Essential Genes Identification in Streptococcus Sanguinis and Comparison among Streptococci

Ping Xu and Xiuchun Ge
The Philips Institute
Virginia Commonwealth University, USA

1 Introduction

Streptococci belonging to the phylum Firmicutes represent a large number of important Gram-positive bacteria. Many streptococcus genera play important roles in human living. For example, many streptococci are important human pathogens for infectious diseases. *Streptococcus pyogenes*, group A streptococcus (GAS), is the leading cause of uncomplicated bacterial pharyngitis and tonsillitis (strep throat) (Churchward, 2007). GAS also relate with sinusitis, otitis, mastoiditis, pneumonia, necrotizing fasciitis or myositisPatterson, 1996 7162 /id. *Streptococcus agalactiae*, group B streptococci (GBS), is the common cause of meningitis and sepsis in newborns (Manning, 2003). This organism is the cause of serious infections in immunocompromised adults. *Streptococcus equi subsp. zooepidemicus* is one of the Group C streptococci that has also been isolated from clinical cases in humans and other animals (Waller *et al.,* 2011). *Streptococcus pneumoniae* is the most common cause of sinusitis, acute bacterial otitis media, and conjunctivitis beyond early childhood. It has more than 90 different antigenic types and is also known to cause pneumonia, meningitis, and bacteremia (Cobey & Lipsitch, 2012). *Streptococcus thermophilis* is important for fermentation industries and commonly used for fermented dairy and cheese products(Kang *et al.,* 2012). *Streptococcus mutans* is a member of the viridans streptococci group and a normal inhabitant in dental plaque. It is a significant contributor to tooth decay or dental caries (Ajdic *et al.,* 2002).

Streptococcus sanguinis, a member of the viridans streptococci group, endocarditis when it gains entrance to the human bloodstream (Xu *et al.,* 2007). The complete S. *sanguinis* SK36 genome sequence is 2.39 Mb and annotated 2270 is a normal inhabitant in dental plaque. It is one of the organisms causing infective putative genes. Based on this genome information, we recently attempted to construct genome-wide single-gene deletion mutants in S. *sanguinis* SK36 (Xu *et al.,* 2011). The two most common methods utilized to construct gene knock-out in bacteria are transposon mutagenesis (van Opijnen *et al.,* 2009) and single-gene deletion (Gerdes *et al.,* 2003). While transposon mutagenesis is less time-consuming, less costly, and does not require completed genome information, it has two weaknesses: (1) the possibility of disparate mutants in the mixed mutant library that counter-selects mutants with decreased competition; and (2) the possibility of partial gene inactivation whereby genes do not entirely lose their function following the insertion of a transposon. We established a high-throughput technique for single-gene deletion using

S. *sanguinis* as a model organism and identified all essential genes. We report here the identified essential genes and their comparisons among streptococci.

2 Methodology

2.1 Bacterial growth and transformation

The genome of S. *sanguinis* strain SK36 including gene annotation was stored in our local computer server(Xu *et al.,* 2007). All of the SK 36 strain and its mutants were cultured at 37°C in a brain heart infusion (BHI) broth (BD) as described previously(Ge *et al.,* 2008). Unless otherwise stated, all streptococcal strains were cultured under microaerobic conditions (6% O_2, 7.2% CO_2, 7.2% H_2 and 79.6% N_2). Two other streptococcal species, S. *pneumoniae* TIGR4 and S. *mutans* UA159, were grown in a Todd Hewitt (TH; BD) broth plus 0.5% yeast extract and BHI broth, respectively.

In order to prepare competent bacterial cells that would obtain high transformation efficiency in the genome-wide deletions, the following procedures were used. A TH medium with pH 7.6 was prepared using 10N NaOH. The medium was then boiled and allowed to cool to room temperature. A 0.22 μm polystyrene filter was used to sterilize the cooled medium. Heat-inactivated horse serum was added to the TH medium to a final concentration of 2.5% to create the TH+HS medium. The TH+HS medium were aliquot into a tube of 2 ml and 4 tubes of 10 ml in 15-ml conical tubes. A stock frozen S. *sanguinis* SK36 strain at -80°C was inoculated into a 2 ml TH+HS medium. After the tube was capped tightly, the culture was incubated at 37°C. The bacterial cells were inoculated with 1/200 dilution into pre-incubated 10 ml TH+HS medium after growing overnight. After further incubation at 37°C for 3 h, the bacterial cells were immediately used for transformation. Two microliters of 70 ng S. *sanguinis* SK36 competence stimulating peptide (CSP) and 2 l linear recombinant PCR amplicon (~50 ng) were added to Eppendorf tubes on 96-well block and pre-warmed at 37°C. This peptide sequence of the cleaved form CSP was DLRGVPNPWGWIFGR. Three hundred and thirty microliters of bacterial cells incubated for 3 hours in TH+HS medium were added to each tube. After the mixtures were incubated at 37°C for 1 hour, the transformants were spread on BHI agar plates with kanamycin for selection.

2.2 High throughput PCR deletion mutagenesis

Primers were designed based on the complete S. *sanguinis* SK36 genome sequence (Xu *et al.,* 2007). Three sets of primers (F1/R1, F2/R2 and F3/R3) were designed using in house script to amplify 1 kb up-stream flanking region of the target gene, a kanamycin-resistant (Km^r) gene cassette, and 1-kb down-stream flanking region of the target gene, respectively. ePrimer3 in the EMBOSS suite of the programs (`http://emboss.sourceforge.net/apps/cvs/emboss/apps/index.html`) was used to design all primers. R1 and F3 primers contain 25-bp adaptor sequences at their 5' end that are complementary to the Km^r gene cassette. The R1 and F3 primers were designed to delete the coding region from 6 bp after the start codon to 30 bp before the stop codon in order to precisely delete the target genes. Our previous report describes the details of the primer designs (Xu *et al.,* 2011). In order to confirm the mutants, the P1, P2, and various T1 primers were designed for sequencing. All primers were designed with the melting temperatures (Tm) to be as close as possible to 60 ^{o}C for multiple mutant constructions in high throughput PCR amplification.

