ibc Interdisciplinary Bio Central

IBC 2012;4:11, 1-8 • DOI: 10.4051/ibc.2012.4.4.0011

Open Access

RAPID REPORT

Prediction of an Essential Gene with Potential Drug Target Property in *Streptococcus suis* Using Comparative Genomics

Aubhishek Zaman*

Department of Genetic engineering and Biotechnology, University of Dhaka, Dhaka, Bangladesh

Subject areas; Bioinformatics/Computational biology/Molecular modeling, Omics (Physiomics/metabolomics/proteomics/genomics)

***Correspondence** and requests for materials should be addressed to A.Z. (aubhishek@gmail.com).

Editor; Hong Gil Nam, DGIST, Korea

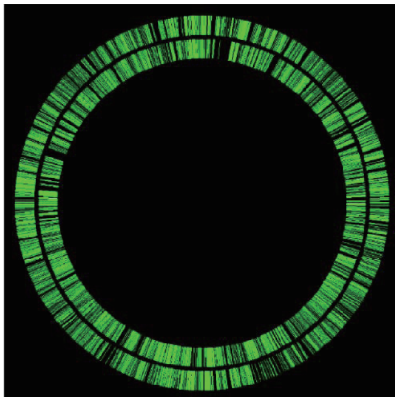
Received September 21, 2012
Accepted October 08, 2012
Published October 10, 2012

Citation; Zaman, A. Prediction of an Essential Gene with Potential Drug Target Property in *Streptococcus suis* using comparative genomics. IBC 2012, 4:11, 1-8.
doi: 10.4051/ibc.2012.4.4.0011

Competing interest; All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Genes that are indispensable for survival are referred to as essential gene. Due to the momentous significance of these genes for cellular activity they can be selected potentially as drug targets. Here in this study, an essential gene for *Streptococcus suis* was predicted using coherent statistical analysis and powerful genome comparison computational method. At first the whole genome protein scatter plot was generated and subsequently, on the basis of statistical significance, a reference genome was chosen. The parameters set forth for selecting the reference genome was that the genome of the query (*Streptococcus suis*) and subject must fall in the same genus and yet they must vary to a good degree. *Streptococcus pneumoniae* was found to be suitable as the reference genome. A whole genome comparison was performed for the reference (*Streptococcus pneumoniae*) and the query genome (*Streptococcus suis*) and 14 conserved proteins from them were subjected to a screen for potential essential gene property. Among those 14 only one essential gene was found to be with impressive similarity score between reference and query. The essential gene encodes for a type of 'Clp protease'. Clp proteases play major roles in degrading misfolded proteins. Results found here should help formulating a drug against *Strptococcus suis* which is responsible for mild to severe clinical conditions in human. However, like many other computational studies, the study has to be validated furthermore through in vitro assays for concrete proof.



Key Words: *Strptococcus suis*; *Streptococcus pneumonia*; essential gene; Clp protease; reference genome

INTRODUCTION

Essential genes are genes that are indispensable to support cellular life. A minimal gene set that is required for a living cell falls under this group of genes. Identifying essential genes in bacteria supports to identify potential drug targets and an understanding of minimal requirements for a synthetic cell. However, experimentally assaying the essentiality of their coding genes is resource intensive and not feasible for all bacterial organisms, in particular if they are infective.

Streptococcus suis is a peanut-shaped, Gram-positive bacterium[1]. *S. suis* is a notorious pathogen for pigs. It is a chief cause of zoonotic disease and transmission from pigs to humans has also been reported[2]. Human infection may be of varied intensity- mild to severe. Severe clinical conditions range from meningitis, septicaemial shock, cardiac inflammation and deafness[3-10]. However, although not impossible, *S. suis* infection usually does not turn lethal in humans. Humans get infected with *S. suis* when they handle infected pig carcasses or meat, especially with exposed cuts and abrasions on their hands. The whole genome of the bacteria *Streptococcus suis* has been sequenced by Wellcome Trust Sanger Institute. They have sequenced the whole genome sequences of three *Streptococcus suis* strains from the same lineage: one from European pigs, and two from human cases from China and Vietnam. The strains were: P1/7, BM407 and SC84. Although pathogenicity of the bacteria is well reported, to date, the mechanism of pathogenesis is not fully understood. In this regard studying the genome information of the bacteria would be of great importance. The recent availability of bacterial genome sequence information permits the identification of conserved genes that are potential targets for novel antibiotic drug discovery. Database of Essential Genes (DEG) is a database that enlists essential genes of different prokaryotes and eukaryotes. The functions encoded by essential genes are considered a foundation of life and therefore are likely to be common to all cells. Analysis of essential genes could help to answer the question of what are the core set of genes necessary to support cellular life. DEG is freely accessible from the website http://tubic.tju.edu.cn/deg/[11-13]. One challenge presented by large-scale genome sequencing efforts is effective display of uniform information to the scientific community. The Comprehensive Microbial Resource (CMR) was organized with a view to face this challenge with meaningful sharing of sequence data amongst scholars and scientists. CMR contains robust annotation of all complete microbial genomes and allows for a wide variety of data retrievals. The bacterial information has been placed on the Web at http://cmr.jcvi.org/tigr-scripts/CMR/CmrHomePage.cgi for retrieval using standard web browsing technology. Retrievals can be based on protein properties such as molecular weight or hydrophobicity, GC-content, functional role assignments and taxonomy[14].

