This is a repository copy of *Identification of two new core chromosome-encoded superantigens in Streptococcus pyogenes; speQ and speR*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/142818/

Version: Accepted Version

Article:
Reglinski, M., Sriskandan, S. and Turner, C. orcid.org/0000-0002-4458-9748 (2019) Identification of two new core chromosome-encoded superantigens in Streptococcus pyogenes; speQ and speR. Journal of Infection. ISSN 0163-4453

https://doi.org/10.1016/j.jinf.2019.02.005

Article available under the terms of the CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Identification of two new core chromosome-encoded superantigens in Streptococcus pyogenes; speQ and speR

Mark Reglinski[1], Shiranee Sriskandan[1], Claire E Turner[1,2].

[1]Department of Infectious Diseases, Imperial College London, London, UK, W12 0NN

[2]Department of Molecular Biology & Biotechnology, The Florey Institute, University of Sheffield, Sheffield, UK, S10 2TN

Running head: New S. pyogenes superantigens speQR

Corresponding author:

Dr Claire Turner

Department of Molecular Biology & Biotechnology

University of Sheffield

Sheffield

S10 2TN

Tel: 0114 222 2819

Email: c.e.turner@sheffield.ac.uk

Abstract

Superantigens are ubiquitous within the Streptococcus pyogenes genome, which suggests that superantigen-mediated T-cell activation provides a significant selective advantage. S. pyogenes can carry a variable complement of the 11 known superantigens. We have identified two novel S. pyogenes superantigens, denoted speQ and speR, adjacent to each other in the core-chromosome of isolates belonging to eleven different emm-types. Although distinct from other superantigens, speQ and speR were most closely related to speK and speJ respectively. Recombinant SPEQ and SPER were mitogenic towards human peripheral blood mononuclear cells at ng/ml concentrations, and SPER was found to be more mitogenic than SPEQ.

Key words:

Streptococcus pyogenes; group A Streptococcus; superantigens; scarlet fever; toxic shock syndrome; mitogen

Background

The human pathogen, Streptococcus pyogenes, produces numerous virulence factors, including the extracellular superantigen toxins which are associated with the development of streptococcal toxic shock syndrome and scarlet fever. Superantigens are able to cross-link the major histocompatibility complex (MHC) class II with the T cell receptor, bypassing the normal antigen presenting process and resulting in a high level of T cell activation, cytokine release and inflammation [1]. Although superantigens are thought to promote invasive disease, associated with high mortality rates, they may have a localized role in productive infection of the nasopharynx [2].

Genetically and structurally related superantigens are also produced by Staphylococcus aureus and some other streptococcal species including the group C/G streptococci. There are 14 known streptococcal superantigens, 11 of which have been found in S. pyogenes; speA, speC, speG-speM, smeZ and ssa. All, except speJ and ssa, have been found in other streptococcal species as well. Three superantigen genes, szeN, szeP and szeF have been found only in S. equi subsp zooepidemicus [3]. Commons et al. later suggested renaming these to speN, speP and speO respectively to standardize the nomenclature across all streptococci [1].

The superantigen genes speG, speJ and smeZ are encoded on the core chromosome but are not ubiquitous among S. pyogenes isolates. The other eight identified S. pyogenes superantigens are associated with prophages which have the potential to be mobile, introducing variability among isolates. As there is variability in the complement of superantigens carried by S. pyogenes isolates, along with mobility and sharing across other streptococcal species, there may be streptococcal superantigens that are yet to be identified.

Whilst testing for the presence of the 11 known superantigens in whole genome sequence (WGS) data from two S. pyogenes isolates from the pre-antibiotic era, we identified two new adjacent potential superantigen genes. We subsequently confirmed the presence of both genes in WGS from modern S. pyogenes isolates of different emm genotypes. The two new potential superantigen genes were not associated with recognised mobile genetic elements but were limited to certain emm-types. We have termed the genes speQ and speR, to follow the proposed nomenclature, and confirmed that they are indeed mitogenic towards human mononuclear cells.

Methods

Bacterial strains and growth conditions

Two S. pyogenes emm60 isolates (H865 and H870) were first isolated in 1938 from puerperal sepsis patients at Queen Charlottes Hospital London. S. pyogenes isolates were cultured on Columbia blood agar plates or statically in Todd Hewitt broth at 37 °C with 5% CO_2. E. coli were cultured in LB at 37 °C with 225 rpm agitation and supplemented with 100ug/ml ampicillin where appropriate.

Gene identification

The speQ and speR genes were first identified from whole genome sequence (WGS) data of emm60 isolates H865 and H870 (short read archive ERR485817 and ERR485821, respectively). The presence and sequence of speQR were confirmed by PCR with primers spanning from the upstream ideS gene and the downstream hypothetical gene (speQR-region primers listed in Supplementary Table 1), and Sanger sequencing.

Whole genome sequence analysis

Publicly available genome sequence fastq data for other S. pyogenes strains representing 86 different emm-types [4-6] were obtained and assembled de novo for identifying the presence or absence of speQ and speR.

Fastq reads were assembled de novo using Velvet [7] or as previously described [6]. Assembly statistics are available at Mendeley Data (doi:10.17632/b89yzfcxp8.1 and doi:10.17632/f2d39nsfwk.1) or in the original study [6]. The genomic region spanning the potential speQR locus was extracted from de novo assemblies and examined for the presence or absence of complete superantigen genes, but in some cases the quality of assembled sequence data was too low for adequate confirmation of the complete speQR locus or the allelic sequence.

Multi-locus sequence (MLST) data were obtained from de-novo assemblies of some UK data [8]. The presence of the 11 known streptococcal superantigens was determined by mapping of the short read sequence data to a pseudosequence of concatenated genes and confirmed through BLAST analysis of de-novo assemblies. Other MLST, emm-genotype and superantigen data were obtained from the original studies. We excluded isolates where the emm type could not be definitively assigned. This included WGS from Kapatai et al. [5] where emm-type determined by WGS was reported to be different to the original Sanger sequenced emm-type, and emm-negative isolates from Chochua et al. [4].

For phylogenetic analysis, fastq data was mapped to the completed S. pyogenes emm89 reference genome H293 (HG316453.2) [9] and single nucleotide polymorphisms extracted from the core genome using SNP-sites [10] to generate a maximum likelihood phylogeny with RAxML [11].

Different speQR DNA sequence alleles were identified and submitted to Genbank (accession numbers: BK010649- BK010666, BK010692, BK010693).

Recombinant protein expression

BamHI-ended coding sequences for speQ.1 and speR.1 were amplified from H865 gDNA using the primers listed in Supplementary Table 1 and cloned into the overexpression vector pET-19b (Novagen). Recombinant proteins were expressed in One Shot BL21(DE3) Chemically Competent E. coli (Life Technologies) and purified to apparent homogeneity using the Ni-NTA purification system (Novagen). Protein concentrations were measured using the Pierce Coomassie Plus (Bradford) Assay Kit. 1 µg aliquots of each protein were separated by SDS-PAGE and visualized by InstantBlue staining (Expedeon). Recombinant IdeS was produced as previously described [12] and purified alongside recombinant SPEQ and SPER. Recombinant SPEC was purchased from Toxin Technology, Inc (Sarasota, Florida).

SPEQ and SPER antisera

Antisera towards SPEQ and SPER were raised by immunizing mice intramuscularly with 10 µg of recombinant protein, emulsified 1:1 with Freund's complete adjuvant, and booster immunizations at 21 and 35 days in Freund's incomplete adjuvant. Blood was collected on day 42 and the resulting antiserum was pooled.

Human mononuclear cell proliferation assay.

Healthy donor human mononuclear cells (MNCs) were purified as previously described [13] and diluted to $1x10^6$ cells/ml in RPMI media (Life Technologies) (+10% FCS). Cells were seeded into 96 well plates at a concentration of $2x10^5$/well and incubated with decreasing concentrations of recombinant protein for 48 h. Cell proliferation was measured with a Colorimetric Cell Proliferation BrdU ELISA (Roche).

Immunoblotting

S. pyogenes culture supernatants collected at different time points were concentrated 16-fold by TCA precipitation. Separated proteins were transferred to PVDF membrane (Hybond-LFP, GE Healthcare) which were blocked with 5% non-fat milk (Sigma Aldrich) in PBS and probed with a 1:1000 dilution of mouse antiserum raised against SPEQ or SPER. Bound antibodies were detected using a 1:80,000 dilution of HRP-conjugated goat anti-mouse IgG (Abcam).