For high throughput PCR deletion mutagenesis, all PCR reactions were performed at a 96-well

format. For PCR amplification of up- and down-stream regions, cocktail mixtures were prepared for final concentration including 1 X High Fidelity PCR Buffer, 0.8 μl of 10 mM dNTP mixture, 0.4 μl of 50 mM MgSO4, 0.8 μl of 10 ng/μl S. *sanguinis* SK36 genomic DNA and 0.1 μl Platinum® Taq DNA polymerase high fidelity, and different primer pairs 1 μl of each from the 10 M working solutions. All PCR were performed at 94°C for 1 min, 30 cycles of 94°C for 30 sec, 54°C for 30 sec, and 68°C for 1.5 min. A DNA fragment containing the Kmr cassette was digested using EcoRI for the PCR template. The PCR was amplified at 94 ^{o}C for 1min, and 30 cycles of 94°C for 30 sec, 55°C for 30 sec and 68^{o}C for 1 min. A bulk of the Kmr cassette amplicon from 10 individual PCR reactions was purified by the QIAquick PCR purification kit. The Kmr cassette amplicon was adjusted to a final concentration of 10 $\mu g/\mu l$ for recombinant PCR. Finally, three PCR amplicons (up-stream regions, Kmr gene, and down-stream regions) were combined with nearly equal-molar amounts into one PCR plate well for creating final recombinant amplicons. PCR amplifications were performed using F1 and R3 primers at 94°C for 2 min, 30 cycles of 94°C for 30 sec, 55°C for 30 sec, and 68^{o}C for 3.5 min, and finally the extension cycle at 68°C for 4 min. To avoid remaining non-specific primers and dNTP residues between different PCR, all amplicons were purified by a PureLink 96 PCR purification kit (Invitrogen) using centrifugation according to the manufacturers instructions. The final recombinant PCR amplicons binding on the PureLink resin were eluted by 40 μl ddH_2O and the DNA concentrations were determined by Nanodrop before transformation.

2.3 Gene function analysis

The gene functions were from *sanguinis* genome annotation (Xu *et al.*, 2007). For genes of metabolic pathways, the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was used to download the functions of the genes in the S. *sanguinis* genome. Analysis of the assigned KO numbers and the pathway for the S. *sanguinis* genes was performed. The missing essential genes in other streptococcal species were further confirmed by both a BLAST search with a lower cut-off E value and manual examination of their genome annotations.

2.4 Gene Comparison

To compare genes among bacterial genomes, forty eight completed streptococcal genomes were downloaded from NCBI and annotated protein sequences were stored in our local database in order to compare streptococcal genomes. BLASTP program was used for the protein sequence comparisons as previously described (Xu *et al.,* 2004). In order to identify gene orthologs among streptococcal genomes, BLASTP with $E < 1e-5$ was used to find the reciprocal best hits of database searches. For those genes absent in one or two strains, TBLASTN search program was performed using the identified gene sequences in the strains to recover miss-annotated genes. Gene paralogs in S. *sanguinis* genome were identified using BLASTP search with $E < 1e-5$. Protein sequence conservation was calculated by using the percent amino acid identity of the orthologs.

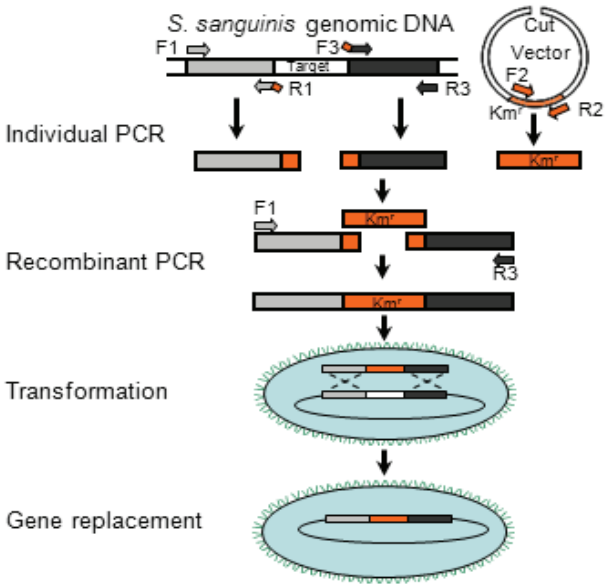


Figure 1: Gene replacement by a recombinant PCR in S. sanguinis. The up- and down-stream flanking regions of the target gene and Kmr were PCR amplified by F1/R1, F3/R3 and F2/R2 primers, respectively. A combined PCR was performed using F1/R3 to obtain the recombinant linear DNA for target gene replacement.

3 Results and discussions

3.1 Screening essential genes

We created a PCR recombinant method to produce mutants (Fig. 1). High throughput format was applied for genome-wide mutant construction. All ORFs annotated in the S. *sanguinis* genome were mutated by using over 10,000 primers and over 9000 PCR amplicons (Xu *et al.,* 2011). During the genome-wide S. sanguinis gene deletions, we examined essential genes. We found two types of essential genes. The first type of essential gene was that the gene deletions failed to produce a transformant. In order to ensure that the gene deletion failure was not a result of low transformation efficiency of the competent cells, we performed a second cycle of construction in which a new amplicon of this gene was re-amplified and re-transformed in parallel with a non-essential gene, SSA_0169 that encodes a hypothetical protein. The gene was classified as an essential gene if it had failed to be deleted up to five independent deletion attempts. We finally obtained 60 of these types of essential genes for S. *sanguinis* (Xu *et al.,* 2011).

