91•1057•400	299•254•160•30•75•108•619•3
S. mitis B6	349•6•204•105•42•257•28•140•417	–	91•55•138•84•320•172•117•363•21•123•64	16•49•123•126•296•112•204•312•307•3
S. oralis Uo5	–	1243•85•220	91•55•213•9•492•195•429•64	16•727•802•3
S. parasanguinis FW213	667•299•585	–	149•93•36•93•492•117•507•64	332•819•288•109•3
S. pasteurianus ATCC 43144	13•246•300•210•194•28•438•119	–	91•1057•336•64	299•254•190•75•108•619•3
S. pneumoniae 70585	–	–	–	16•244•54•15•414•324•369•109•3
S. pneumoniae D39	–	–	91•55•714•117•507•64	16•172•534•204•141•369•109•3
S. pneumoniae G54	355•204•404•28•140•417	–	146•222•609•507•64	16•298•15•738•369•109•3
*S. pneumoniae*I NV104	–	–	91•55•222•492•117•507•64	–
S. pneumoniae R6	–	–	112•55•714•117•507•64	5•32•172•534•204•141•369•109•3
S. pyogenes MGAS5005	–	–	176•129•93•972•178	25•21•172•416•19•303•321•271
S. pyogenes NZ131	55•711•177•20•168•169•248	–	–	–
S. salivarius 57.I	963•28•411•146	–	146•93•36•9•84•492•624•64	722•426•397•3
S. sanguinis SK36	267•711•28•140•417	117•1427•19	798•77•78•210•264•136	329•477•754•3
S. suis D12	252•307•147•257•168•417	102•1141•258•47	368•533•439•208	16•283•422•97•730
S. suis ST1	252•220•24•21•189•257•118•50•298•119	–	91•707•103•439•208	16•172•416•117•827
S. suis T15	198•54•220•24•21•189•257•109•59•417	–	–	16•588•117•70•27•618•112
S. suis TL13	–	–	–	16•283•305•117•70•757
S. uberis 0140 J	13•950•28•432•66•21•38	–	368•66•1114	299•15•155•84•190•75•727•3

Symbol (•) indicates RE site in the gene sequences

[a] No unique pattern observed

Table 3 Unique in silico restriction endonuclease digestion patterns (5′–3′) of common genes (other than *purH*) of *Streptococcus* genomes

Streptococcus spp.	Gene	RE	RE digestion pattern
S. pneumoniae SPNA45	*dnaA*	*HpyCH4 V*	622•348•329
S. pneumoniae SPN034156	*dnaK*	*TaqI*	557•555•33•198•487
S. pneumoniae 670-6B		*Tru9I*	428•380•160•399•258•84•115
S. suis ST3		*Tru9I*	968•36•555•265
S. pneumoniae ATCC 700669	*fabG*	*Tru9I*	100•193•57•57•186•128•11
S. pneumoniae SPN994038	*mraY*	*AluI*	139•207•8•166•189•155•117
S. pneumoniae Taiwan19F-14		*AluI*	139•215•166•189•91•64•117
S. pneumoniae P1031	*purK*	*AluI*	5•509•174•69•335
S. pneumoniae OXC141		*BfaI*	125•249•5•120•180•422
S. pyogenes MGAS315		*AluI*	202•8•238•9•141•68•354
S. pyogenes MGAS9429		*AluI*	78•142•72•8•247•141•68•354
S. suis BM407		*Tru9I*	418•496•163•3
S. pneumoniae JJA	*pyrH*	*HpyCH4 V*	103•84•89•45•127•290
S. pyogenes MGAS1882		*Tru9I*	25•4•443•64•14•79•100
S. suis D9		*AluI*	279•115•66•272

Symbol (•) indicates RE site in the gene sequences

Table 4 Identification of *Streptococcus* strains using in silico restriction endonuclease digestion patterns (5′–3′) of two common genes