RNA extraction and PCR

RNA was extracted from early, mid and late logarithmic growth phases of the two emm60 strains (H865 and H870) using a hot acidic phenol method as previously described [14]. RNA samples were treated with Turbo DNA-free (Ambion) DNase and 1 µg was converted into cDNA using Transcriptor reverse transcriptase (Roche) and random hexaoligos (RT+ samples). To control for contaminating genomic DNA equivalent reverse transcriptase negative (RT-) samples were generated with another 1 µg but the Transcriptor reverse transcriptase was excluded from the reaction. PCR for speQR co-transcription was performed on 200 ng of the RT+ cDNA and RT- samples using speQR primers (Supplementary Table 1) and visualized on an agarose gel.

Ethics

Normal human donor cells were acquired from an approved sub-collection of the Imperial College Tissue Bank. All murine procedures were approved by the local ethical review process at Imperial College London and conducted in accordance with the relevant, UK Home Office approved, project license.

Results

Identification of two new potential superantigen genes; speQ and speR

We sequenced the genomes of two viable emm60 isolates, originally collected in 1938 from two puerperal sepsis patients, and analyzed the genomes for the presence of superantigens. We could not detect the presence of any of the known streptococcal superantigens by short read sequence mapping or BLAST analysis of de novo assembled genomes. The analysis did, however, indicate the presence of sequence in the genomes of both emm60 isolates with partial homology to speK. We identified this homologous sequence to be within one of two hypothetical genes located immediately downstream of the gene ideS (also known as mac) encoding for an immunoglobulin cleaving protease (Figure 1A). BLAST indicated that these genes were closely related to other streptococcal superantigens and carried the typical superantigen C terminal β-grasp domain [1]. We therefore predicted that these would be superantigen genes and denoted them speQ and speR. PCR and Sanger sequencing confirmed the WGS data. BLASTn and BLASTp of completed available S. pyogenes genomes also identified speQ and speR in an emm87 strain NGAS743 (DI45_05770 and DI45_05775 respectively; Genbank CP007560.1) [16]. In isolates where full length speQ and speR genes were absent, a C-terminal fragment of speR was present immediately downstream of ideS (Figure 1A). We also performed BLASTp analysis of the entire NCBI database, excluding S. pyogenes, but did not identify SPEQ or SPER in any other available genomes including other streptococcal species.

Phylogenetic analysis of the amino acid sequences of SPEQ, SPER, and all other available superantigen alleles from all streptococcal species [1] demonstrated that, although phylogenetically distinct, SPEQ is closely related to the prophage-associated SPEK sharing 84% amino acid identity, and SPER is most closely related to the chromosomal SPEJ sharing 64% amino acid identity (Figure 1B). Comparisons were

made between SPEQ, SPEK, SPEJ and SPER to identify two superantigen signature amino acid motifs (Supplementary Figure 1) [1, 3]. SPER, like SPEJ, had the motif Y-G-G-(LIV)-T-x_4-N (Prosite PS00277) but only a partial match for this was identified in SPEQ and SPEK. All four superantigens had the motif K-x_2-(LIVF)-x_4-(LIVF)-D-x_3-R-x_2-L-x_5-(LIV)-Y (Prosite PS00278) and a C-terminal zinc binding domain (HxD).

To determine the presence of speQR in other S. pyogenes genotypes, publicly available WGS fastq data were obtained from the short read archive for UK isolates [5, 6] and USA isolates [4] totaling 4,262 genomes tested covering 86 different genotypes (Supplementary Table 2). Complete speQ and speR were identified in the assembled genome sequence of isolates belonging to the emm-types emm9, 15, 18, 42, 53, 58, 60, 77, 87, 94 and 169 (Supplementary Table 3). However, not all isolates belonging to some of these genotypes carried the complete speQR locus, which was unexpected given the lack of association with mobile genetic elements. Only one out of 41 emm18 (USA isolate 20154046) had complete speQR, as did 21/24 emm58 isolates and 49/72 emm77 isolates. The presence or absence of complete speQR in these genotypes appeared to be associated with divergent lineages and multi-locus sequence types (MLST) within these emm-types (Supplementary Figure 2, Supplementary Table 3), indicative of the same emm gene on completely different genetic backgrounds. In contrast, all emm94 isolates were MLST-89, but 2/50 did not carry the complete speQR allele and formed a separate sub-lineage. As these isolates were still relatively closely related there may have been a horizontal gene transfer event of the speQR region.

The majority of isolate genomes that were positive for speQR were also positive for at least one other superantigen gene (Supplementary Table 3). The exceptions to this were

4/5 emm60, 1/72 emm77, and 2/2 emm169 isolates where no superantigen genes other than speQR were detected [4].

From the WGS analysis, thirteen DNA alleles for speQ and seven DNA alleles for speR were identified. The variation was limited to single nucleotide polymorphisms, except a region in speQ which varied in the number of a 15bp/5aa repeat (Supplementary Figure 3). This 15bp/5aa region repeated twice, four times and five times in three alleles, speQ.2, speQ.4 and speQ.5 respectively; these alleles were found only in genotype emm9 isolates. Based on amino acid sequence, SPEQ.1, SPEQ.6, SPEQ.8, SPEQ.11 and SPEQ.12 were identical. SPEQ.3, SPEQ.7, SPEQ.9 and SPEQ.13 each differ from SPEQ.1 by one amino acid residue and SPEQ.10 differs by two amino acid residues. For SPER, SPER.1, SPER.2, SPER.4, and SPER.5 were identical by amino acid sequence, but SPER.3, SPER.6 and SPER.7 each differ by one amino acid.

Recombinant SPEQ and SPER induced proliferation of human mononuclear cells

To determine if SPEQ and SPER were capable of inducing proliferation of human T cells, we recombinantly expressed both proteins in E. coli (Figure 2A). These recombinant toxins represented gene alleles speQ.1 and speR.1. Purified toxins were then used to stimulate human mononuclear cells (MNCs) and proliferation was measured by BrdU uptake (Figure 2B). Both SPEQ and SPER induced proliferation, although a 10-fold greater concentration of SPEQ than SPER was required to generate an equivalent response. Proliferation after stimulation with another streptococcal superantigen, SPEC, required 100-1000-fold lower concentration than SPEQ and SPER. As a control, the non-mitogenic IdeS was recombinantly expressed and purified in the same manner as SPEQ and SPER but failed to stimulate any proliferation, as expected.

Both speQ and speR are expressed by S. pyogenes during culture

To confirm speQ and speR expression by S. pyogenes, transcription and protein expression were measured. RNA was extracted at early, mid and late-logarithmic phases of growth of two emm60 strains and converted to cDNA for PCR. Primers that spanned across both speQ and speR confirmed the two genes are co-transcribed (Figure 2C). Although only semi-quantitative, transcription appeared greatest at early and mid-logarithmic phases of growth.

Culture supernatants from the same two strains of emm60 S. pyogenes were probed by Western blot for SPEQ and SPER using an antibody raised in mice against recombinant proteins. SPEQ could be detected at late-logarithmic phase and following overnight culture in both emm60 strains (Figure 2D). Using rSPEQ at known concentrations the estimated concentration of SPEQ was ~90-127 ng/ml in late-logarithmic phase culture and increased to ~155-163 ng/ml by overnight culture. We were, however, unable to detect SPER using the rSPER murine antibody in either strain at any growth phase, which was unexpected given the co-transcription.

Discussion

We identified two potential superantigen genes present in the chromosomes of two 1930's S. pyogenes emm60 isolates and subsequently identified the same genes in isolates belonging to 10 other emm-types in modern international isolates. We termed these genes speQ and speR and confirmed that they were capable of inducing proliferation of human cells.

We tested the genomes over 4000 different isolates representing 86 emm-types and detected speQR in strains belonging to emm9, 15, 18, 42, 53, 58, 60, 77, 87, 94 and 169, although both speQR positive and negative lineages existed within these genotypes

(Supplementary Figure 2). Both emm77 and emm87 have been reported as common causes of invasive disease in various countries [17].

Like the majority of other superantigens, both speQ and speR carry two of the three classic superantigen motifs. The third was absent in speQ, as also observed in the closest relative speK, although present in speR and may relate to the different mitogenic potential; 10-fold more SPEQ than SPER was required to generate an equivalent mitogenic response. The mitogenic activity of both SPEQ and SPER was 10-100 fold lower than that of SPEC. This may limit contribution of SPEQ/R to virulence in the presence of much more potent superantigens. The majority of isolates whose genomes tested positive for speQ/R also carried at least one other superantigen genes.