The second type of essential gene was that the gene deletion produced mutant colonies but all mutants contained two amplicons (double-band) in agarose gel electrophoresis when PCR amplifications were performed using F1 and R3 primers. The size of one DNA band typically corresponded to the size expected for the replacement mutant while the other matched with the wild-type gene. These mutants may be produced from pressures of both selections for kanamycin resistance and the essential gene for bacterial viability. A genome recombination might occur under both pressures and produce a double-band mutant.

One DNA band was from the wild type while the other was from the mutated DNA containing Km^r cassette. We applied longer agarose gel electrophoresis or designed internal T1 primers to determine those double-band mutants. We finally identified a total of 158 double-band essential genes from both longer gel electrophoresis and T1 internal primer PCR amplification performance (Xu *et al.,* 2011).

3.2 Validation of essential genes

During our essential gene screen, we validated essential genes. In our essential gene identification, we considered two false result genes that would interfere with our study, the false positive essential genes and the false negative essential genes. The false positive genes are when constructs result in no transformants because of certain technical limitations, for example, the failed transformation. The false negative genes are when constructs result in transformants but the target gene remains in the mutant, for example, the double-band mutants as mentioned above. To validate essential genes, we examined all mutants to confirm the deletion of the target genes. Different methods were applied to validate mutants. We first applied colony PCR to identify the correct transformants. The PCR amplicons were then sequenced to confirm that the Km^r cassette was correctly inserted. The mutants were further examined to identify the double-band mutants as mentioned above. For no growth mutants on kanamycin antibiotic selection plates, we performed five gene deletions independently before we concluded that they were essential genes. Initially, a promoterless Kmr cassette was used to create the deletion mutants. However, we found that many of the genes for which mutants could not be obtained were annotated as acquired via horizontal gene transfer or as encoding hypothetical proteins. We considered the possibility that those failed mutants might result from the non-expression of the Km^r cassette based on our microarray analysis (Xu *et al.,* 2011). We further applied promoter Km^{r}cassette to create the deletion mutants. After examination of 142 genes that did not generate recoverable mutants using promoter Km^r cassette, 60 additional mutants were identified (Xu *et al.,* 2011). The results indicated that these 60 additional mutants may either have no functions or may not be expressed well in S. *sanguinis*. Through these results, insight was gained into the gene expressions for S. *sanguinis* genome annotation.

3.3 Essential genes clustered for three basic pathways

After validating all essential genes in S. *sanguinis*, we tried to find the causes for gene essentiality. All essential genes were examined as a whole using systems biology approach. We linked all essential genes together and found that many essential genes were associated with metabolic pathways. The KEGG database was used to examine the functions of the essential genes. Using KEGG pathway maps, the distribution of essential genes in specific categories was studied (Kanehisa *et al.,* 2004). This revealed that the essential genes were a component of various differing biological functions, such as carbohydrate metabolism, nucleotide metabolism, and DNA replication. However, numerous genes were also discovered to be clustered together in certain pathways, including glycolysis (Fig. 2), pentose phosphate pathway (PPP), and peptidoglycan biosynthesis. There were 2270 genes in the S. *sanguinis* genome. The probability of one gene for a pathway to be due to chance will be one out of 2270 genes. In figure 2, there were 9 essential genes clustered together for converting α-D-glucose-6P to pyruvate in the glycolysis/gluconeogenesis pathway. The probability by chance will be (1/2270)9 = 1-30. It is nearly impossible by chance. This result of 9 essential genes clustering together also suggested the accuracy of our essential gene identification.

In order to study the associations among essential genes as a whole, systems biology was utilized.