In this study, an essential gene has been predicted for the bacteria using comparative genomic techniques. The found results are novel for the given bacterial species and can be of considerable importance to design drugs against potential drug targets.

RESULTS

Reference genome selection

To deduce an optimum reference genome a scatter protein plot was performed between potential reference genomes- genomes of *Streptococcus* genus deposited in the CMR database. From the total protein scatter diagram the image was captured and digitized using Engauge Digitizer tool. The digitized image was then converted into a regression curve that goes through most of the diagonally scattered points. This generated a liner regression curve. The regression curve was generated for 3 pairs of genomes;

- *Streptococcus suis* vs *Streptococcus mutans*: the most liner graph indicating best homology.
- *Streptococcus suis* vs *Streptococcus pyogens*: liner graph with mediocre homology.
- *Streptococcus suis* vs *Streptococcus pseudomonas*: liner graph with greatest divergence.

According to the principle mentioned afore, *Streptococcus pseudomonas* was chosen as the reference genome. The analysis is described in detail in the following section for perspicuity purpose (Figure 1). The black patches that were found after the scatter plot were linked to each other to form a straight line. Regression Analysis for the scatter plot between *S. mutans* and *S. suis* was conducted by naming the Y values of the given scatter plot was termed as Curve1 in the regression analysis graph whereas the X values were termed simply as X. For the given notations the regression equation is:

$$\text{Curve1} = -63.6 + 1.06 \times$$

To check whether the regression line was good or bad from a statistical point of view, *P*-value was measured. The coefficient value for P (*P*-value) tells you whether or not the association between the response and predictor(s) is statistically significant. P below the α-value (level of significance, take on a 5% confidence level) means that they are significant (Table 1). S, R and adjusted R are measures of how well the model fits the data. These values can help one select the model with the best fit. S is measured in the units of the response variable and represents