Streptococcus spp.	RE digestion patterns	
	argR-Tru9I	*purK-Tru9I*
S. suis P1/7	25•397•19	251•167•496•163•3
S. suis B407[a]	25•397•19	418•496•163•3
	argR-BfuCI	*purH-BfuCI*
S. suis SC84	29•445	435•699•219•195
S. suis D12[a]	29•445	436•699•414
	argR-Tru9I	*pyrH-Tru9I*
S. pyogenes MGAS15252	25•51•30•189•27•75•74	25•4•443•64•14•179
S. pyogenes MGAS1882[b]	25•51•30•189•27•75•74	25•4•443•64•14•79•100

Symbol (•) indicates RE site in the gene sequences

[a] Strains had unique RE pattern with *purH* (Table 2)

[b] Strain had unique RE pattern with *pyrH* (Table 3)

REs revealed unique features, on the basis of which the *Streptococcus* genomes can be easily distinguished. Although all the 27 genes proved to have certain unique features, however, only 7 genes, namely *purH*, *dnaA*, *dnaK*, *fabG*, *mraY*, *purK* and *pyrH* can be recommended for usage in identification process.

purH

In silico digestion of *purH* gene (1548 nts) was observed with 8 REs: *AluI*, *BfaI*, *BfuCI*, *CviAII*, *HpyCH4V*, *RsaI*, *TaqI*, and *Tru9I*. REs—*AluI*, *BfuCI*, *CviAII*, *TaqI* and *Tru9I* were effective in providing unique digestion patterns in *purH* gene present in 18–21 genomes. Together these gene-RE combinations encompassed 30 genomes, which represented all the 16 species used in this study. The unique feature of this gene is that the number of RE sites varied from 2 to 6 and exceptionally it varied up to 13. The fragment sizes ranged from 18 to 500 nucleotides (nts) in most of the cases and exceptionally up to 1100 nts (Table 2a, b).

dnaA, dnaK, fabG, mraY, purK and pyrH

Gene—*purH* did not provide information on unique RE patterns in the rest of the 20 genomes: *S. pneumoniae* (11/20 strains), *S. pyogenes* (4/7 strains), and *S. suis* (5/9 strains). We looked for unique RE digestion patterns in other genes that can be used for identification. This analysis revealed that the following combinations can be used: (a) *dnaA*:*HpyCH4V* for *S. pneumoniae* SPNA45; (b) *dnaK*: (1) *Tru9I* for *S. pneumoniae* 670-6B, (2) *TaqI* for *S. pneumoniae* SPN034156, and (3) *Tru9I* for *S. suis* ST3; (c) *fabG*: *Tru9I* for *S. pneumoniae* ATCC700669; (d) *mraY*: (1) *AluI* for *S. pneumoniae* SPN994038, and (2) *AluI* for *S. pneumoniae* Taiwan 19F-14; (e) *purK*: (1) *BfaI* for *S. pneumoniae* OXC141, (2) *AluI* for *S. pneumoniae* P1031; (3) *AluI* for *S. pyogenes* MGAS315 (4) *AluI* for *S. pyogenes* MGAS9429; (5) *Tru9I* for *S. suis* BM407; (f) *pyrH*: (1) *HpyCH4V* for *S. pneumoniae* JJA, (2) *Tru9I* for *S. pyogenes* MGAS1882, (3) *AluI* for *S. suis* D9 (Table 3). This strategy allowed us to distinguish 15 out of 20 genomes, which could not be segregated using *purH* gene.

Table 5 Unique in silico Restriction Endonuclease digestion patterns (5′–3′) of *recA* gene of *Streptococcus* genomes