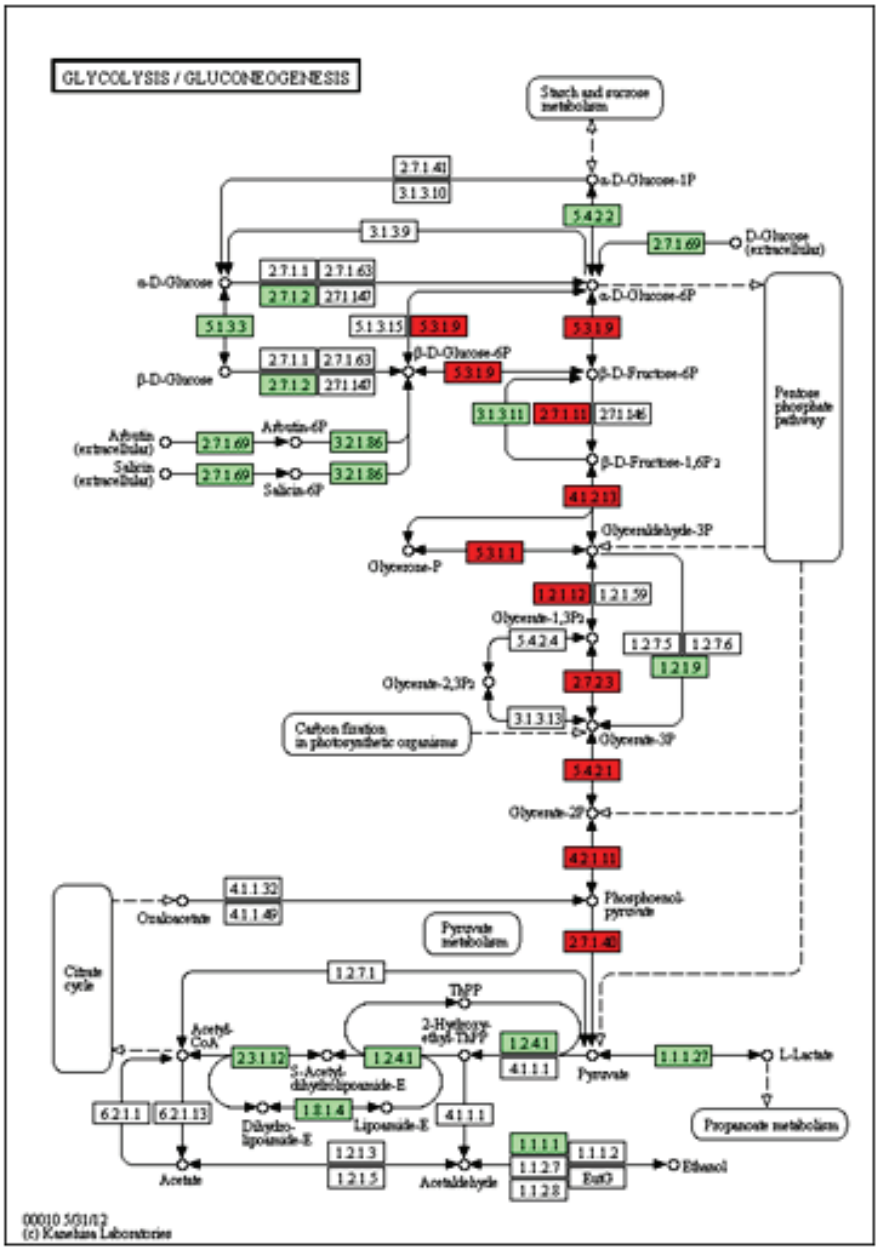


Figure 2: Essential genes clustered in KEGG glycolysis/gluconeogenesis pathway. The essential genes clustered together for converting α-D-glucose-6P to pyruvate in KEGG glycolysis pathway. Red color, essential gene EC (enzyme commission) number; Green color, non-essential genes EC number; White color, absent gene

All essential genes were assigned to biochemical pathways based on their annotations made in the KEGG pathways. As many essential genes as possible were linked together through their involved biochemical pathways. All associations among the essential genes in KEGG categories were then examined. It became apparent that many essential genes were correlated with specific pathways (Fig. 3) for three basic biological functions: maintenance of the cell envelope, energy production, and processing of genetic information. This result suggested the importance of studying essential genes as a whole using systems biology.

3.4 Confirmation of essential genes associated with three basic biological functions

In order to confirm our discovery that essential genes correspond to three basic biological functions, additional experiments were conducted. Various experiments were developed to test the association between the essential genes and the three biological functions. Four different experiments were performed. All of these experimental results supported our finding (Xu *et al.,* 2011). We first examined the isozymes and gene paralogs. We initially focused on the non-essential genes that interrupted the associations of essential genes in pathways. Genes that were expected to be important due to their correlations with other essential genes were found to be non-essential. Upon closer inspection of these non-essential genes in genome annotation, it was evident that the majority of them had complimentary genes with similar functions. It was revealed that many of them had either isozymes or paralogs based on our genome annotation. It was assumed that the deletion of one of these genes could be complemented by another gene with similar function in our single gene deletion mutagenesis. In order to test this hypothesis, we identified all gene paralogs based on protein sequence homology search using BLASTP. We also collected all potential isozyme genes from genome annotation and literature searches. We then selected four gene sets that had only two potential isozymes or paralogs for double gene knockout. Of the four selected, three sets were confirmed to be essential in double gene knockout. A potential alternative pathway might have caused the last set of genes to be shown as non-essential in this double gene knockout. Secondly, we examined the growth of mutants in nutrition limit medium. Genes involving the de novo biosynthesis for amino acids and nucleotides should be essential based on our hypothesis of essential genes associated with the three biological functions. However, for de novo biosynthesis of amino acids and nucleotides, we still obtained 124 strains of gene mutations in a BHI rich medium. In order to determine if the BHI medium provided the required amino acids and nucleotide precursors, we cultured 124 mutant strains in a nutrition limit chemically defined medium (CDM) and compared their growth to those in the BHI. The results indicated that the BHI did provide the required amino acids and nucleotide precursors, since the mutants could not grow well in the CDM, except when there were paralogous genes. The results proved our hypothesis that genes for de novo biosynthesis of amino acids and nucleotides are essential.

Thirdly, we predicted essential genes in S. *pneumoniae* TIGR4. There were two genome-wide essential gene screens of S. *pneumonia* (Thanassi *et al.,* 2002; Song *et al.,* 2005). Based on the pathway functions, eight previously unknown essential genes in the S. *pneumoniae* TIGR4 strain were predicted. We examined these gene essentialities. All eight genes in S. *pneumoniae* were shown to be essential in this experiment.

Finally, we examined apparent inconsistent essential genes based on homologous protein sequences in two other gram-positive bacteria, *Staphylococcus aureus* NCTC8325 and *Bacillus subtilis* 168 (Kobayashi *et al.,* 2003; Chaudhuri *et al.,* 2009). We found that some essential genes in S. *sanguinis* were non-essential in S. *aureus* and B. *subtilis*, and vice versa. Re-examination of all inconsistent genes that correlated to the