Across the entire collection of 1441 USA isolates (previously all typed for the 11 known superantigen genes), the prevalence of speQR was 6%, similar to speL (5%) and speM (6%) [4]. The most commonly found superantigen gene within the USA collection was speG (93%) followed by smeZ (91%), speC (51%), speJ (41%), speA (26%), speH (25%), speI (23%), ssa (10%), speK (9%) [4]. Of those that were positive for speQR, the prevalence of smeZ was still high (94%) and similar for speK (11%) and speC (40%), but fewer were positive for speG (54%) as well as speA (4%), speH (4%), speI (1%), speJ (28%), and more were positive for ssa (41%), speL (16%) and speM (16%). This may, however, reflect an association of superantigen complement with emm-type. At least one superantigen gene was detected in all 1441 USA isolates, except for four (of 5) emm60 isolates and one (of 54) emm77 isolate [4]. We identified that these five 'superantigen negative' isolates were positive for speQR, consistent with our initial finding that two 1930s emm60 were only positive for speQR and no other known superantigens.

Interestingly, in all isolates where full length speQ and speR genes were absent, immediately downstream of ideS was a C-terminal fragment of speR (Figure 1A). This suggests that speQR genes were present in the most recent common ancestor of all emm-types but were lost as the emm-types diverged. Despite being chromosomally encoded, speQ and speR were less frequent among contemporary sequenced isolates than the other chromosomally encoded superantigens speG, speJ and smeZ. Quite why speQ and speR have persisted in the emm types that we identified is unclear; this may reflect a requirement of S. pyogenes to express at least one superantigen.

Although superantigens are implicated in the pathogenesis of scarlet fever and streptococcal toxic shock, it is widely recognized that the production of superantigens must play a role in S. pyogenes fitness [18, 19]. This has been borne out in epidemiological studies that show alterations of superantigen gene content to be implicated in emergence or expansion of new lineages in the population [20, 22, 23], and in animal models where superantigens are shown to be necessary for successful infection [2, 19].

Superantigens crosslink the major histocompatibility molecule (MHC) class II on antigen presenting cells with the T cell receptor. The binding of the T cell receptor occurs through the variable (V) β-region, with a specificity towards different Vβ variants [18]. As each superantigen can activate a different repertoire of Vβ, the more superantigens expressed by S. pyogenes the greater the heterogeneity of T cell expansion. Further work is required to determine the Vβ preference of SPEQ and SPER but it may be that even in the presence of more potent superantigens, they still contribute to pathogenesis through the expansion of a different T cell repertoire.

295 Genetic comparison of speQ and speR with the other known streptococcal
296 superantigens indicated that they were closely related to the prophage-associated speK
297 and the core chromosomal speJ, respectively. As SPEQ shares 84% amino acid identity
298 with SPEK, it is possible that speK originated from speQ that was picked up by a
299 bacteriophage. We did not identify any speQR orthologues in the available genomes of
300 other streptococcal species.

301 Although we identified thirteen different alleles of speQ, the most common was speQ.1,
302 found in 81% (174/216) of speQR-positive isolate genomes where we could confirm
303 the allele. It was also the original allele we identified in the emm60 isolates from the
304 1930s. The alleles speQ.2, speQ.4 and speQ.5 that vary by a repeat region were
305 restricted to the emm9 genotype. It is unclear as to the significance of this or the impact
306 the repeat region may have on mitogenic activity of these alleles. The original 1930s
307 emm60 speR allele, speR.1 was found in 40% (82/216) but the most common was
308 speR.2, found in 56% (120/216) of speQR-positive isolate genomes; these two alleles
309 should, however, encode identical proteins.

310 Co-transcription of speQR was detected with greater expression in the early stages of
311 exponential growth. Surprisingly, we were unable to detect SPER protein in S.
312 pyogenes culture supernatant, although we were able to detect SPEQ at levels similar
313 to SPEC expressed by emm3 S. pyogenes [22]. Other superantigens have been shown
314 to be sensitive to SPEB degradation [23], and this may be the case for SPER. It is also
315 possible that the antibody we raised, while able to detect recombinant SPER, was
316 unable to detect native SPER.

317 We have shown that there are now 16 streptococcal superantigen genes; each one may
318 play a role in promoting S. pyogenes virulence. While the benefits of superantigen

production remain incompletely understood, the ubiquitous presence of superantigen genes within the S. pyogenes genome suggests that they do play a significant role in S. pyogenes disease.

Funding

This work was supported by the NIHR Biomedical Research Centre awarded to Imperial College NHS Healthcare Trust (SS, MR), and Imperial College JRF (CET).

Acknowledgements The authors acknowledge the NIHR Biomedical Research Centre Tissue Bank at Imperial College

References

1. Commons RJ, Smeesters PR, Proft T, Fraser JD, Robins-Browne R, Curtis N. Streptococcal superantigens: categorization and clinical associations. Trends Mol Med **2014**; 20:48-62.
2. Zeppa JJ, Kasper KJ, Mohorovic I, Mazzuca DM, Haeryfar SMM, McCormick JK. Nasopharyngeal infection by Streptococcus pyogenes requires superantigen-responsive Vbeta-specific T cells. Proc Natl Acad Sci U S A **2017**.
3. Paillot R, Darby AC, Robinson C, et al. Identification of three novel superantigen-encoding genes in Streptococcus equi subsp. zooepidemicus, szeF, szeN, and szeP. Infect Immun **2010**; 78:4817-27.
4. Chochua S, Metcalf BJ, Li Z, et al. Population and whole genome sequence based characterization of invasive group A streptococci recovered in the United States during 2015. MBio **2017**; 8.

344 5. Kapatai G, Coelho J, Platt S, Chalker VJ. Whole genome sequencing of group A
345 Streptococcus: development and evaluation of an automated pipeline for emm
346 gene typing. PeerJ **2017**; 5:e3226.
347 6. Turner CE, Bedford L, Brown NM, et al. Community outbreaks of group A
348 Streptococcus revealed by genome sequencing. Sci Rep **2017**; 7:8554.
349 7. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de
350 Bruijn graphs. Genome Res **2008**; 18:821-9.
351 8. Page AJT, B.; Keane, J.A. Multilocus sequence typing by blast from de novo
352 assemblies against PubMLST. Journal of Open Source Software **2016**; 1.
353 9. Turner CE, Abbott J, Lamagni T, et al. Emergence of a New Highly Successful
354 Acapsular Group A Streptococcus Clade of Genotype emm89 in the United
355 Kingdom. MBio **2015**; 6:e00622.
356 10. Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs
357 from multi-FASTA alignments. Microb Genom **2016**; 2:e000056.
358 11. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis
359 of large phylogenies. Bioinformatics **2014**; 30:1312-3.
360 12. Reglinski M, Lynskey NN, Sriskandan S. Modification of the classical Lancefield
361 assay of group A streptococcal killing to reduce inter-donor variation. J
362 Microbiol Methods **2016**; 124:69-71.
363 13. Unnikrishnan M, Altmann DM, Proft T, et al. The bacterial superantigen
364 streptococcal mitogenic exotoxin Z is the major immunoactive agent of
365 Streptococcus pyogenes. J Immunol **2002**; 169:2561-9.
366 14. Turner CE, Kurupati P, Jones MD, Edwards RJ, Sriskandan S. Emerging role of the
367 interleukin-8 cleaving enzyme SpyCEP in clinical Streptococcus pyogenes
368 infection. J Infect Dis **2009**; 200:555-63.

15. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods **2011**; 8:785-6.

16. Athey TB, Teatero S, Sieswerda LE, et al. High incidence of invasive group A Streptococcus disease caused by strains of uncommon emm types in Thunder Bay, Ontario, Canada. J Clin Microbiol **2016**; 54:83-92.

17. Gherardi G, Vitali LA, Creti R. Prevelent emm types among invasive GAS in Europe and North America since year 2000. Front Public Health **2018**; 6:59.

18. Proft T, Fraser JD. Streptococcal Superantigens: Biological properties and potential role in disease. In: Ferretti JJ, Stevens DL, Fischetti VA, editors. Streptococcus pyogenes : Basic Biology to Clinical Manifestations [Internet]. Oklahoma City. **2016.**

19. Kasper KJ, Zeppa JJ, Wakabayashi AT et al. Bacterial superantigens promote acute nasopharyngeal infection by Streptococcus pyogenes in a human MHC Class II-dependent manner. PLoS Pathog **2014;** e1004155.