Streptococcus spp.	Restriction endonucleases			
	AluI	*BfaI*	*BfuCI*	*CviAII*
(a)				
S. agalactiae 2603 V/R	136•144•261•282•245•22•50	706•178•256	753•342•45	271•646•223
S. dysgalactiae subsp. equisimilis RE378	61•111•57•138•89•7•228•65•267	–[a]	113•109•645•27•84•45	157•57•198•311•300
S. gordonii str. Challis	64•97•217•103•60•264•347	22•336•765•29	166•170•645•66•92•13	156•115•57•233•276•315
S. intermedius JTH08	136•34•140•171•389•276	–	336•443•109•93•27•138	156•115•57•198•116•195•309
S. macedonicus ACA-DC 198	28•84•58•116•57•35•163•36•293•285	–	–	–
S. mitis B6	161•182•227•346•138•27•50•24	–	336•117•702	117•154•57•233•81•195•318
S. oralis Uo5	161•380•36•293•279	–	86•693•109•64•29•99•69	117•211•233•276•80•232
S. parasanguinis FW213	161•416•228•282•59	–	86•23•109•118•98•345•173•85•109	117•39•115•57•497•12•309
S. pasteurianus ATCC 43144	28•84•58•116•57•35•163•36•246•47•285	–	–	–
S. pyogenes MGAS5005	138•21•90•57•138•96•102•191•304	–	–	234•57•198•35•276•337
S. pyogenes MGAS9429	–	–	336•117•639•45	–
S. pyogenes NZ131	–	–	336•645•111•45	–
S. salivarius 57.I	175•203•54•49•89•300•40•230	172•186•526•227•29	952•29•27•132	271•57•198•35•579
S. sanguinis SK36	175•257•189•184•65•40•69•170	–	336•443•202•168	156•172•198•311•312
S. suis D12	164•84•35•75•77•172•219•326	–	–	120•39•172•198•116•195•312
S. suis ST1	164•84•35•75•77•172•219•255•71	–	339•98•319•32•331•33	159•172•198•20•15•81•195•312
S. suis T15	173•75•35•75•77•172•219•255•68	–	–	120•39•172•198•20•15•81•195•309
S. suis TL13	164•84•35•75•77•391•326	574•543•35	–	120•39•115•371•195•312
S. uberis 0140 J	265•87•788	67•57•54•714•248	51•693•237•159	156•115•57•198•391•223

Streptococcus spp.	Restriction endonucleases			
	HpyCH4 V	*RsaI*	*TaqI*	*Tru9I*
(b)				
S. agalactiae 2603 V/R	424•54•59•379•110•114	–	57•167•243•363•310	365•219•38•169•80•4•265
S. dysgalactiae subsp. equisimilis RE378	–[a]	–	203•150•670	251•219•30•117•60•24•210•109•3
S. gordonii str. Challis	307•15•102•54•59•346•269	488•213•140•311	467•36•504•135•10	731•51•33•337
S. intermedius JTH08	76•111•9•111•171•99•27•542	488•250•408	467•669•10	80•285•249•69•132•87•244
S. macedonicus ACA-DC 198	–	–	–	85•280•318•108•24•60•280
S. mitis B6	150•124•57•93•113•618	488•353•81•233	224•93•18•132•672•16	80•735•340
S. oralis Uo5	307•45•96•48•41•346•266	488•235•118•81•227	89•135•93•150•426•240•16	80•735•331•3
S. parasanguinis FW213	136•138•33•42•188•346•263	149•339•353•20•61•103•121	89•135•93•54•66•30•36•504•33•106	614•69•132•60•268•3

Table 5 continued

Streptococcus spp.	Restriction endonucleases			
	HpyCH4 V	*RsaI*	*TaqI*	*Tru9I*
S. pasteurianus ATCC 43144	196•78•150•113•618	–	–	85•280•219•99•108•24•60•280
S. pyogenes A20	–	–	–	13•352•219•30•117•60•24•198•12•109•3
S. pyogenes MGAS5005	–	–	–	328•219•30•117•60•24•198•12•109•24•13•3
S. pyogenes MGAS9429	–	841•20•158•118	–	–
S. salivarius 57.1	112•24•60•126•126•30•18•198•446	488•213•12•306•6•115	317•99•724	365•249•168•9•24•325
S. sanguinis SK36	478•59•142•470	488•353•81•51•176	317•186•646	80•285•249•177•84•274
S. suis D12	352•75•54•18•41•157•455	–	182•138•186•646	818•331•3
S. suis T15	–	–	–	617•201•328•3
S. suis TL13	277•75•75•72•41•40•117•319•136	–	–	–
S. uberis 0140 J	196•111•42•6•69•180•201•111•224	–	89•378•76•159•438	79•286•318•48•51•9•24•60•262•3