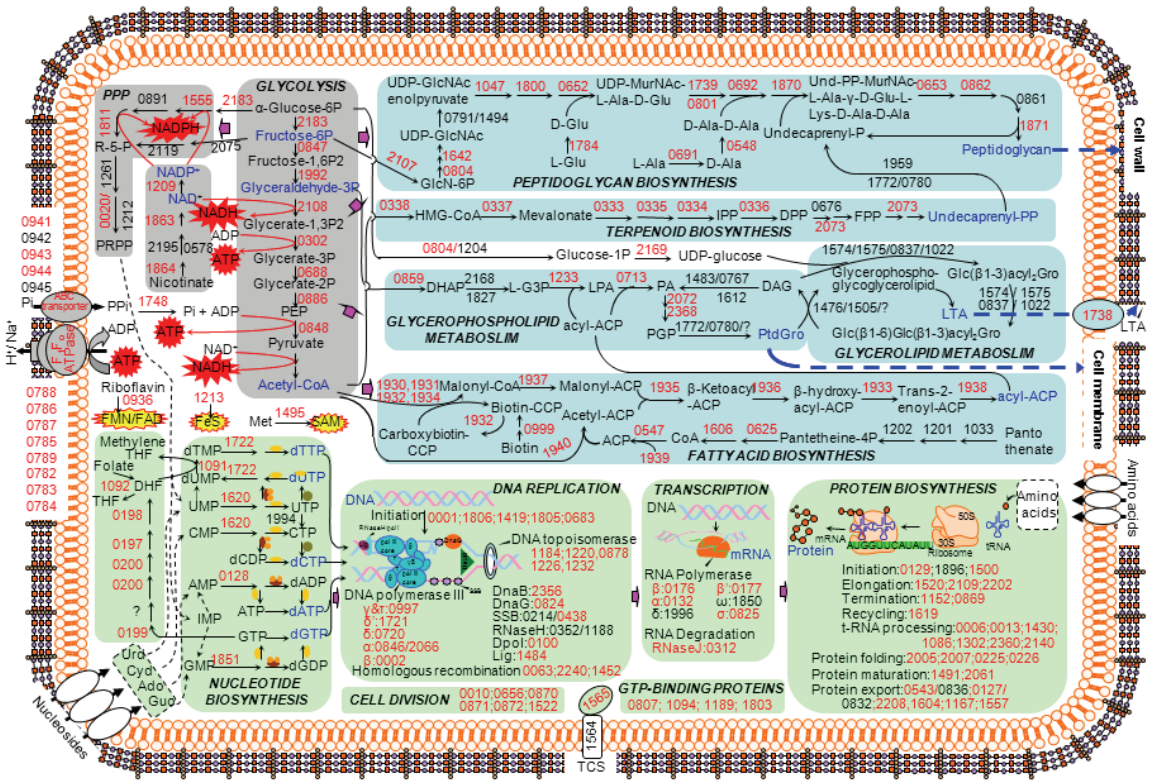
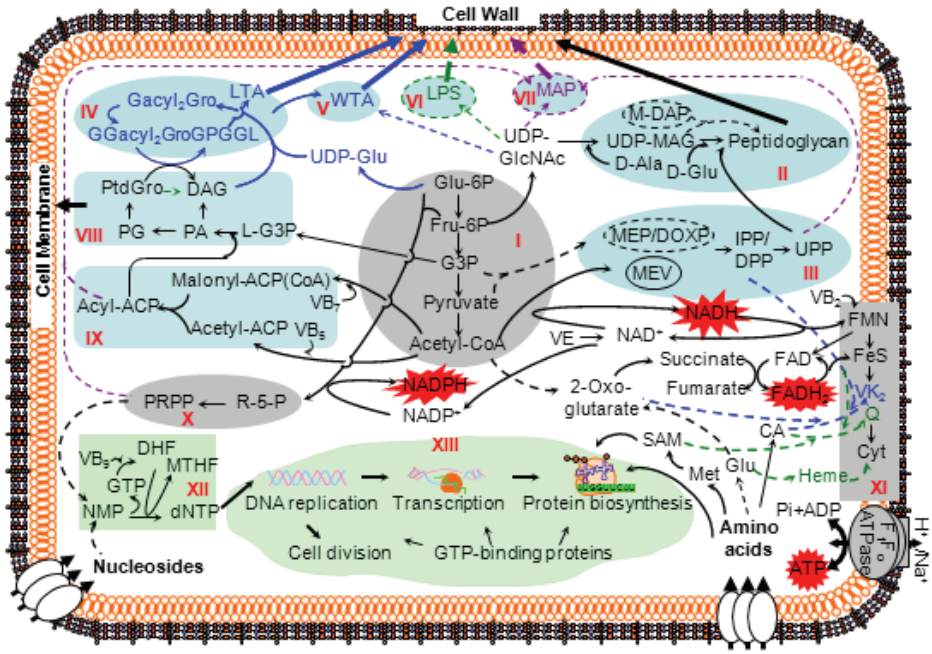


Figure 3: pathways of S. sanguinis SK36 essential genes. The three functions associated with essential pathways are indicated by color: pale blue, cell envelope; gray, energy production; light green, processing of genetic information. Numbers are SSA#; red number, essential gene; black number, nonessential gene; solid arrow, an enzymatic reaction; dashed arrow, multistep pathway; blue dashed arrow, products involved in cell wall and membrane formation; block arrow, product from one pathway serving as input to another pathway; oval with bold arrow, transporter; slash, separating paralogs. Ado, adenosine; Cyd, cytidine; DAG, 1,2-diacylglycerol; DHAP, dihydroxyacetone phosphate; DHF, dihydrofolate; L-G3P, sn-glycerol 3-phosphate; Guo, guanosine; LPA, lysophosphatidic acid; LTA, lipoteichoic acid; PA, phosphatidic acid; PGP, phosphatidylglycerophosphate; PPP, pentose phosphate pathway; TCS, two-component system; THF, tetrahydrofolate; Urd, uridine. In nucleotide biosynthesis, symbols are used to represent essential genes that participate in multistep reactions: SSA_1865 (green circle), SSA_0848 (yellow oval), SSA_2263, SSA_0771 (orange circles) and SSA_0768 and SSA_0770 (brown oval). Essential genes associated with ribosome and aminoacyl-tRNA biosynthesis are not indicated here, nor are SSA_0575, SSA_0800 and SSA_1903, which could not be categorized into these pathways.