20. Nasser W, Beres SB, Olsen RJ et al. Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. PNAS **2014**; 111:E1768-E1776.

21. Al-Shahib A, Underwood A, Afshar B et al. Emergence of a novel lineage containing a prophage in emm/M3 group A Streptococcus associated with upsurge in invasive disease in the UK. Microb Genom **2016**; e000059.

22. Afshar B, Turner CE, Lamagni TL, et al. Enhanced nasopharyngeal infection and shedding associated with an epidemic lineage of emm3 group A Streptococcus. Virulence **2017**; 8:1390-400.

23. Kansal RG, Nizet V, Jeng A, Chuang WJ, Kotb M. Selective modulation of superantigen-induced responses by streptococcal cysteine protease. J Infect Dis **2003**; 187:398–407.

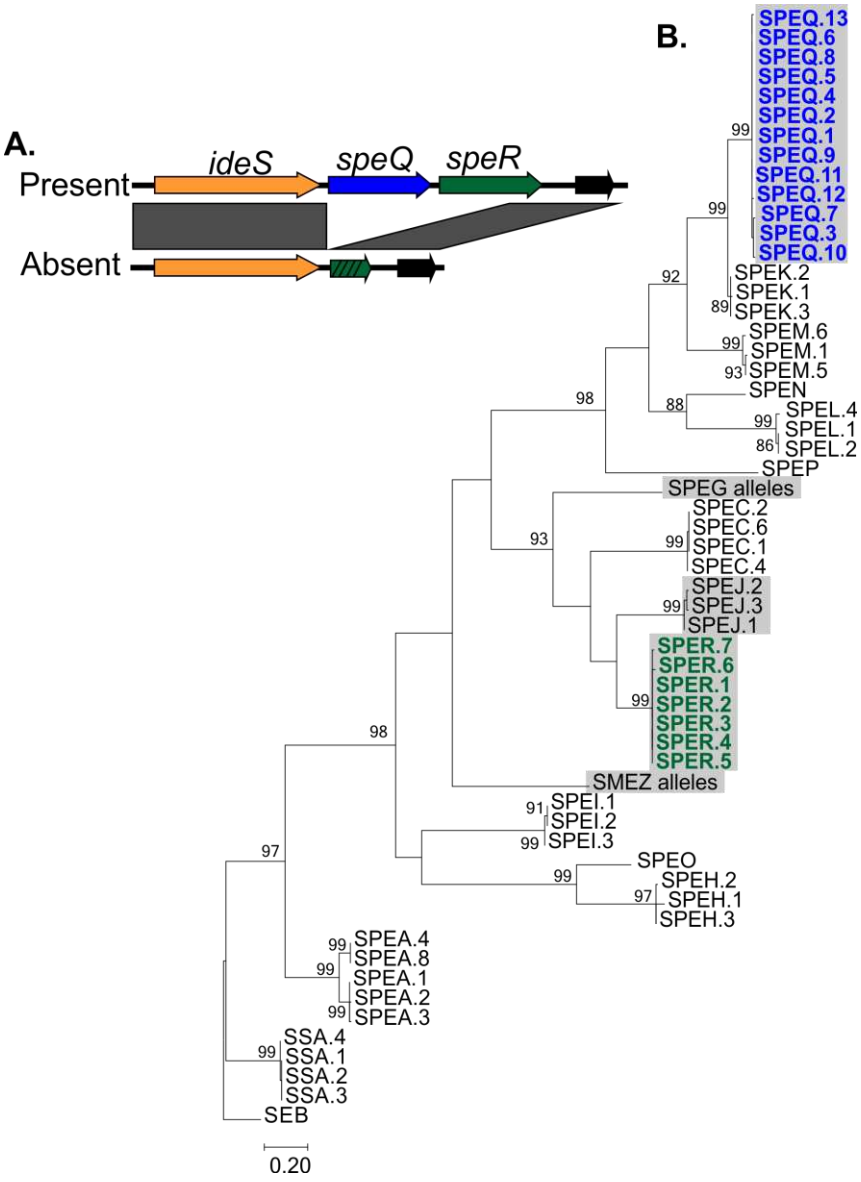


Figure 1. Schematic representation of the chromosomal location of speQ and speR and similarity to other superantigens. (A) In emm-types where the two superantigens genes speQ (blue) and speR (green) were present in the chromosome they were located immediately downstream of the gene ideS (orange) encoding for the immunoglobulin cleaving protease, and upstream of a gene encoding for a hypothetical protein (black). In emm-types where the full length superantigen genes were absent in the chromosome, a C-terminal ~364 base pair fragment of speR was present downstream of ideS. (**B**) Available sequences of the superantigen alleles determined by Commons et al [1] were

obtained and translated to amino acids. The signal sequences were identified by SignalP [15] and excluded before alignment with the predicted mature protein sequences of SPEQ (blue) and SPER (green) alleles. Core chromosome (i.e. not associated with known prophage elements) S. pyogenes superantigens are shaded grey. The Staphylococcus aureus superantigen SEB sequence was also included and used to root the maximum likelihood tree. The 28 SPEG alleles and 39 SMEZ alleles were condensed to a single branch for illustration purposes. Scale represents amino acid substitutions per site. Bootstrap values greater than 80% are shown.