Symbol (•) indicates RE site in the gene sequences

[a] No unique pattern observed

As no single gene-RE combination worked for identifying the 5 genomes of *S. pneumoniae* SPN034183, *S. pneumoniae* SPN994039, *S. suis* P1/7, *S. suis* SC84, and *S. pyogenes* MGAS15252, combinations of 2–3 gene-RE patterns were used. Two strains viz., *S. suis* P1/7, *S. suis* SC84, were found to have identical RE patterns with *argR* and *dnaA* genes of *S. suis* B407, and *S. suis* D12, respectively, whereas *S. pyogenes* MGAS15252 and *S. pyogenes* MGAS1882 shared their RE patterns for the following genes: *argR*, *argS*, *cysS*, *dnaK*, *glyA*, *gyrB*, *parE*, *purH*, *purK*, and *purR*. These three groups of two strains each could be distinguished by using additional gene-RE combinations: *purK-Tru9I*, *purH-BfuCI*, and *pyrH-Tru9I* (Table 4). Using this strategy, we could not distinguish *S. pneumoniae* SPN034183 and *S. pneumoniae* SPN994039, from each other.

The strategy of screening all the 50 genomes for searching genes which were common to all of them and subjecting each one of them to 10 different Res, allowed us to find unique RE digestion patterns in a few genes. *purH* alone proved effective in segregating 30 out of 50 genomes. Identification of an additional 18 genomes, was possible by employing other gene-RE combinations. No unique gene—RE combinations could be deduced for two strains of *S. pneumoniae*.

Discussion

Bacterial identification methods have graduated from those based on morphological and metabolic characteristics to molecular methods. Among the various genes based identification methods, the most widely employed has been the usage of *rrs* gene [12–16]. It has proved instrumental in bacterial identification; however, the major difficulty encountered is in the cases where the organism has multiple copies of *rrs* within the genome [13, 14, 18–20]. The issue becomes more complicated when *rrs* genes from different species show high sequence similarity among themselves. In order to circumvent these issues, the information is complemented by employing other highly conserved genes. However, it involves more inputs and selection of other genes out of a few thousand genes within the genome is not an easy task. In spite of the fact that around 23 genes have been used frequently in many studies carried out for identifying *Streptococcus*, no consensus gene has been identified [3, 4, 11, 21–35]. It has also not been realized that except *recA*, the rest of the 22 genes are not present in all the species of *Streptococcus*. Although, *recA* is one of those genes which are used widely for identification of *Streptococcus* [27], however, our analysis revealed that it is not among the best candidates, which can be exploited for

Table 6 Unique in silico restriction endonuclease digestion patterns (5′–3′) of *gyrB* gene of *Streptococcus* genomes

Streptococcus spp.	Restriction endonucleases			
	AluI	*BfaI*	*BfuCI*	*CviAII*
(a)				
S. agalactiae 2603 V/R	82•102•111•315•489•36•264•51•90•123•45•213•32	1711•54•185•3	483•1323•81•66	90•12•220•51•486•1094
S. dysgalactiae subsp. equisimilis RE378	121•156•195•423•291•9•132•165•461	812•363•778	112•371•838•334•124•27•147	322•51•425•22•11•1044•6•72
S. gordonii str. Challis	793•378•9•132•12•123•90•141•27•245	191•221•286•474•540•238	109•55•327•13•89•1087•96•27•59•88	817•39•15•530•111•288•150
S. intermedius JTH08	76•622•92•198•180•12•141•179•192•255	44•131•426•94•474•332•446	501•765•507•108•66	367•458•43•405•596•78
S. macedonicus ACA-DC 198	184•306•214•191•279•225•51•56•202•80•165	812•350•238•553	382•1397•27•147	514•284•22•39•545•177•372
S. mitis B6	698•611•105•180•252•101	136•559•551•165•177•235•124	106•1209•458•27•81•66	418•48•326•22•11•28•15•1079
S. mutans LJ23	82•96•117•879•141•84•309•245	–[a]	24•516•732•140•271•96•108•66	322•51•425•22•11•28•15•530•177•222•150
S. mutans UA159	82•96•117•311•568•141•84•309•80•165	–	24•516•732•411•96•108•66	322•476•22•11•28•15•530•177•222•150
S. oralis Uo5	76•102•520•290•189•68•64•42•596	136•39•13•221•192•94•551•701	106•371•57•781•334•28•123•81•66	418•48•326•33•28•15•242•288•48•351•150
S. parasanguinis FW213	109•395•188•213•413•58•132•57•60•151•108•66	793•99•41•238•528•251 817•11•28•15•641•438	73•118•413•94•716•536	
S. pasteurianus ATCC 43144	82•102•306•214•32•159•204•75•81•144•51•56•202•245	812•588•553	1341•438•27•147	373•141•284•22•39•545•111•66•372
S. pneumoniae TCH8431/19A	395•659•69•647•174	55•75•262•578•69•12•240•294•105•231•23	100•83•1241•75•445	235•72•584•543•75•66•25•185•78•81
S. pneumoniae Hungary19A-6	1309•285•108•80•165	–	–	–
S. pyogenes MGAS5005	–	–	514•57•533•610•123•81•35	–
S. pyogenes MGAS9429	–	–	–	90•283•249•176•33•28•15•641•438
S. pyogenes NZ131	–	–	–	90•283•458•28•15•641•438
S. salivarius 57.I	606•444•49•258•93•21•482	607•793•17•525•11	112•1571•270	514•306•39•545•399•150
S. sanguinis SK36	411•171•119•92•40•100•114•85•60•132•273•188•70•95	1048•124•778	537•1347•66	319•150•359•43•530•111•360•78
S. suis D12	–	–	1157•622•174	–
S. suis TL13	–	–	540•1205•34•174	–
S. uberis 0140 J	295•501•40•59•270•9•9•12•120•12•30•63•533	50•131•721•375•676	483•57•1192•47•27•81•66	322•351•147•39•15•641•438