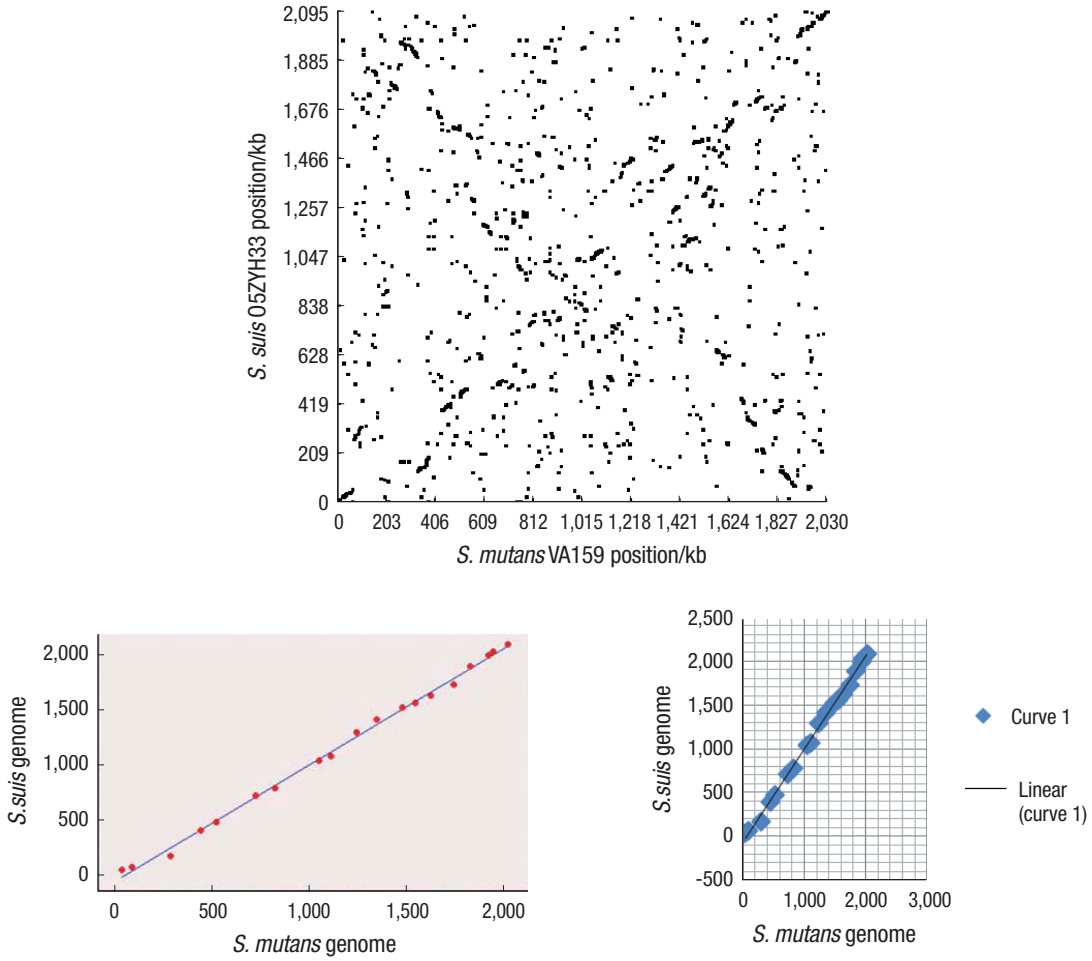


Figure 1. The Scatter plot generated for the whole proteome of *S. mutans* (X axis) and *S. suis* (Y axis). A regression curve was drawn subsequently joining the dark patches in the scatter diagram.

Table 1. Standard Error (SE) coefficient and *P* value for generated regression curve between *S. mutans* and *S. suis*

Predictor	Coef	SE Coef	T	*P*
Constant	-63.59	17.94	-3.55	0.002
x	1.05700	0.01368	77.27	0.000

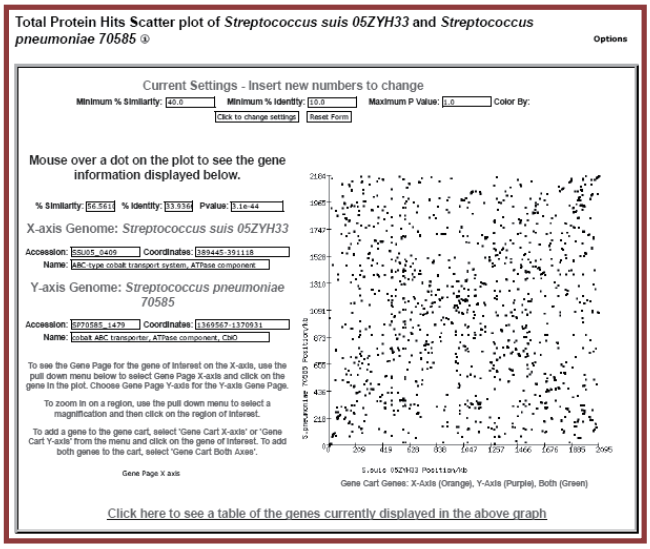
Table 2. Analysis of varience for the generated regression curve between *S. mutans* and *S. suis*

Source	DF	SS	MS	F	*P*
Regression	1	8,541,847	8,541,847	5,971.13	0.000
Residual Error	17	24,319	1,431	-	-
Total	18	566,166	-	-	-

the standard distance data values fall from the regression line. For a given study, the better the equation predicts the response, the lower the value of S is. R (R-Sq) describes the amount of variation in the observed response values that is explained by the predictor(s). R always increases with additional predictors. For example, the best five-predictor model will always have a higher R than the best four-predictor model. Therefore, R is most useful when comparing models of the same size. Adjusted R is a modified R that has been adjusted for the number of terms in the model. If one includes unnecessary terms, R can be artificially high. Unlike R, adjusted R may get smaller when terms to the model are added. Following are the values for the scattered plot illustrated above;

S = 37.8223 R-Sq = 99.7% R-Sq(adj) = 99.7%

Analysis of Variance was determined by statistical package Minitab-15 (Table 2). Similarly the scatter plot for *S. pneumoniae* and *S. suis* was generated and similar statistical analysis was performed (Figure 2). Regression Analysis for the scatter plot between *S. pneumoniae* and *S. suis* was conducted by naming the Y values of the given scatter plot was termed as Curve 2 in the regression analysis graph whereas the X values were termed simply as X (Table 3). For the given notations the regression equation is:



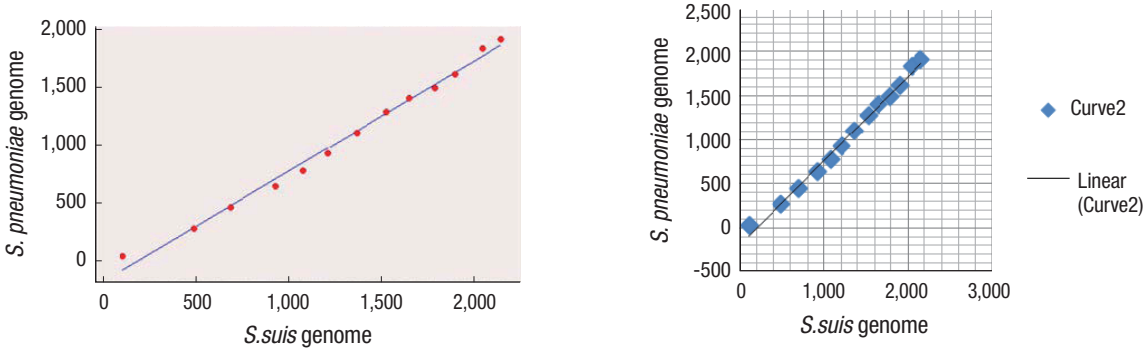


Figure 2. The Scatter plot generated for the whole proteome of *S. pneumoniae* (Y axis) and *S. suis* (X axis). A regression curve was drawn subsequently joining the dark patches in the scatter diagram.

Table 3. Standard Error (SE) coefficient and *P* value for generated regression curve between *S. pneumoniae* and *S. suis*

Predictor	Coef	SE Coef	T	*P*
Constant	-186.50	36.41	-5.12	0.000
x	0.95324	0.02535	37.60	0.000

S = 55.0294, R-Sq = 99.2%, R-Sq(adj) = 99.2%.

Table 4. Analysis of varience for the generated regression curve between *S. pneumoniae* and *S. suis*

Source	DF	SS	MS	F	*P*
Regression	1	4,281,464	4,281,464	1,413.85	0.000
Residual Error	11	33,311	3,028	-	-
Total	12	4,314,775	-	-	-

$$\text{Curve 2} = -186 + 0.953 \times$$

Analysis of Variance was determined by statistical package Minitab-15 (Table 4). The S value found here was higher for *Streptococcus suis* vs *Streptococcus pneumoniae* scatter plotthan that of the S value of the *Streptococcus suis* vs *Streptococcus mutans* scatter plot. As the lower the S value the higher the regression fit tendency of the scatter plot, it can be interpreted that *S. mutans* shares a more similar genome sequence with *S. pneumoniae*. As a more distant genome was looked for to screen out too many proteins hits while comparing the proteome, *S. pneumoniae* was chosen in this regard.

Comparative genome analysis using multigenome comparion algorithm

The genome information of the *Streptococcus suis* (*strain: O5ZYH33*) was compared with the *Streptococccus pneumoniae* using multi genome homology comparison algorithm. The algorithm sets a user defined reference DNA molecule and compares this with a query molecule present in the database (Figure 3). Based on the homology the locus/proteins that show good similarity to each other were retrieved. The cut off value set for the homology was set to an absolute 100% similarity score. 14 sequences were selected on the basis of the results (Table 5).

Search for essential genes within the highly similar proteins

The proteins that showed high similarity in *S. pneumoniae* were subjected to a screening for their inevitability for cellular processes. This was done by searching essential database gene against the query proteins- proteins selected earlier from *S. pneumoniae* based on the similarity scores. 'Clp protease was found to be an essential protein (Figure 4A). This result indicates that Clp protease is an essential gene for the bacteria and quite specific for the *S. suis*. Clp protease was again searched in the scatter plot and is marked by red in the (Figure 4B). From this figure we can see that the Clp protease structure fits well into the regression curve.

Figure 3. Whole genome comparison map for *S. pneumoniae* and *S. suis*. Total 2,167 proteins were found to match to some extent to the reference genome (*S. pneumoniae*) among which only protein that showed 100% similarity score were chosen.

DISCUSSION

The results found in this study was based on a careful reference genome selection for *S. suis* and a subsequent genome-wide proteome search for a drug target- an essential gene product without which the bacteria cannot survive. Given the economic and clinical harm *S. suis* inflicts upon, the study can have far reaching impacts; especially to design a drug targeting the Clp protease found as a potential drug target from this study.