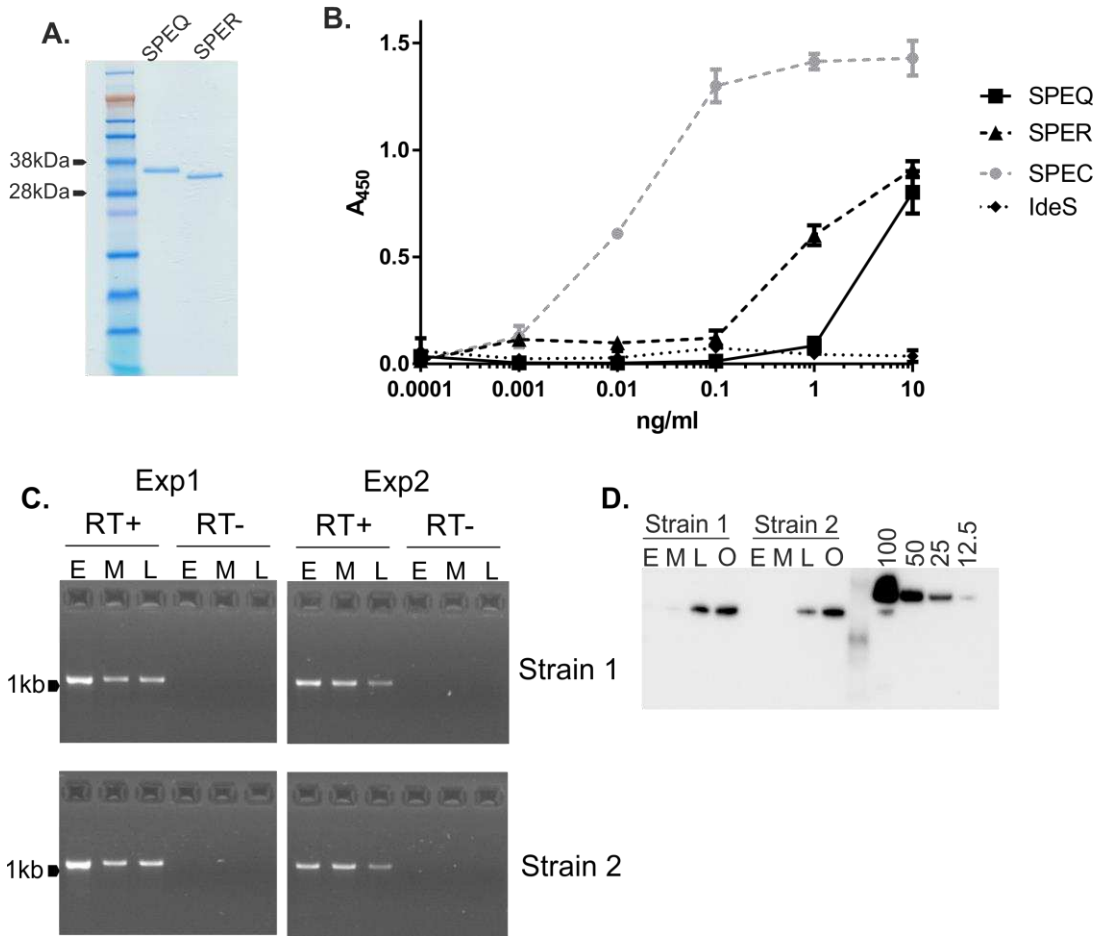


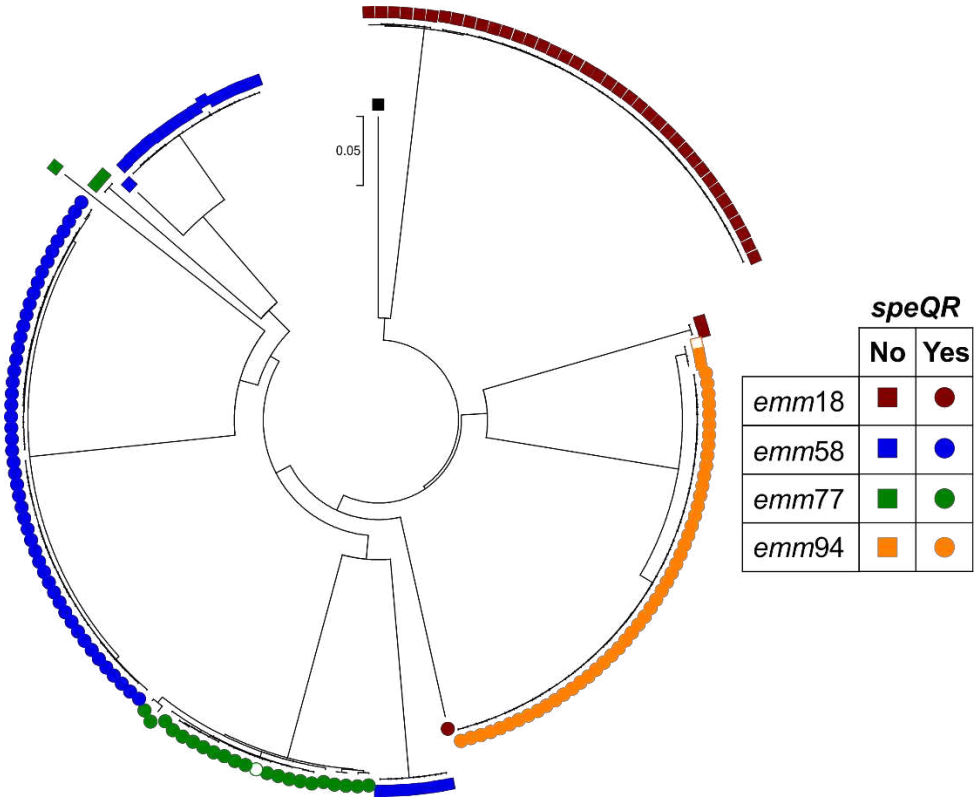
Figure 2. Activity and expression of SPEQ/R. (**A**) Instant blue stained SDS-PAGE gel of purified recombinantly-expressed SPEQ and SPER which were then used to stimulate human mononuclear cells (MNCs). (**B**) Proliferation of MNCs, measured by BrdU assay, required 10 fold more SPEQ (solid line, squares) than SPER (dashed line, triangles). The streptococcal superantigen SPEC (grey dashed line, circles) induced proliferation at a concentration ~100 fold lower than SPER. The protein IdeS (dotted line, diamonds) which was expressed and purified in the same manner as SPEQ and SPER, induced no proliferation of MNCs at any concentration used. Data represent mean and SD from a single MNC donor representative of data from two other donors. (**C**) RNA was extracted from two strains of S. pyogenes emm60 cultured to early (E), mid (M) and late (L) logarithmic phase and converted to cDNA. This was repeated on

two separate occasions (Exp1 and Exp2) and gave very similar results. PCR was performed on 200ng of cDNA using primers that spanned speQ and speR and, for both strains and in both experiments, a band was detected at ~1kb following PCR on the RT+ samples, confirming co-transcription of speQ and speR. No bands were present in samples where the reverse transcriptase had been excluded from the RNA to cDNA reverse transcription reaction (RT-), suggesting no contamination of genomic DNA. (**D**) SPEQ was detected by Western blot in the culture supernatant of two emm60 strains (Strain 1 and Strain 2) at late logarithmic phase (L) and following overnight culture (O), but not at early (E) and only faintly at mid (M) logarithmic phase. A standard curve of 100, 50, 25 and 12.5 ng of rSPEQ was included to provide quantification. Culture supernatants were concentrated 20-fold.

Supplementary Figures

	PS00277	PS00278	Zn
SPEQ	DGGIIKTSDV	NNMVTLQEIDVRLRKSLMGDSKIK	HFD
SPEK	DGGIIKTSDV	NNIVTLQEIDVRLRKSLMGDSKIK	HFD
SPER	YGGITPSTDD	KDIVTIQEFDFKIRKFLMESKEIY	HFD
SPEJ	YGGVTPSVNS	KPIFTIQEFDFKIRQYLMQTYKIY	HFD
	YGG*TxxxxN	Kxx*xxxx*DxxxRxxLxxxxx*Y	HxD
	(LIV)	(LIVF)(LIVF) (LIV)	

Supplementary Figure 1. Identification of three signature domains found in superantigens. The amino acid sequences of SPEQ and SPER were compared to their closest relatives SPEK and SPEJ to identify three motifs typically found in superantigens. The domain PS00277 (Prosite) was found in SPER and SPEJ with only one mismatch to the consensus sequence Y-G-G-(LIV)-T-x_4-N. Only three bases of this consensus sequence matched in SPEQ and SPEK indicating that this domain is not present. The domain PS00278 was identified in all four superantigens with only two mismatches to the consensus sequence K-x_2-(LIVF)-x_4-(LIVF)-D-x_3-R-x_2-L-x_5-(LIV)-Y in SPEQ and SPEK and no mismatches in SPER and SPEJ. The zinc binding domain (Zn) was also present in all strains. Consensus matching residues are indicated in red. * represents one of the 3-4 possible amino acids indicated in brackets on the line directly below.



Supplementary Figure 2. Intra-emm type lineages within genotypes carrying speQR. Isolates belonging to one of the four genotypes, emm18 (red), 58 (blue), 77 (green) and 94 (orange), could be negative (square) or positive (positive) for the complete speQR depending on the lineage. Fastq data for isolates was mapped to the completed emm89 genome H293 (Black square) and single nucleotide polymorphisms used to generate a maximum-likelihood tree. Scale bar represents substitutions per site. Unfilled symbols; could not confirm presence or absence of speQR from assemblies.

```
SPEQ.1* DTYITNDIRNSEDIYFPRQDKDGILDNKRLKDIYG--------------------KEIIEKTNIPINAKQ
SPEQ.2  DTYITNDIRNSEDIYFPRQDKDGILDNKRLKDIYGKDIYG---------------KEIIEKTNIPINAKQ
SPEQ.4  DTYITNDIRNSEDIYFPRQDKDGILDNKRLKDIYGKDIYGKDIYGKDIYG-----KEIIEKTNIPINAKQ
SPEQ.5  DTYITNDIRNSEDIYFPRQDKDGILDNKRLKDIYGKDIYGKDIYGKDIYGKDIYGKEIIEKTNIPINAKQ
        30                                                                    56
```

Supplementary Figure 3. **Varying number of repeats within SPEQ.** A block of 5 amino acids (bold) varied in number between alleles of SPEQ up to 5 repeats in SPEQ.5. Other alleles of SPEQ (3, 6-13) had just one block of these 5 amino acids, like SPEQ.1, but differed by other single nucleotide polymorphisms in other sites across the gene. Region shown is between 30 and 56 amino acids of the predicted mature protein (without the signal sequence). *Represents SPEQ.1, 3 and 6-13.

Supplementary Tables

Supplementary Table 1: Primers used in this study

Primer	Sequence (5'-3')[a]	Use
speQR-regionF	AAGATAGTTGGAATCAGACC	speQR confirmation and sequencing
speQR-regionR	CTGAGCAGTTTTAGATTTGG	speQR confirmation and sequencing
SPEQ-F	CCGGATCCAATGTGCTCATCAATTATTTTATATACTACTAAGG	Recombinant SPEQ expression
SPEQ-R	CCGGATCCACTAGGCCAATATTTTATTTATTCATAAAGTATATC	Recombinant SPEQ expression
SPER-F	CCGGATCCGATGTTTATTTATACATTTAAAACAG	Recombinant SPER expression
SPER-R	CCGGATCCCGTCAATTGATCATATTTATCCAAAGC	Recombinant SPER expression
speQR-F	ACCGATAAATGCCAAACAGG	speQ/R co-transcription
speQR-R	CTGTTGAAGGTGTGATTCCAC	speQ/R co-transcription
speQR-regionF	AAGATAGTTGGAATCAGACC	speQR confirmation and sequencing
speQR-regionR	CTGAGCAGTTTTAGATTTGG	speQR confirmation and sequencing

[a]BamHI restriction sites are underlined

Supplementary Table 2. Details of genome sequence data obtained, assembled and tested for speQR.