Streptococcus spp.	Restriction endonucleases			
	HpyCH4 V	*RsaI*	*TaqI*	*Tru9I*
(b)				
S. agalactiae 2603 V/R	193•207•401•40•342•144•144•378•104	891•857•205	115•6•986•764•19•63	203•192•84•137•295•50•27•19•63•491•97•295
S. dysgalactiae subsp. equisimilis RE378	135•649•17•382•288•378•104	527•1411•15	158•57•51•140•482•321•92•70•63•98•384•37	461•167•90•193•30•47•82•314•315•251•3
*S. gordonii*str. Challis	132•666•193•135•57•663•104	516•196•176•660•387•15	163•4•467•106•160•75•256•137•161•384•37	419•585•63•570•313
S. intermedius JTH08	24•163•102•546•288•54•267•168•83•148•104	100•493•292•1047•15	223•13•268•1406•37	934•130•314•256•182•128•3
S. macedonicus ACA-DC 198	135•237•429•394•81•425•252	106•354•410•21•660•115•272•15	87•71•12•59•13•1059•615•37	395•516•50•46•63•491•79•18•164•128•3
S. mitis B6	24•771•334•144•674	799•86•660•387•15	109•114•37•22•500•513•231•358•26•37	389•191•975•97•36•259
S. mutans LJ23	–[a]	43•63•785•829•218•15	288•500•295•870	–
S. mutans UA159	–	43•63•785•829•218•15	788•295•870	–
S. oralis Uo5[b]	366•429•670•210•272	799•65•21•1047•15	109•506•9•32•639•247•342•26•37	580•1054•54•259
S. parasanguinis FW213[b]	798•526•144•234•248	888•1047•15	163•76•174•46•248•268•105•218•231•358•26•27•10	583•354•48•19•63•588•295
S. pasteurianus ATCC 43144	193•128•480•394•81•425•252	106•764•21•660•115•272•15	87•71•71•1072•615•37	395•516•50•46•63•491•97•164•128•3
S. pneumoniae TCH8431/19A	154•107•643•107•933	489•79•115•104•1016•141	103•55•33•1398•355	23•357•210•41•1310•3
S. pyogenes MGAS9429	–	–	115•184•1002•652	–
S. salivarius 57.I	135•160•77•429•817•63•27•245	86•231•202•351•21•660•197•190•15	158•12•72•24•396•605•34•70•161•339•45•37	13•382•27•476•13•96•63•570•18•295
S. sanguinis SK36	369•429•514•210•324•104	510•6•80•1339•15	112•100•922•779•27•10	–
S. suis ST1	–	–	87•71•12•402•766•140•54•373•11•37	–
S. uberis 0140 J	135•666•724•176•226•26	35•71•211•187•15•939•480•15	23•64•79•63•13•46•455•433•21•104•570•24•58	395•66•50•117•90•193•30•129•314•177•97•295