Permission to reuse figure granted by Nature Publishing Group

Figure 4: Deduced essential pathways based on S. sanguinis essential genes. Colors indicate the three functions associated with essential pathways, as in Figure 3. I, glycolysis and pyruvate oxidation; II, peptidoglycan biosynthesis; III, terpenoid backbone biosynthesis; IV, glycerolipid metabolism (LTA biosynthesis); V, WTA biosynthesis; VI, lipopolysaccharide biosynthesis; VII, MAP biosynthesis; VIII, phosphoglycerolipid metabolism; IX, fatty acid biosynthesis; X, electron transport chain; XI, pentose phosphate pathway; XII, nucleotide biosynthesis; XIII, genetic information processing including DNA replication, transcription, protein biosynthesis, GTP-binding proteins and cell division. Black arrow, all bacteria; black dashed arrow or oval, alternative pathways; blue arrow, gram-positive bacteria only; green arrow, gram-negative bacteria only; violet arrow, mycobacteria only; bold arrow, synthesis of cell envelope. CA, chorismate; Cyt, cytochrome; DAG, 1,2-diacylglycerol; DG-DAG, diglucosyl-diacylglycerol; G3P,D-glyceraldehyde 3-phosphate; GPGGL, glycerophosphoglycoglycerolipid; L-G3P, glycerol-3-phosphate; LPS, lipopolysaccharide; Malonyl-ACP(CoA), malonyl-ACP or malonyl-CoA; M-DAP, meso-2,6-diaminopimelate; MEV, mevalonate; MG-DAG, monoglucosyl-diacylglycerol; PA, phosphatidate; PG, phosphatidylglycerophosphate; R-5-P, D-Rribulose 5-phosphate; PPP, pentose phosphate pathway; PtdGro, phosphatidylglycerol; UDP-MAG, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate; UPP, undecaprenyl diphosphate.

three basic biological functions in both strains was conducted. It was revealed that most of the differences could be accounted for gene paralogs, isozymes, or alternative pathways (Xu *et al.,* 2011).

3.5 Simple model for prediction essential genes of bacterial species

As a consequence of climate and environmental changes and of increased global trade and travel frequency, emerging pathogens are increasing and become a threat to the global economy by affecting human and animal health. Also, many antimicrobial resistant strains and superbugs have emerged among pathogenic bacteria after the wide-board applications of antibiotics. It becomes a challenge to combat emerging pathogens. A clear perception of gene essentiality in bacterial pathogens is important for identifying drug targets to combat emerging pathogens and antibiotic-resistant bacteria. Current development of genomics and bioinformatics provide opportunities to identify drug targets more efficiently. The genomic information of the emerging pathogens can be quickly obtained by next generation sequencing techniques. Based on our finding we have developed a model to predict essential genes in bacterial species (Xu *et al.,* 2011) (Fig. 4). By this model, the essential genes of a bacterial strain are ready to be identified when its genome becomes available. This model may be used for emerging pathogens or for bacterial species without genetic manipulation systems. These putative essential genes can be further confirmed by gene knockout experiments. The identified essential genes will become targets for chemotherapeutic agents and immunotherapeutic development.

3.6 Comparison of essential genes among streptococci

Streptococci are a large group of important bacteria including severe pathogens for humans and animals. Essential genes are critical for streptococci survival and are associated with virulence. Comparisons of essential genes among streptococci may reveal the evolution of genes for bacterial survival in different niches. We used comparative genomics (Xu *et al.,* 2004) to examine the conservation of the genes among streptococcal species we identified as essential. To compare, we downloaded 48 complete streptococcal genomes and their annotated genes that are publicly available from the NCBI database and compared them with the S. *sanguinis* essential genes (Fig. 5). The result indicated that most of S. *sanguinis* essential genes had orthologs in other 48 streptococcal genomes. By comparison of annotated protein genes, we found that 16 S. *sanguinis* essential genes were absent in some of streptococcal strains (Xu *et al.,* 2011). We carefully examined streptococcal genomes recently and recovered 13 missed genes that previously were found absent in other streptococcal strains (Fig. 5). Those genes were most likely missed in the annotations. We applied TBLASTN search against the original complete genome of those strains and identified the missed gene sequences. The nucleotide sequence homologs for 13 genes were found in those strains. Most of these sequences were not annotated as genes in those genome annotations. Thus, the result was in agreement with the expectation that the vast majority of the essential genes should be highly conserved within streptococci. In general, we found that essential genes were conserved in our comparison. Most of S. it sanguinis essential genes (215 of 218) were found present in other 48 streptococcal genomes. Protein sequences of essential genes were also found to be more conserved among streptococci than that of non-essential genes. The average identity for the essential gene sequences among 49 streptococcal genomes was 79.71% as compared to the non-essential gene sequences of 58.96%. It is in agreement with the expectation that essential genes change less during evolution.

Only 3 genes had no ortholog in some groups of streptococci. The essential gene SSA_0862 was