	Number assembled				
emm-type	UK	USA	Total	Tested speQR*	Confirmed positive for speQR
1	623	316	939	815	0
2	44	28	72	72	0
3	524	39	563	554	0
4	128	52	180	177	0
5	62	2	64	60	0
6	115	22	137	132	0
8	1	3	4	4	0
9	9	10	19	19	19 (100%)
11	55	50	105	104	0
12	343	134	477	469	0
15	0	1	1	1	0
18	40	2	42	41	1 (2%)
19	1	0	1	1	0
22	19	5	24	24	0
23	1	0	1	1	0
24	0	1	1	1	0
25	1	2	3	3	0
27	1	2	3	3	0
28	225	95	320	320	0
29	0	1	1	1	0
32	2	0	2	2	0
33	2	1	3	3	0
41	0	3	3	3	0
42	0	1	1	1	1 (100%)
43	2	0	2	2	0
44	34	4	38	38	0
49	7	20	27	27	0
53	1	0	1	1	1 (100%)
54	0	1	1	1	0
57	1	0	1	1	0
58	19	6	25	24	21 (88%)
59	0	53	53	53	0
60	0	5	5	5	5 (100%)
63	1	6	7	5	0
66	1	0	1	1	0
68	5	5	10	10	0
73	19	5	24	23	0
74	2	0	2	2	0
75	131	20	148	145	0
76	16	18	34	34	0

77	18	54	72	72	49 (68%)
78	1	0	1	1	0
80	1	1	2	2	0
81	26	26	52	52	0
82	10	101	111	111	0
83	0	14	14	14	0
85	1	1	2	2	0
86	1	0	1	1	0
87	60	26	86	81	81 (100%)
88	3	0	3	3	0
89	312	185	497	490	0
90	15	5	20	20	0
91	0	3	3	3	0
92	6	58	64	64	0
93	1	0	1	1	0
94	50	1	51	50	48 (96%)
95	1	0	1	1	0
99	1	0	1	1	0
100	2	0	2	2	0
101	0	5	5	5	0
102	5	1	6	6	0
103	4	1	5	5	0
104	1	2	3	3	0
106	1	2	3	3	0
108	7	2	9	9	0
109	1	0	1	1	0
110	4	0	4	4	0
111	0	1	1	1	0
112	1	0	1	1	0
113	2	1	3	3	0
114	0	2	2	2	0
118	1	25	26	26	0
123	1	0	1	1	0
134	1	0	1	1	0
151	0	3	3	3	0
165	1	1	2	2	0
168	19	0	19	19	0
169	2	0	2	2	2 (100%)
171	1	0	1	1	0
216	0	1	1	1	0
217	1	0	1	1	0
218	2	0	2	2	0
227	0	1	1	1	0
232	1	1	2	2	0
234	0	1	1	1	0
238	0	2	2	2	0

Total	3000	1440	4437	4262	228

*low coverage or poor quality assembly data prevented confirmation of speQR presence or absence for some isolates

Supplementary Table 3: Isolates within genotypes positive for speQR

Strain name	Accession number	emm-type	Reference	speQR	MLST	speQ allele	speR allele	Other superantigen genes
20155373	SRR5854238	9	[4]	Yes	75	4	3	speG, ssa, smeZ
20161091	SRR5853656	9	[4]	Yes	75	4	3	speG, ssa, smeZ
20161741	SRR5853688	9	[4]	Yes	75	4	3	speG, ssa, smeZ
20162662	SRR5853345	9	[4]	Yes	75	4	3	speG, ssa, smeZ
20163036	SRR5853687	9	[4]	Yes	75	4	3	speG, ssa, smeZ
GASEMM0302	ERR1732625	9	[5]	Yes	75	4	3	speC, speG, ssa, smeZ
GASEMM0360	ERR1734801	9	[5]	Yes – BLAST[#]	75			speG, ssa, smeZ
GASEMM0925	ERR1733536	9	[5]	Yes	75	4	3	speG, ssa, smeZ
GASEMM1010	ERR1733752	9	[5]	Yes	75	5	1	speG, ssa, smeZ
GASEMM1070	ERR1734143	9	[5]	Yes	75	2	1	speG, speL, speM, ssa, smeZ
GASEMM1136	ERR1732759	9	[5]	Yes - BLAST	75			speG, speL, speM, ssa, smeZ
GASEMM1829	ERR1732548	9	[5]	Yes	204	2	1	speG, speK, smeZ
GASEMM2004	ERR1734206	9	[5]	Yes	75	4	3	speC, speG, ssa, smeZ
20155204	SRR5853974	9.2	[4]	Yes	891	2	1	speG, smeZ
20155603	SRR5854200	9.2	[4]	Yes	891	2	1	speG, smeZ
20156713	SRR5854163	9.2	[4]	Yes	75	2	1	speC, speG, speL, speM, smeZ
20160970	SRR5853365	9.2	[4]	Yes	891	2	1	speG, speL, speM, smeZ
20162407	SRR5853438	9.2	[4]	Yes	891	2	1	speG, speL, speM, smeZ
GASEMM1718	ERR1733280	9.2	[5]	Yes	891	2	1	speG, smeZ
20162136	SRR5853544	15.1	[4]	Yes	872	1	5	speA, speC, speG, smeZ
GASEMM0285	ERR1735410	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ

GASEMM0540	ERR1733461	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM0561	ERR1734827	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM0871	ERR1733017	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1160	ERR1735235	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1247	ERR1733168	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1418	ERR1734972	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1598	ERR1732549	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1656	ERR1735241	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1720	ERR1732620	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1766	ERR1734945	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1906	ERR1734065	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1910	ERR1734348	18	[5]	No	41*	-	-	speC, speG, speL, speM, smeZ
GASEMM1967	ERR1734605	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2059	ERR1733917	18	[5]	No	41*	-	-	speC, speG, speL, speM, smeZ
GASEMM2067	ERR1732772	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2130	ERR1733147	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2190	ERR1735052	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2229	ERR1733524	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2230	ERR1733448	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2651	ERR1734606	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2655	ERR1732562	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2713	ERR1732983	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2729	ERR1734538	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2867	ERR1734893	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2895	ERR1733782	18	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM3031	ERR1734236	18	[5]	No	41*	-	-	speC, speG, speL, speM, smeZ
20154046	SRR5858678	18.12	[4]	Yes	535	1	5	speA, speG, speK, smeZ

GASEMM1747	ERR1735058	18.21	[5]	No	402	-	-	speA, speG, smeZ
GASEMM1752	ERR1733080	18.21	[5]	No	402	-	-	speA, speG, smeZ
GASEMM0528	ERR1733900	18.29	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM0573	ERR1732511	18.29	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM0802	ERR1734573	18.29	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM1288	ERR1734851	18.29	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2071	ERR1733665	18.29	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2122	ERR1734499	18.29	[5]	No	41*	-	-	speC, speG, speL, speM, smeZ
GASEMM2087	ERR1732492	18.38	[5]	No	42	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2351	ERR1733901	18.39	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2453	ERR1732505	18.39	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2693	ERR1734737	18.39	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
GASEMM2376	ERR1732816	18.4	[5]	No	41*	-	-	speA, speC, speG, speL, speM, smeZ
20156711	SRR5854193	18.7	[4]	No	42*	-	-	speC, speG, speL, speM, ssa, smeZ
20161097	SRR5853770	42	[4]	Yes	80	6	1	speG, speH, speI, smeZ
GASEMM0926	ERR1733963	53	[5]	Yes	363	13	7	speG, speH, speI, speM, smeZ
20152179	SRR5858386	58	[4]	Yes	176	1	1	speG, ssa, smeZ
20155380	SRR5853860	58	[4]	Yes	176	1	1	speG, ssa, smeZ
20156103	SRR5853785	58	[4]	Yes	176	1	1	speG, ssa, smeZ
20156726	SRR5854259	58	[4]	Yes	176	1	1	speG, ssa, smeZ
GASAR0061	ERS361761	58	[6]	Yes	176	3	1	speG, speH, speI, ssa, smeZ
GASEMM0384	ERR1733815	58	[5]	Yes	176	3	1	speG, ssa, smeZ
GASEMM0530	ERR1735030	58	[5]	Yes	176	3	1	speG, ssa, smeZ
GASEMM0679	ERR1735323	58	[5]	Yes	176	3	1	speG, ssa, smeZ
GASEMM0880	ERR1734067	58	[5]	Yes	176	3	1	speG, ssa, smeZ
GASEMM0900	ERR1735007	58	[5]	Yes	176	3	1	speG, ssa, smeZ
GASEMM1156	ERR1732612	58	[5]	Yes- BLAST	176			speG, ssa, smeZ