Symbol (•) indicates RE site in the gene sequences

[a] No unique pattern observed

[b] With RE-*Hin1I,* unique digestion patterns were recorded for *S. oralis* Uo5: 282•1658•7, and *S. parasanguinis* FW213: 44•1050•856

detection of *Streptococcus* infections especially in the cases of *S. mutans* (2 strains), *S. pneumoniae* (20 strains), *S. pyogenes* (3/7 strains), and *S. suis* (5/9 strains) (Table 5a, b). In the similar manner, analysis of *gyrB* gene, commonly used as biomarker for identification in general [13], was also not very effective as the unique RE patterns could not be deduced in the following cases: *S. pneumoniae* (18/20 strains), *S. pyogenes* (4/7 strains), and *S. suis* (6/9 strains) (Table 6a, b).

The present study has shown that (1) *purH* bears unique RE digestion characteristics for identifying 60 % of the strains representing all the 16 species, (2) a few other genes can be used for identifying another 36 % of the strains. For identifying the clinical isolates, the following protocol may be adopted. DNA extracted from the infected sample can be used to amplify the biomarker gene(s) with specific primer sets using standard molecular techniques. The amplified gene product can be either subjected to RE digestion and run on the gel or the gene can be sequenced and subjected to in silico RE digestion, and compare the patterns to identify the strain. This approach of using genes which are common to all the species enhances the chances of identifying the potential organism, as has been proposed for *Clostridium*, *Yersinia*, and *Vibrio*. Although, *purH* is also present in *Staphylococcus*, the RE digestion patterns did not match with that of *Streptococcus* (data not shown). These biomarkers have the potential for being applied and used in diagnostic kits for *Streptococcus*, a deadly pathogen for which drug targets are being search furiously [37, 38].

Acknowledgments We are thankful to the Director of CSIR-Institute of Genomics and Integrative Biology (IGIB), and CSIR projects—GENESIS (BSC0121) and INDEPTH (BSC0111) for providing the necessary funds, facilities and moral support. Authors are also thankful to the Academy of Scientific & Innovative Research (AcSIR), New Delhi.

