Sequencing, Assembly, and Annotation of the *Kaistella koreensis* Genome and Comparison to Closely Related Organisms

Presented to the faculty of Lycoming College in partial fulfillment of the requirements for Departmental Honors in Biology

by

Timothy Hostelley

Lycoming College

April 22, 2013

Approved by:

_____

(Signature)

_____

(Signature)

_____

(Signature)

_____

(Signature)

Abstract

Advances in DNA sequencing technology have made DNA sequencing cheaper and more efficient. As a result, there has been an enormous increase in the number of genomes being sequenced. The sequence data can be assembled into complete genomes and annotated in order to reveal information about the organism's physiology. In this study the DNA of the bacterium *Kaistella koreensis* was sequenced and assembled into 578 contigs, these contigs were then uploaded to Rapid Annotation Using Subsystems Technology for annotation in order to compute the Average Nucleotide Identity between *K. koreensis* and closely related organisms in order to dispute the reclassification of *K. koreensis* as *Chryseobacterium koreense*. Phenotypic tests including Biolog GenII, API ZYM, and Fatty Acid Methyl Ester analysis were also done in order to supplement the ambiguous results of the ANI. The results of these tests reveal a number of significant differences between *K. koreensis* and its closest related neighbors that suggests that *K. koreensis* does not belong in the *Chryseobacterium* genus or the closely related *Lycomia* genus. Instead, *K. koreensis* should be reclassified back to its original classification in the *Kaistella* genus. This would dispute the proposal made by Kämpfer *et al*. to reclassify *Kaistella koreensis* into the *Chryseobacterium* genus.

Introduction

DNA sequencing has rapidly evolved from the earliest sequencing efforts using only whole genome shotgun-cloning based sequencing of the 1990's, to the further advances in Sanger sequencing in the early 2000's. This made possible the sequencing of larger and more complex genomes. The next development in sequencing took place around 2005, with the advent of next generation sequencing technologies that could sequence entire genomes at a much lower cost. This made the technology more accessible to researchers (Henson, Tischler and Ning 2012). The rapid advancement of DNA sequencing technology has also been driven by the goal of the $1,000 human genome, by which any person could have their entire genome sequenced for $1,000, making it possible as a normal medical test.

In this project, the next generation genome sequencing was done using Ion-Torrent technology. Ion-Torrent determines the DNA sequence via sequencing by synthesis method. In this method, the sequence of single stranded DNA is determined by synthesizing the complementary strand. The DNA is prepared for sequencing using emulsion PCR which isolates single stranded DNA molecules with primer coated beads in aqueous droplets. PCR then coats each bead with copies of the DNA molecule. The beads are then inserted into a well, which contains the enzyme DNA polymerase in order to synthesize the complementary strand, on a plate in the instrument (Life Technologies Corporation). The instrument then alternates flooding the plate with the adenine, cytosine, thymine, and guanine nucleotide bases. If the base is complementary to the next base on the template strand (adenine pairs with thymine and cytosine pairs with guanine) it gets incorporated into the strand being synthesized by

the DNA polymerase. The instrument is able to determine which base has been added because, when a base gets incorporated into the strand, a pyrophosphate and a hydrogen ion are released (Life Technologies Corporation). The release of the hydrogen ion causes the pH to become slightly more acidic. This change in pH is recognized by the Ion-Torrent instrument, which acts as a pH meter. In homopolymeric regions of the DNA, multiple complementary bases will be incorporated at the same time. This results in an even greater change in pH, and the instrument determines how many bases have been added by the degree of change in pH (Life Technologies Corporation). One problem this can cause is that, in longer homopolymeric regions, the instrument can experience difficulties determining exactly how many bases were incorporated. The next generation sequencers do not, however, give the entire sequence of the genome as a single piece of data. The sequencers give reads that consist of parts of the genome which then must be assembled to give the complete genome. The average read length for Ion-Torrent sequencing is 200 bases, however, different sequencers yield different average lengths including; 400-900 bases using Sanger sequencing, 50-250 bases for Illumina sequencing, and an average of 2,900 bases for Pacific Bio sequencing.

Herein lies the problem facing researchers today; assembly technology is not quite up to par with sequencing technology. Assemblers have to put together the sequence data like a jigsaw puzzle that could have some pieces missing and some pieces containing errors. As a result, the assemblers can introduce error into the assembly by incorrectly pairing pieces together, or by incorrectly discarding pieces as mistakes or repeats. This can call into question the quality of the assemblies (Baker

2012). Nevertheless, genomes are continuing to be sequenced at a rapidly growing rate. Once the genome has been assembled the sequence can then be annotated to see which genes are present.