GASEMM1511	ERR1732902	58	[5]	Yes	176*		3		1	speG, smeZ
GASEMM1624	ERR1733931	58	[5]	Yes	176		7		1	speG, ssa, smeZ
GASEMM1915	ERR1733025	58	[5]	Yes	176		3		1	speG, speH, smeZ
GASEMM1985	ERR1733185	58	[5]	Yes- BLAST	176*					speG, ssa, smeZ
GASEMM2416	ERR1732690	58	[5]	Yes	176		3		1	speG, ssa, smeZ
GASEMM2441	ERR1734784	58	[5]	Yes	176		3		1	speG, ssa, smeZ
GASEMM3021	ERR1733289	58	[5]	Yes	176		3		1	speC, speG, speH, smeZ
20154011	SRR5858614	58.2	[4]	Yes	176		1		1	speC, speG, speH, smeZ
20156359	SRR5854013	58.2	[4]	Yes	176		10		1	speC, speG, speH, smeZ
GASEMM1431	ERR1734695	58.7	[5]	Yes	176		3		1	speC, speG, smeZ
GASEMM1086	ERR1735087	58.8	[5]	No	985	-		-		speC, speG, smeZ
GASEMM1108	ERR1735468	58.8	[5]	No	985	-		-		speC, speG, smeZ
GASEMM1933	ERR1734866	58.8	[5]	No	549	-		-		speG, speK, smeZ
20154028	SRR5858394	60.2	[4]	Yes	53		1		1	None
20161832	SRR5853686	60.2	[4]	Yes	53		1		1	speG, smeZ
20162105	SRR5853728	60.2	[4]	Yes	53		1		1	None
20162139	SRR5853537	60.2	[4]	Yes	53		1		1	None
20162155	SRR5853329	60.2	[4]	Yes	53		1		1	None
20152512	SRR5858716	77	[4]	Yes	63		8		1	speA, smeZ
20153541	SRR5858459	77	[4]	No	399	-		-		speG, smeZ
20153553	SRR5858462	77	[4]	No	399	-		-		speG, smeZ
20154045	SRR5858677	77	[4]	Yes	63		1		1	speL, speM, smeZ
20154151	SRR5858630	77	[4]	Yes	63		1		1	smeZ
20154558	SRR5858309	77	[4]	Yes	63		1		1	speK, smeZ
20154612	SRR5858670	77	[4]	No	399	-		-		speG, smeZ
20154792	SRR5858505	77	[4]	Yes	63		1		1	None
20155018	SRR5858422	77	[4]	Yes	63		1		1	speC, smeZ

20155030	SRR5858695	77	[4]	Yes	63		1		1	speC, smeZ
20155035	SRR5858539	77	[4]	Yes	63		1		1	smeZ
20155065	SRR5853826	77	[4]	Yes	63		1		1	speC, smeZ
20155357	SRR5854098	77	[4]	Yes	63		1		1	smeZ
20155578	SRR5853970	77	[4]	Yes	63		1		1	speL, speM, smeZ
20155640	SRR5854282	77	[4]	Yes	63		1		1	speC, smeZ
20155641	SRR5854281	77	[4]	Yes	63		1		1	speC, smeZ
20155652	SRR5854144	77	[4]	Yes	63		9		1	speL, speM, smeZ
20156012	SRR5854073	77	[4]	No	399	-		-		speG, smeZ
20156017	SRR5853922	77	[4]	Yes	63		1		1	speK, smeZ
20156106	SRR5853788	77	[4]	Yes	63		1		1	speK, smeZ
20156164	SRR5854197	77	[4]	Yes	63		1		1	smeZ
20156171	SRR5854064	77	[4]	Yes	63		1		1	smeZ
20156324	SRR5854015	77	[4]	No	399	-		-		speG, smeZ
20156364	SRR5854021	77	[4]	No	399	-		-		speC, speG, speH, speI, smeZ
20156388	SRR5854152	77	[4]	Yes	63		1		1	smeZ
20156402	SRR5853904	77	[4]	Yes	63		1		1	speK, smeZ
20156442	SRR5853843	77	[4]	No	399	-		-		speG, speK, speL, smeZ
20156624	SRR5854091	77	[4]	Yes	63		1		1	speK, smeZ
20156633	SRR5853933	77	[4]	No	399	-		-		speG, speH, speI, smeZ
20156792	SRR5854166	77	[4]	Yes	63		1		1	speL, speM, smeZ
20160025	SRR5854032	77	[4]	Yes	63		1		6	speK, smeZ
20160171	SRR5853946	77	[4]	Yes	63		1		1	speK, smeZ
20160245	SRR5854087	77	[4]	Yes	63		1		1	speC, smeZ
20160322	SRR5853916	77	[4]	Yes	63		1		1	speL, speM, smeZ
20160750	SRR5853422	77	[4]	Yes	63		1		1	speL, speM, smeZ
20160953	SRR5853589	77	[4]	No	904	-		-		speG, smeZ

20161075	SRR5853432	77	[4]	No	399	-		-	speG, smeZ
20161076	SRR5853433	77	[4]	Yes	63		1	1	speC, smeZ
20161082	SRR5853552	77	[4]	Yes	63		1	1	speC, smeZ
20161205	SRR5853474	77	[4]	Yes	63		1	1	speL, speM, smeZ
20161238	SRR5853603	77	[4]	No	399	-		-	speG, smeZ
20161721	SRR5853401	77	[4]	Yes	63		1	1	speL, speM, smeZ
20161722	SRR5853402	77	[4]	No	399	-		-	speG, speJ, speL, speM, smeZ
20161724	SRR5853404	77	[4]	Yes	63		1	1	speL, speM, smeZ
20161838	SRR5853507	77	[4]	Yes	63		1	1	speL, speM, smeZ
20162121	SRR5853652	77	[4]	Yes	63		1	1	speL, speM, smeZ
20162428	SRR5853748	77	[4]	Yes	63		1	1	speC, smeZ
GASEMM0134	ERR1732814	77	[5]	Yes	63		1	1	speC, smeZ
GASEMM0135	ERR1733196	77	[5]	Yes	63		1	1	speC, smeZ
GASEMM0352	ERR1734572	77	[5]	Yes	63		1	1	speC, smeZ
GASEMM0846	ERR1734088	77	[5]	No	399	-		-	speG, speL, speM, smeZ
GASEMM1061	ERR1733365	77	[5]	Yes	63		1	1	smeZ
GASEMM1453	ERR1735072	77	[5]	No	399	-		-	speC, speG, speL, speM, smeZ
GASEMM1494	ERR1733572	77	[5]	Yes	63		1	1	speC, smeZ
GASEMM1700	ERR1734961	77	[5]	Yes	63		1	1	smeZ
GASEMM1729	ERR1734104	77	[5]	No	399	-		-	speG, speL, speM, smeZ
GASEMM1761	ERR1734232	77	[5]	Yes	63		1	1	smeZ
GASEMM1833	ERR1735076	77	[5]	Yes	63		1	1	speC, smeZ
GASEMM2373	ERR1734646	77	[5]	Yes	63		1	1	speL, smeM, smeZ
GASEMM2748	ERR1733450	77	[5]	No	399	-		-	speG, speL, speM, smeZ
GASEMM2849	ERR1735297	77	[5]	Yes	63*		1	1	smeZ
GASEMM2882	ERR1734850	77	[5]	Yes	63		1	1	speL, speM, smeZ
GASEMM2883	ERR1732745	77	[5]	Yes	63		1	1	speL, speM, smeZ