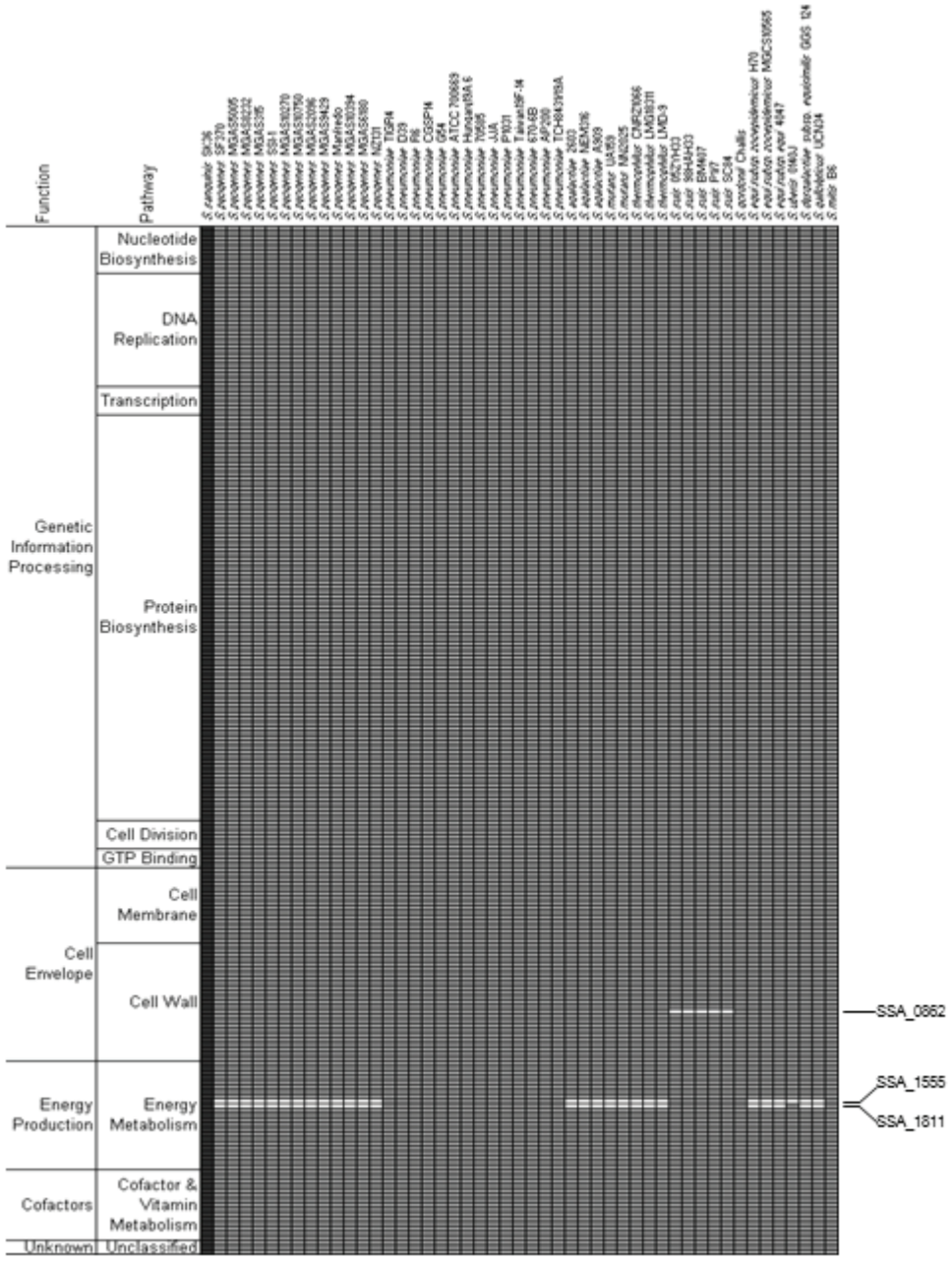


Figure 5: Essential gene comparisons among streptococci. S. sanguinis essential genes were compared with 48 other streptococcal genomes. The genes related functions and pathways were indicated. Grey rectangle, ortholog present; White rectangle, ortholog absent. Three genes for SSA_0862, SSA_1555 and SSA_1811 were absent in some strains of streptococci.

absent in all 5 strains of the compared *Streptococcus suis*. Two essential genes, SSA_1555 and SSA_1811, were absent in 27 and 26 other streptococcal genomes, respectively. Both genes were absent in 13 *S. pyogenes*, 3 *S. agalactiae*, 2 *S. mutans*, 3 *S. thermophiles*, 3 *S.equi*, 1 *S. dysgalactiae* and 1 *S. gallolyticus* strains. SSA_1811 but not SSA_1555 was found in S. *uberis*. SSA_1811 was found in S. *uberis* with a low identity of 36.27%. In one of the streptococcal species, one essential gene may be missing in a different streptococcal species. For example, three of the essential genes in S. *sanguinis* were found missing in some of the streptococcal species. The *murM* gene (SSA_0862) in streptococci is involved in the last several steps of peptidoglycan biosynthesis. This gene has similar functions to *femAB* genes in staphylococci. Although 5 strains of the examined S. *suis* lack the ortholog of the SSA_0862 gene, this ortholog was found in two other S. *suis* strains, R61 and D12, in the NCBI database. This suggests that the*murM* gene may still participate in peptidoglycan biosynthesis in S. *suis* and there is possibly another gene substituted for the function of the *murM* gene in those 5 S. *suis* strains. Based on gene and pathway annotation, it was determined that SSA_1555 and SSA_1811 are located in the oxidative phase from D-glucose 6-phosphate to D-ribulose 5-phosphate of the pentose phosphate pathway. However, the enzymes of the oxidative phase in the pentose phosphate pathway were completely absent in streptococci without orthologs of SSA_1555 and SSA_1811. Both glucose-6-phosphate 1-dehydrogenase and 6-phosphogluconate dehydrogenase, encoded by SSA_1555 and SSA_1811 in S. *sanguinis*, respectively, are involved in NADPH production. NADPH production supplies energy and is crucial to cell metabolism during bacterial cell envelope biosynthesis and genetic information processing. For example, NADPH will serve as an electron donor in the reactions during fatty acid biosynthesis. NADPH production will be precipitated by glucose-6-phosphate, 1-dehydrogenase and 6-phosphogluconate dehydrogenase in streptococci without an oxidative phase. As such, it is possible that other essential genes responsible for NADPH production exist in those streptococci. It is important to identify essential genes not only for identification of drug targets against emerging pathogens and antibiotic resistance bacteria, but also for synthetic biology in the designing of new bacterial cells that requires an understanding of the minimal gene set (Glass *et al.,* 2006; Forster & Church, 2006). From our finding, the essential genes of a bacterial strain are ready to be identified when its genome becomes available. We developed a simple model to predict essential genes via bacterial genome annotation (Xu *et al.,* 2011). This model may be used for emerging pathogens or for bacterial species without genetic manipulation systems. These putative essential genes may be further confirmed by gene knockout experiments to eliminate potential unknown isozymes or alternative pathways. In this study, we obtained an ordered, comprehensive set of non-essential gene mutants of S. *sanguinis* SK36. This mutant library will become a great resource for gene biological functional studies for S. *sanguinis* SK36. The comprehensive set of mutants will be especially useful in systems biology research as we discussed in this chapter for the essential gene identification. Such a global analysis of genes as a whole using systems biology approach may provide a comprehensive understanding of gene biological functions or characteristics that cannot be revealed by any single or small group of gene mutations. This systems biological approach may pave the way to a new, global research emphasis in the post-genomic era and may serve as novel approach in the study of microbiology and bacterial gene functions.