References

1. Blaschke AJ (2011) Interpreting assays for the detection of *Streptococcus pneumoniae*. Clin Infect Dis 52:S331–S337. doi:10.1093/cid/cir048
2. Spellerberg B, Brandt C (2011) *Streptococcus*. In: Versalovic J, Carroll KC, Funke G, Jorgensen JH, Landry ML, Warnock DW (eds) Manual of clinical microbiology, 10th edn. ASM Press, Washington, pp 331–349
3. Dubois D, Segonds C, Prere MF, Marty N, Oswald E (2013) Identification of clinical *Streptococcus pneumoniae* isolates among other alpha and nonhemolytic streptococci by use of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. J Clin Microbiol 51:1861–1867. doi:10.1128/JCM.03069-12
4. Scholz CFP, Poulsen K, Kilian M (2012) Novel molecular method for identification of *Streptococcus pneumoniae* applicable to clinical microbiology and 16S rRNA sequence-based microbiome studies. J Clin Microbiol 50:1968–1973. doi:10.1128/JCM.00365-12
5. Chi F, Nolte O, Bergmann C, Ip M, Hakenbeck R (2007) Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. Int J Med Microbiol 297:503–512. doi:10.1016/j.ijmm.2007.02.009
6. Kellogg JA, Bankert DA, Elder CJ, Gibbs JL, Smith MC (2001) Identification of *Streptococcus pneumoniae* revisited. J Clin Microbiol 39:3373–3375. doi:10.1128/JCM.39.9.3373-3375.2001
7. Ferroni A, Suarez S, Beretti JL, Dauphin B, Bille E, Meyer J, Bougnoux M-E, Alanio A, Berche P, Nassif X (2010) Real-time identification of bacteria and *Candida* species in positive blood culture broths by matrix-assisted laser desorption ionization–time of flight mass spectrometry. J Clin Microbiol 48:1542–1548. doi:10.1128/JCM.02485-09
8. Stevenson LG, Drake SK, Murray PR (2010) Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 48:444–447. doi:10.1128/JCM.01541-09
9. Foster AGW (2013) Rapid identification of microbes in positive blood cultures by use of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. J Clin Microbiol 51:3717–3719. doi:10.1128/JCM.01679-13
10. Kawamura Y, Hou Sultana X-GF, Miura H, Ezaki T (1995) Determination of 16S rRNA sequences of *Streptococcus mitis*and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. Int J Syst Bacteriol 45:406–408. doi:10.1099/00207713-45-2-406
11. Abdeldaim G, Herrmann B, Mölling P, Holmberg H, Blomberg J, Olcén P, Strålin K (2010) Usefulness of real-time PCR for *lytA*, *ply*, and *Spn9802* on plasma samples for the diagnosis of pneumococcal pneumonia. Clin Microbiol Infect 16:1135–1141. doi:10.1111/j.1469-0691.2009.03069.x
12. Bhushan A, Joshi J, Shankar P, Kushwah J, Raju SC, Purohit HJ, Kalia VC (2013) Development of genomic tools for the identification of certain *Pseudomonas* up to species level. Indian J Microbiol 53:253–263. doi:10.1007/s12088-013-0412-1
13. Bhushan A, Mukherjee T, Joshi J, Shankar P, Kalia VC (2015) Insights into the origin of *Clostridium botulinum* strains: evolution of distinct restriction endonuclease sites in *rrs* (16S rRNA gene). Indian J Microbiol 55:140–150. doi:10.1007/s12088-015-0514-z
14. Kalia VC, Mukherjee T, Bhushan A, Joshi J, Shankar P, Huma N (2011) Analysis of the unexplored features of *rrs* (16S rDNA) of the genus *Clostridium*. BMC Genom 12:18. doi:10.1186/1471-2164-12-18
15. Lal D, Verma M, Lal R (2011) Exploring internal features of 16S rRNA gene for identification of clinically relevant species of the genus *Streptococcus*. Ann Clin Microbiol Antimicrob 10:28. doi:10.1186/1476-0711-10-28
16. Porwal S, Lal S, Cheema S, Kalia VC (2009) Phylogeny in aid of the present and novel microbial lineages: diversity in *Bacillus*. PLoS ONE 4:e4438. doi:10.1371/journal.pone.0004438
17. Kalia VC (2015) Let's explore the latent features of genes to identify bacteria. J Mol Genet Med 9:e105. doi:10.4172/1747-0862.1000E105
18. Kalia VC, Kumar P (2015) Genome wide search for biomarkers to diagnose *Yersinia* infections. Indian J Microbiol 55:366–374. doi:10.1007/s12088-015-0552-6
19. Kalia VC, Kumar P, Kumar R, Mishra A, Koul S (2015) Genome wide analysis for rapid identification of *Vibrio* species. Indian J Microbiol 55:375–383. doi:10.1007/s12088-015-0553-5
20. Kekre A, Bhushan A, Kumar P, Kalia VC (2015) Genome wide analysis for searching novel markers to rapidly identify *Clostridium* strains. Indian J Microbiol 55:250–257. doi:10.1007/s12088-015-0535-7