GASEMM2955	ERR1734264	77	[5]	Yes	63	1	1	speC, smeZ
GASEMM3037	ERR1733626	77	[5]	Yes	63	1	1	speC, smeZ
20155579	SRR5853969	77.4	[4]	No	133	-	-	speG, speM, smeZ
20155580	SRR5853964	77.4	[4]	No	133	-	-	speG, speM, smeZ
20155585	SRR5853820	77.4	[4]	No	133	-	-	speG, speM, smeZ
20161863	SRR5853342	77.4	[4]	No	133	-	-	speG, speM, smeZ
20161864	SRR5853343	77.4	[4]	No	133	-	-	speG, speM, smeZ
20161872	SRR5853623	77.4	[4]	No	133	-	-	speG, speM, smeZ
20165973	SRR5853525	77.4	[4]	No	133	-	-	speG, speM, smeZ
20152749	SRR5858550	87	[4]	Yes	62	1	2	speG, speJ, ssa, smeZ
20153605	SRR5858351	87	[4]	Yes	62	1	2	speG, speJ, ssa, smeZ
20154292	SRR5858368	87	[4]	Yes	62	12	2	speC, speG, speJ, ssa, smeZ
20155202	SRR5853973	87	[4]	Yes	62	1	2	speC, speG, ssa, smeZ
20155269	SRR5853953	87	[4]	Yes	62	12	2	speC, speG, speJ, ssa, smeZ
20155393	SRR5854175	87	[4]	Yes	62	1	2	speG, speJ, ssa, smeZ
20155602	SRR5854241	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20155658	SRR5854063	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20156112	SRR5853794	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20156396	SRR5854160	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20156412	SRR5854044	87	[4]	Yes	62	1	2	speG, speJ, ssa, smeZ
20156782	SRR5854122	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20160141	SRR5853847	87	[4]	Yes	62	12	2	speC, speG, speJ, ssa, smeZ
20160495	SRR5853645	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20160499	SRR5853530	87	[4]	Yes	62	12	2	speC, speG, speJ, ssa, smeZ
20160747	SRR5853526	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20160975	SRR5853469	87	[4]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20160986	SRR5853485	87	[4]	Yes	62	1	2	speC, speG, speJ, speK, ssa, smeZ

20161088	SRR5853698	87	[4]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
20161436	SRR5853392	87	[4]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
20162409	SRR5853434	87	[4]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
20162410	SRR5853435	87	[4]	Yes	890		1	2	speC, speG, speJ, ssa, smeZ
20162420	SRR5853559	87	[4]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
20162426	SRR5853750	87	[4]	Yes	62		1	2	speG, speJ, ssa, smeZ
20162648	SRR5853630	87	[4]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASAR0007	ERS361812	87	[6]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0188	ERR1733872	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0189	ERR1733138	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0234	ERR1734059	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0264	ERR1733408	87	[5]	Yes	62		1	2	speG, speJ, ssa, smeZ
GASEMM0375	ERR1733437	87	[5]	Yes	62		1	2	speC, speG, speH, speI, speJ, ssa, smeZ
GASEMM0391	ERR1733949	87	[5]	Yes	62	Cannot confirm allele	Cannot confirm allele		speG, speJ, ssa, smeZ
GASEMM0406	ERR1733763	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0416	ERR1733324	87	[5]	Yes	62		1	2	speG, speJ, ssa, smeZ
GASEMM0422	ERR1734129	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0460	ERR1733077	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0470	ERR1734760	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0558	ERR1735476	87	[5]	Yes	62		1	2	speG, speJ, ssa, smeZ
GASEMM0563	ERR1734436	87	[5]	Yes	62		1	2	speC, speG, ssa, smeZ
GASEMM0639	ERR1734864	87	[5]	Yes	62	Cannot confirm allele	Cannot confirm allele		speC, speG, speJ, ssa, smeZ
GASEMM0696	ERR1734006	87	[5]	Yes	62		1	2	speC, speG, speJ, ssa, smeZ
GASEMM0736	ERR1732833	87	[5]	Yes	62		1	2	speG, speJ, ssa, smeZ

GASEMM0781	ERR1733581	87	[5]	Yes	62	Cannot confirm allele		Cannot confirm allele		speC, speG, speJ, ssa, smeZ
GASEMM0796	ERR1735275	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM0799	ERR1732942	87	[5]	Yes	62	Cannot confirm allele		Cannot confirm allele		speC, speG, speJ, ssa, smeZ
GASEMM0800	ERR1734904	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM0920	ERR1735085	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM0968	ERR1734309	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa smeZ
GASEMM0972	ERR1732701	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM0978	ERR1733699	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1074	ERR1734225	87	[5]	Yes	62		1		2	speC, speG, speJ, speK, ssa, smeZ
GASEMM1215	ERR1734719	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1258	ERR1733634	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1295	ERR1732685	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1313	ERR1733067	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1315	ERR1733719	87	[5]	Yes	62		1		2	speG, speJ, ssa, smeZ
GASEMM1349	ERR1735221	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1365	ERR1732481	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1466	ERR1732652	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1493	ERR1733772	87	[5]	Yes	62		1		2	speG, speJ, ssa, smeZ
GASEMM1512	ERR1733251	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1572	ERR1734273	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1573	ERR1734975	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1575	ERR1733810	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1577	ERR1733532	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1601	ERR1735368	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ
GASEMM1688	ERR1735057	87	[5]	Yes	62		1		2	speC, speG, speJ, ssa, smeZ

GASEMM1705	ERR1735155	87	[5]	Yes	62	1	2	speC, speG, ssa, smeZ
GASEMM1813	ERR1735180	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM1959	ERR1733991	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2010	ERR1732466	87	[5]	Yes	62	1	2	speC, speG, speJ, speK, ssa, smeZ
GASEMM2377	ERR1735186	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2378	ERR1734363	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2379	ERR1733864	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2551	ERR1733014	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2568	ERR1732810	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2846	ERR1734886	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2942	ERR1732604	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM3012	ERR1734744	87	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
GASEMM2424	ERR1734959	87.1	[5]	Yes	62	1	2	speC, speG, speJ, ssa, smeZ
20162638	SRR5853498	87.3	[4]	Yes	62	11	2	speC, speG, ssa, smeZ
20152838	SRR5858329	94	[4]	No	89	-	-	speG, speH, smeZ
GASEMM0336	ERR1735266	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM0542	ERR1733387	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM0725	ERR1735331	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM0801	ERR1735414	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM0818	ERR1734261	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1157	ERR1734377	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1158	ERR1734859	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1343	ERR1733447	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1360	ERR1733284	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1374	ERR1732869	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1379	ERR1732790	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1438	ERR1734494	94	[5]	Yes	89	1	2	speG, speH, smeZ

GASEMM1522	ERR1734834	94	[5]	Yes	89	1	2	speG, speH, speL, speM, smeZ
GASEMM1553	ERR1733938	94	[5]	No	89	-	-	speG, speH, smeZ
GASEMM1608	ERR1734955	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1612	ERR1735069	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1643	ERR1735361	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1692	ERR1735041	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1711	ERR1734820	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1727	ERR1734329	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1789	ERR1734110	94	[5]	Yes- BLAST	89			speG, speH, smeZ
GASEMM1832	ERR1734589	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1864	ERR1732589	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1874	ERR1733144	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1885	ERR1733359	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1900	ERR1734193	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1911	ERR1733866	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1969	ERR1734399	94	[5]	Yes	89	1	2	speC, speG, speH, smeZ
GASEMM2062	ERR1733775	94	[5]	Yes	89	1	2	speC, speG, speH, smeZ
GASEMM2124	ERR1734669	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2148	ERR1733192	94	[5]	Yes- BLAST	89			speG, speH, smeZ
GASEMM2192	ERR1732991	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2195	ERR1734642	94	[5]	Yes - BLAST	89			speG, speH, smeZ
GASEMM2212	ERR1732773	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2301	ERR1734922	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2363	ERR1733377	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2467	ERR1733475	94	[5]	Yes - BLAST	89			speG, speH, smeZ
GASEMM2535	ERR1733579	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2725	ERR1733407	94	[5]	Yes	89	1	2	speG, speH, smeZ

GASEMM2800	ERR1735255	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2847	ERR1733164	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2866	ERR1733363	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2868	ERR1732521	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2871	ERR1732873	94	[5]	Yes - BLAST	89			speG, speH, smeZ
GASEMM2890	ERR1733409	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2936	ERR1734650	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2937	ERR1735320	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM2964	ERR1732797	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM3042	ERR1734749	94	[5]	Yes	89	1	2	speG, speH, smeZ
GASEMM1880	ERR1733790	169.3	[5]	Yes	53	6	4	None
GASEMM2308	ERR1733632	169.3	[5]	Yes	53	6	4	None

BLAST; presence of the gene could be confirmed by BLAST but not from the de novo assembly therefore an allele was not assigned.

* MLST closest match