References

Ajdic, D., McShan, W.M., McLaughlin, R.E., Savic, G., Chang, J., Carson, M.B., Primeaux, C., Tian, R., Kenton, S., Jia, H., Lin, S., Qian, Y., Li, S., Zhu, H., Najar, F., Lai, H., White, J., Roe, B.A. & Ferretti, J.J. (2002). Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. *Proc Natl Acad Sci U S A*, 99, 14434-14439.

Chaudhuri, R.R., Allen, A.G., Owen, P.J., Shalom, G., Stone, K., Harrison, M., Burgis, T.A., Lockyer, M., Garcia-Lara, J., Foster, S.J., Pleasance, S.J., Peters, S.E., Maskell, D.J. & Charles I.G. (2009) Comprehensive identification of essential *Staphylococcus aureus* genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics*, 10, 291.

Churchward, G. (2007). The two faces of Janus: virulence gene regulation by CovR/S in group A streptococci. *Mol Microbiol*, 64, 34-41.

Cobey, S. & Lipsitch, M. (2012). Niche and neutral effects of acquired immunity permit coexistence of pneumococcal serotypes. *Science*, 335, 1376-1380.

Forster, A.C. & Church, G.M. (2006) Towards synthesis of a minimal cell. *Mol Syst Biol*, 2, 45.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balazsi, G., Ravasz, E., Daugherty, M.D., Somera, A.L., Kyrpides ,N.C., Anderson, I., Gelfand, M.S., Bhattacharya, A., Kapatral, V., D'Souza, M., Baev, M.V., Grechkin, Y., Mseeh, F., Fonstein, M.Y., Overbeek, R., Barabsi, A.L., Oltvai, Z.N. & Osterman, A.L. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of Bacteriology*, 185, 5673-5684.

Ge, X., Kitten, T., Chen, Z., Lee, S.P., Munro, C.L. & Xu, P. (2008). Identification of Streptococcus sanguinis genes required for biofilm formation and examination of their role in endocarditis virulence. *Infect Immun*, 76, 2551-2559.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O. & Venter J.C. (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A*, 103, 425-430.

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. & Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res*, 32, Database issue: D277-D280.

Kang, X., Ling, N., Sun, G., Zhou, Q., Zhang, L. & Sheng, Q. (2012). Complete Genome Sequence of Streptococcus thermophilus Strain MN-ZLW-002. *J Bacteriol*, 194, 4428-4429.

Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M. et al. (2003). Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A*, 100, 4678-4683.

Manning, S.D. (2003). Molecular epidemiology of *Streptococcus agalactiae* (group B Streptococcus). *Front Biosci*, 8, s1-18.

Song, J.H., Ko, K.S., Lee, J.Y., Baek, J.Y., Oh, W.S., Yoon, H.S., Jeong J.Y. & Chun J. (2005). Identification of essential genes in *Streptococcus pneumoniae* by allelic replacement mutagenesis. *Mol Cells*, 19, 365-374.

Thanassi, J.A., Hartman-Neumann, S.L., Dougherty, T.J., Dougherty, B.A. & 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.

van Opijnen, T., Bodi,K. L., and Camilli, A. (2009). Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods*, 6, 767-772.

Waller, A.S., Paillot, R. & Timoney, J.F. (2011). *Streptococcus equi*: a pathogen restricted to one host. *J Med Microbiol*, 60, 1231-1240.

Xu, P., Alves, J.M., Kitten, T., Brown, A., Chen, Z., Ozaki, L.S., Manque, P., Ge, X., Serrano, M.G., Puiu, D., Hendricks, S., Wang ,Y., Chaplin, M.D., Akan, D., Paik, S., Peterson, D.L., Macrina, F.L. & Buck, G.A. (2007). Genome of the opportunistic pathogen *Streptococcus sanguinis. J Bacteriol*. 189, 3166-3175

Xu, P., Ge, X., Chen, L., Wang, X., Dou, Y., Xu, J.Z., Patel, J.R., Stone, V., Trinh, M., Evans, K., Kitten, T., Bonchev, D. & Buck, G.A.. (2011). Genome-wide essential gene identification in *Streptococcus sanguinis. Scientific Reports*, 1, doi:10.1038/srep00125 .

Xu, P., Widmer, G., Wang, Y., Ozaki, L.S., Alves, J.M., Serrano, M.G., Puiu, D., Manque, P., Akiyoshi, D., Mackey, A.J., Pearson, W.R., Dear ,P.H., Bankier, A.T., Peterson, D.L., Abrahamsen, M.S., Kapur, V., Tzipori, S. & Buck, G.A. (2004). The genome of *Cryptosporidium hominis. Nature*, 431, 1107-1112.