CLP protease family is a family of serine peptidases belong to the MEROPS peptidase family S14 (ClpP endopeptidase family, clan SK)[15]. ClpP is an ATP-dependent protease that cleaves a number of proteins, such as casein and albumin. Clp proteases all have a characteristic Clp protease domain associated with it. They cleave peptides in various proteins in a process that requires ATP hydrolysis and have a chymotrypsin-like activity[16]. They hold major roles in the degradation of misfolded proteins and may also have roles of a master protease which is attracted to different substrates by different specificity factors such as

Table 5. Whole genome comparison map for *S. pneumoniae* and *S. suis* in a tabular form

Reference Locus	Reference Common Name	Comparison Locus	Comparison Common Name	Percent Similarity
SSU05_1551	putative ATP-dependent Clp protease, proteolytic subunit	SP70585_0792	Clp protease	100.0
SSU05_0080	Ribosomal protein S17	SP70585_0274	ribosomal protein S17	100.0
SSU05_0276	50S ribosomal protein L33	SP70585_2261	ribosomal protein L33	100.0
SSU05_0093	Ribosomal protein S13	SP70585_0289	ribosomal protein S13p/S18e	100.0
SSU05_0368	Predicted RNA-binding protein containing KH domain, possibly ribosomal protein	SP70585_1787	conserved hypothetical protein	100.0
SSU05_0096	Ribosomal protein L17	SP70585_0292	ribosomal protein L17	100.0
SSU05_0075	SSU ribosomal protein S19P	SP70585_0269	ribosomal protein S19	100.0
SSU05_1734	6-phosphogluconate dehydrogenase	SP70585_0446	6-phosphogluconate dehydrogenase, decarboxylating	100.0
SSU05_1161	Transcriptional regulator	SP70585_2240	transcriptional regulator, GntR family	100.0
SSU05_0083	Ribosomal protein L5	SP70585_0277	50S ribosomal protein L5 (BL6)	100.0
SSU05_0340	Ribosomal protein L28	SP70585_0511	ribosomal protein L28	100.0
SSU05_2046	Tetrahydrodipicolinate N-succinyltransferase	SP70585_2203	galactoside O-acetyltransferase	100.0
SSU05_1089	Ribosomal protein S20	SP70585_0874	ribosomal protein S20	100.0
SSU05_1165	sugar transporter, putative	SP70585_2243	sugar/sodium symporter	100.0

Total 2,167 proteins were found to match to some extent to the reference genome (*S. pneumoniae*) among which only protein that showed 100% similarity score were chosen. 14 such proteins were found and among them clp protease was later found to be an essential gene.

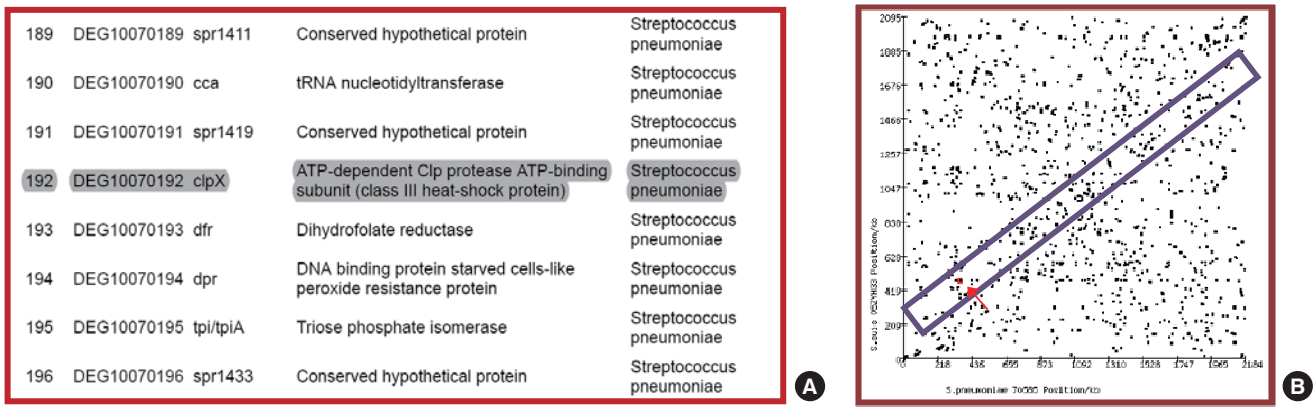


Figure 4. (A) 14 proteins that were selected on the basis of whole genome homology map were searched exhaustively for their inevitability for the bacterial life-cycle. 'Clp protease' (*S. pneumoniae* Essential protein number 192 according to Database of Essential Gene 5.0) that showed a 100% similarity score was also found to be an essential gene. (B) 'Clp protease' is marked by red in the scatter plot generated for the whole proteome of *S. pneumoniae* (X axis) and S.suis (Y axis).