21. Bishop CJ, Aanensen DM, Jordan GE, Kilian M, Hanage WP, Spratt BG (2009) Assigning strains to bacterial species via the Internet. BMC Biol 7:3. doi:10.1186/1741-7007-7-3
22. Carvalho MG, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, Steigerwalt A, Whaley M, Facklam RR, Fields B, Carlone G, Ades GW, Dagan R, Sampson JS (2007) Evaluation and improvement of real-time PCR assays targeting *lytA*, ply, and *psaA* genes for detection of pneumococcal DNA. J Clin Microbiol 45:2460–2466. doi:10.1128/JCM.02498-06
23. Chen CC, Teng LJ, Chang TC (2004) Identification of clinically relevant viridans group streptococci by sequence analysis of the 16S-23S ribosomal DNA spacer region. J Clin Microbiol 42:2651–2657. doi:10.1128/JCM.42.6.2651-2657.2004
24. Drancourt M, Roux V, Fournier PE, Raoult D (2004) *rpoB* gene sequence-based identification of aerobic Gram-positive cocci of the genera *Streptococcus*, *Enterococcus*, *Gemella*, *Abiotrophia*, and *Granulicatella*. J Clin Microbiol 42:497–504. doi:10.1128/JCM.42.2.497-504.2004
25. Picard FJ, Ke D, Boudreau DK, Boissinot M, Huletsky A, Richard D, Ouellette M, Roy PH, Bergeron MG (2004) Use of *tuf* sequences for genus-specific PCR detection and phylogenetic analysis of 28 streptococcal species. J Clin Microbiol 42:3686–3695. doi:10.1128/JCM.42.8.3686-3695.2004
26. Teng LJ, Hsueh PR, Tsai JC, Chen PW, Hsu JC, Lai HC, Lee CN, Ho SW (2002) *groESL* sequence determination, phylogenetic analysis, and species differentiation for viridans group streptococci. J Clin Microbiol 40:3172–3178. doi:10.1128/JCM.40.9.3172-3178.2002
27. Zbinden A, Kohler N, Bloemberg GV (2011) *recA*-based PCR assay for accurate differentiation of *Streptococcus pneumoniae* from other viridans streptococci. J Clin Microbiol 49:523–527. doi:10.1128/JCM.01450-10
28. Glazunova OO, Raoult D, Roux V (2010) Partial *recN* gene sequencing: a new tool for identification and phylogeny within the genus *Streptococcus*. Int J Syst Evol Microbiol 60:2140–2148. doi:10.1099/ijs.0.018176-0
29. Nielsen XC, Justesen US, Dargis R, Kemp M, Christensen JJ (2009) Identification of clinically relevant nonhemolytic streptococci on the basis of sequence analysis of 16S-23S intergenic spacer region and partial *gdh* gene. J Clin Microbiol 47:932–939. doi:10.1128/JCM.01449-08
30. Park HK, Yoon JW, Shin JW, Kim JY, Kim W (2010) *rpoA* is a useful gene for identification and classification of *Streptococcus pneumoniae* from the closely related viridans group streptococci. FEMS Microbiol Lett 305:58–64. doi:10.1111/j.1574-6968.2010.01913.x
31. Poyart C, Quesne G, Coulon S, Berche P, Trieu-Cuot P (1998) Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. J Clin Microbiol 36:41–47
32. Wessels E, Schelfaut JJG, Bernards AT, Claas ECJ (2012) Evaluation of several biochemical and molecular techniques for identification of *Streptococcus pneumoniae* and *Streptococcus pseudopneumoniae* and their detection in respiratory samples. J Clin Microbiol 50:1171–1177. doi:10.1128/JCM.06609-11
33. Hanage WP, Kaijalainen T, Herva E, Saukkoriipi A, Syrjänen R, Spratt BG (2005) Using multilocus sequence data to define the pneumococcus. J Bacteriol 187:6223–6230. doi:10.1128/JB.187.17.6223-6230.2005
34. Innings Å, Krabbe M, Ullberg M, Herrmann B (2005) Identification of 43 *Streptococcus* species by pyrosequencing analysis of the *rnpB* gene. J Clin Microbiol 43:5983–5991. doi:10.1128/JCM.43.12.5983-5991.2005
35. Kilian M, Poulsen K, Blomqvist T, Havarstein LS, Bek-Thomsen M, Tettelin H, Sorensen UB (2008) Evolution of *Streptococcus pneumoniae* and its close commensal relatives. PLoS ONE 3:7. doi:10.1371/journal.pone.0002683
36. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98
37. Sajid A, Arora G, Singhal A, Kalia VC, Singh Y (2015) Protein phosphatases of pathogenic bacteria: role in physiology and virulence. Annu Rev Microbiol 69:527–547. doi:10.1146/annurev-micro-020415-111342
38. Koul S, Prakash J, Mishra A, Kalia VC (2015) Potential emergence of multi-quorum sensing inhibitor resistant (MQSIR) bacteria. Indian J Microbiol. doi:10.1007/s12088-015-0558-0