ClpA or ClpX[17]. Hydrolysis of proteins to small peptides in the presence of ATP and magnesium[18-21]. Alpha-casein is the usual test substrate. In the absence of ATP, only oligopeptides shorter than five residues are hydrolyzed (such as succinyl-Leu-Tyr-|-NHMec; and Leu-Tyr-Leu-|-Tyr-Trp, in which cleavage of the -Tyr-|-Leu- and -Tyr-|-Trp bonds also occurs)[16]. ATP-dependent Clp protease proteolytic subunit is an enzyme that in humans is encoded by the CLPP gene. It is found in mitochondria and is widely distributed in bacterial species[22]. The protein encoded by this gene belongs to the peptidase family S14 and hydrolyzes proteins into small peptides in the presence of ATP and magnesium[23,24].

In *Bacillus subtilis* regulation of the stress-inducible clpE, clpC, and clpP genes was found to be dependent predominantly on the transcriptional repressor, CtsR, encoded by the first gene of the clpC operon[25,26]. Under stress conditions CtsR becomes a target for the ClpCP protease. ATP-dependent proteolysis mediated by Clp proteases is not only important for B. subtilis during stress but also for general stationary-phase phenomena, such as exoenzyme synthesis, motility, competence development and sporulation[27,28].

Clp protease deficient mutants represent a new global approach in the proteomic era and offer a big chance to gain insights into post-translational regulation by these proteolytic systems[29-31]. Clp protease is an essential gene for the bacteria for the *S. pneumoniae*[32-38]. By using powerful genome comparison tools the study succeeds to predict Clp protease as a potential drug target in *S. suis*- which had significant similarity with S. pneumoniae in genomic content.

MATERIALS AND METHODS

Reference genome selection

The concept of reference genome is an important one in comparative genomics[39,40]. A reference genome is a genome with which the genome information of the query genome is compared. Although the whole genome sequence of *Streptococcus suis* is available in public databases, a comprehensive annotation of the genome is lacking in many aspects; many of the proteins in the bacteria are yet to be annotated beyond doubt. However the sequence information allowed us to compare it with the reference genome and thus generate a hypothetical annotation on the basis of the comparison.

While choosing the reference genome it was kept in mind that the genome must not be too variable in sequence- so that there can be significant overlapping between the query and reference. However, the genomes must also not be too similar either; otherwise this will give away no new information at all.

To ensure enough homology between two genomes, as a rule of thumb, the organisms were chosen from the same genus-*Streptococcus*. This choice meant that the range for the most dissimilar sequence has been set- the reference genome sequences cannot be more dissimilar than the genus. However the species name must vary and the variation should be higher to ensure that even within a same genus, within quite conserved genomic sequences, there is at least some sort of dissimilarities between them. This in result would ease the job of distinguishing some potential essential gene targets (usually highly conserved) from background of less conserved proteins.

To deduce an optimum reference genome a scatter protein plot was performed between potential reference genomes- genomes of *Streptococcus genus* deposited in the CMR database. From the total protein scatter diagram the image was captured and digitized using Engauge Digitizer tool. The digitized image was then converted into a regression curve that goes through most of the scattered points. This generated a liner regression curve. The regression curve was generated for 3 pairs of genomes- 1) *Streptococcus suis* vs *Streptococcus mutans* 2) *Streptococcus suis* vs *Streptococcus pyogens,* and 3) *Streptococcus suis* vs *Streptococcus pneumoniae*.

Comparative genome analysis using multigenome comparison algorithm

The genome information of the *Streptococcus suis* (*strain: O5ZYH33*) is publicly available in the CMR database[41]. The genome of the afore-mentioned bacteria was compared with the *Streptococccus pneumoniae* using multi genome homology comparison algorithm. The algorithm sets a user defined reference DNA molecule and compares this with a query molecule present in the database.

Based on the homology the locus/proteins that show good similarity to each other were retrieved. The cut off value set for the homology was set to an absolute 100% similarity score.

Search for essential genes within the highly similar proteins

The proteins that showed high similarity in *S. pneumoniae* were subjected to a screening for their inevitability for cellular processes. This was done by searching databases for protein sequences that were unique to the genus itself.

REFERENCES

1. Palmieri, C., Varaldo, P.E., and Facinelli, B. *Streptococcus suis*, an Emerging Drug-Resistant Animal and Human Pathogen. *Front Microbiol* 2, 235.
2. Tsai, H.Y., Liao, C.H., Liu, C.Y., Huang, Y.T., Teng, L.J., and Hsueh, P.R. *Streptococcus suis* infection in Taiwan, 2000-2011. Diagn Microbiol Infect Dis.
3. Choi, S.M., Cho, B.H., Choi, K.H., Nam, T.S., Kim, J.T., Park, M.S., Kim, B.C., Kim, M.K., and Cho, K.H. Meningitis caused by *Streptococcus suis*:

case report and review of the literature. *J Clin Neurol* 8, 79-82.
4. Takamatsu, D. Diversity and virulence factors of *Streptococcus suis*. *Nihon Saikingaku Zasshi* 66, 7-21.
5. Fernandez-Ferro, J., Lopez-Gonzalez, F.J., Pardo, F., and Pias-Peleteiro, J.M. [Acute *Streptococcus suis* meningitis in a pig breeder]. *Enferm Infecc Microbiol Clin* 29, 396-397.
6. Ngo, T.H., Tran, T.B., Tran, T.T., Nguyen, V.D., Campbell, J., Pham, H.A., Huynh, H.T., Nguyen, V.V., Bryant, J.E., Tran, T.H., et al. Slaughterhouse pigs are a major reservoir of *Streptococcus suis* serotype 2 capable of causing human infection in southern Vietnam. *PLoS One* 6, e17943.
7. Nakayama, T., Takeuchi, D., Akeda, Y., and Oishi, K. *Streptococcus suis* infection induces [corrected] bacterial accumulation in the kidney. *Microb Pathog* 50, 87-93.
8. Muckle, A., Giles, J., Lund, L., Stewart, T., and Gottschalk, M. Isolation of *Streptococcus suis* from the urine of a clinically ill dog. *Can Vet J* 51, 773-774.
9. Xu, M., Wang, S., Li, L., Lei, L., Liu, Y., Shi, W., Wu, J., Rong, F., Sun, G., Xiang, H., et al. Secondary infection with *Streptococcus suis* serotype 7 increases the virulence of highly pathogenic porcine reproductive and respiratory syndrome virus in pigs. *Virol J* 7, 184.
10. Tan, J.H., Yeh, B.I., and Seet, C.S. Deafness due to haemorrhagic labyrinthitis and a review of relapses in *Streptococcus suis* meningitis. *Singapore Med J* 51, e30-33.
11. Zhang, C.T., and Zhang, R. (2008). Gene essentiality analysis based on DEG, a database of essential genes. *Methods Mol Biol* 416, 391-400.
12. Zhang, R., and Lin, Y. (2009). DEG 5.0, a database of essential genes in both prokaryotes and eukaryotes. *Nucleic Acids Res* 37, D455-458.
13. Zhang, R., Ou, H.Y., and Zhang, C.T. (2004). DEG: a database of essential genes. *Nucleic Acids Res* 32, D271-272.
14. Peterson, J.D., Umayam, L.A., Dickinson, T., Hickey, E.K., and White, O. (2001). The Comprehensive Microbial Resource. *Nucleic Acids Res* 29, 123-125.
15. Maurizi, M.R., Thompson, M.W., Singh, S.K., and Kim, S.H. (1994). Endopeptidase Clp: ATP-dependent Clp protease from Escherichia coli. *Methods Enzymol* 244, 314-331.
16. Maurizi, M.R., Clark, W.P., Katayama, Y., Rudikoff, S., Pumphrey, J., Bowers, B., and Gottesman, S. (1990). Sequence and structure of Clp P, the proteolytic component of the ATP-dependent Clp protease of Escherichia coli. *J Biol Chem* 265, 12536-12545.
17. Gottesman, S., Roche, E., Zhou, Y., and Sauer, R.T. (1998). The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev* 12, 1338-1347.
18. Clarke, A.K. The chloroplast ATP-dependent Clp protease in vascular plants - new dimensions and future challenges. *Physiol Plant* 145, 235-244.
19. Wu, H., Ji, Y., Du, J., Kong, D., Liang, H., and Ling, H.Q. ClpC1, an ATP-dependent Clp protease in plastids, is involved in iron homeostasis in Arabidopsis leaves. *Ann Bot* 105, 823-833.
20. Sjogren, L.L., and Clarke, A.K. Assembly of the chloroplast ATP-dependent Clp protease in Arabidopsis is regulated by the ClpT accessory proteins. *Plant Cell* 23, 322-332.
21. Sjogren, L.L., Stanne, T.M., Zheng, B., Sutinen, S., and Clarke, A.K. (2006). Structural and functional insights into the chloroplast ATP-dependent Clp protease in Arabidopsis. *Plant Cell* 18, 2635-2649.
22. Andersson, F.I., Tryggvesson, A., Sharon, M., Diemand, A.V., Classen, M., Best, C., Schmidt, R., Schelin, J., Stanne, T.M., Bukau, B., et al. (2009). Structure and function of a novel type of ATP-dependent Clp protease. *J Biol Chem* 284, 13519-13532.
23. Skinner, M.M., and Trempy, J.E. (2001). Expression of clpX, an ATPase subunit of the Clp protease, is heat and cold shock inducible in Lactococcus lactis. *J Dairy Sci* 84, 1783-1785.
24. Porankiewicz, J., Wang, J., and Clarke, A.K. (1999). New insights into the ATP-dependent Clp protease: *Escherichia coli* and beyond. *Mol Microbiol* 32, 449-458.
25. Gerth, U., Kruger, E., Derre, I., Msadek, T., and Hecker, M. (1998). Stress induction of the *Bacillus subtilis* clpP gene encoding a homologue of the proteolytic component of the Clp protease and the involvement of ClpP and ClpX in stress tolerance. *Mol Microbiol* 28, 787-802.
26. Zellmeier, S., Schumann, W., and Wiegert, T. (2006). Involvement of Clp protease activity in modulating the Bacillus subtilissigmaw stress response. *Mol Microbiol* 61, 1569-1582.
27. Kruger, E., Zuhlke, D., Witt, E., Ludwig, H., and Hecker, M. (2001). Clp-mediated proteolysis in Gram-positive bacteria is autoregulated by the stability of a repressor. *EMBO J* 20, 852-863.
28. Miethke, M., Hecker, M., and Gerth, U. (2006). Involvement of *Bacillus subtilis* ClpE in CtsR degradation and protein quality control. *J Bacteriol* 188, 4610-4619.
29. Gerth, U., Kirstein, J., Mostertz, J., Waldminghaus, T., Miethke, M., Kock, H., and Hecker, M. (2004). Fine-tuning in regulation of Clp protein content in *Bacillus subtilis*. *J Bacteriol* 186, 179-191.
30. Kock, H., Gerth, U., and Hecker, M. (2004). MurAA, catalysing the first committed step in peptidoglycan biosynthesis, is a target of Clp-dependent proteolysis in *Bacillus subtilis*. *Mol Microbiol* 51, 1087-1102.
31. Kirstein, J., Zuhlke, D., Gerth, U., Turgay, K., and Hecker, M. (2005). A tyrosine kinase and its activator control the activity of the CtsR heat shock repressor in *B. subtilis*. *EMBO J* 24, 3435-3445.
32. Thanassi, J.A., Hartman-Neumann, S.L., Dougherty, T.J., Dougherty, B.A., and Pucci, M.J. (2002). Identification of 113 conserved essential genes using a high-throughput gene disruption system in *Streptococcus pneumoniae*. *Nucleic Acids Res* 30, 3152-3162.
33. Molzen, T.E., Burghout, P., Bootsma, H.J., Brandt, C.T., van der Gaast-de Jongh, C.E., Eleveld, M.J., Verbeek, M.M., Frimodt-Moller, N., Ostergaard, C., and Hermans, P.W. Genome-wide identification of *Streptococcus pneumoniae* genes essential for bacterial replication during experimental meningitis. *Infect Immun* 79, 288-297.
34. Song, J.H., Ko, K.S., Lee, J.Y., Baek, J.Y., Oh, W.S., Yoon, H.S., Jeong, J.Y., and Chun, J. (2005). Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol Cells* 19, 365-374.
35. Bijlsma, J.J., Burghout, P., Kloosterman, T.G., Bootsma, H.J., de Jong, A., Hermans, P.W., and Kuipers, O.P. (2007). Development of genomic array footprinting for identification of conditionally essential genes in *Streptococcus pneumoniae*. *Appl Environ Microbiol* 73, 1514-1524.
36. Lee, M.S., Dougherty, B.A., Madeo, A.C., and Morrison, D.A. (1999). Construction and analysis of a library for random insertional mutagenesis in *Streptococcus pneumoniae*: use for recovery of mutants defective in genetic transformation and for identification of essential genes. *Appl Environ Microbiol* 65, 1883-1890.
37. Porankiewicz, J., Schelin, J., and Clarke, A.K. (1998). The ATP-depen-

dent Clp protease is essential for acclimation to UV-B and low temperature in the cyanobacterium Synechococcus. *Mol Microbiol* 29, 275-283.
38. Shapiro, J.A. (1993). A role for the Clp protease in activating Mu-mediated DNA rearrangements. *J Bacteriol* 175, 2625-2631.
39. Kaakoush, N.O., Deshpande, N.P., Wilkins, M.R., Raftery, M.J., Janitz, K., and Mitchell, H. Comparative analyses of Campylobacter concisusstrains reveal the genome of the reference strain BAA-1457 is not representative of the species. *Gut Pathog* 3, 15.
40. O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., and Copeland, N.G. (1993). Anchored reference loci for comparative genome mapping in mammals. *Nat Genet* 3, 103-112.
41. Hu, P., Yang, M., Zhang, A., Wu, J., Chen, B., Hua, Y., Yu, J., Chen, H., Xiao, J., and Jin, M. Complete genome sequence of *Streptococcus suis* serotype 3 strain ST3. *J Bacteriol* 193, 3428-3429.