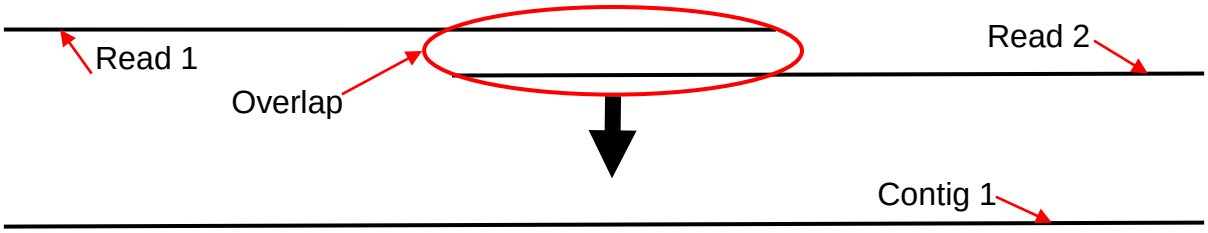


Figure 1. Example contig assembly.

One program that can be used to aid in the genome annotation is Artemis, which was used in this project. Artemis can annotate prokaryotic and eukaryotic genomes and works by allowing the user to find genes within the sequence (Carver *et al.* 2008). Another program for genome annotation that was used in this project is RAST. RAST, or Rapid Annotation using Subsystems Technology, allows for the annotation of archaeal and bacterial genomes only. Annotations using this program will give assessments as to the function of the gene and an initial metabolic reconstruction (Aziz *et al.* 2008).

From the information obtained by the annotation of the whole genome, researchers can learn about the organism's physiology. This information can also be used to find variations within species, such as humans and to develop maps of genetic variation. For example, in Pharmacogenomics researchers are trying to find genetic variations, using information obtained from DNA sequencing, associated with the effectiveness and toxicity of medicines (Henson, Tischler and Ning 2012). This could lead to more personalized medicine in which the selection of drugs can be optimized to

each individual's genome. Also, using the gene information from the annotation, the metabolic pathways used by the organism can begin to be reconstructed. Another use for the annotation information is to compare across species. This could be used to help determine if an organism of study is novel, having large variation of genes compared to closely related species, or what the organism should be classified as, with which genus or species does it show the greatest number of shared genes. The Genome Encyclopedia of Bacteria and Archaea (GEBA) is one such program that is phylogeny-driven to obtain the greatest amount of knowledge from already sequenced genomes and genomes yet to be sequenced (Wu *et al* 2009). One of the major obstacles that GEBA is attempting to remedy is that there are relatively few prokaryotic genomes that have been sequenced for several lineages and type strains because genome sequencing projects tend to focus on particular habitats and relatives of specific organisms. This lack of completed genomes is preventing progress towards classification of microorganisms based on their genomes (Klenk and Göker 2010).

Kaistella koreensis was first isolated from a freshwater stream in Korea and was found to be a rod shaped and yellow pigmented bacterium that is Gram negative and does not form spores (Kim *et al* 2004). This bacterium was proposed to be a novel species in a novel genus due to the large phylogenetic distance between it and its closest related strains. However, *Kaistella koreensis* was recently reclassified as *Chryseobacterium koreense* based on phenotypic similarities, including the polar lipid profile and the fatty acid profile it shared with another bacterium within the *Chryseobacterium* genus, *Chryseobacterium haifense* (Kämpfer *et al* 2009). This bacterium was chosen for sequencing because previous work with this bacterium in Dr.

Newman's lab suggests that *Kaistella koreensis* should not have been reclassified into the *Chryseobacterium* genus. The goal of this project was to compare the genome of *Kaistella koreensis* with those of other members of the *Chryseobacterium* genus that have had their genomes sequenced in order to dispute this reclassification.

One way that genomes can be compared is through Average Nucleotide Identity, or ANI. ANI functions by comparing orthologous genes between the two organisms, in this case comparing *Kaistella koreensis* to 6 other bacteria from closely related genera, in order to determine how similar the DNA sequences are to each other. An ANI of 95% or greater would indicate that the two organisms being compared belong to the same species (Auch *et al* 2010 and Richter and Rosselló-Móra 2009). However, currently there is no clear division between what ANI value separates different genera. An ANI of 80% or greater seems to indicate that the two organisms being compared belong to the same genus, while an ANI in the low 70's seems to indicate that the two organisms belong to different genera (Newman, Personal Communication). The relationship between organisms with an ANI in the mid to upper 70's, however, is unclear.

Methods

The first part of the project was to isolate the DNA of *Kaistella koreensis*, which was isolated as a part of the Molecular Biology class in the fall of 2011 using the Qiagen DNeasy Blood and Tissue Kit. The isolated DNA was then sequenced via Ion-Torrent (Life Technologies Corporation) by the Penn State Genomics Core Facility (http://www.huck.psu.edu/facilities/genomics-up) as a part of the GCAT-Seek consortium (Buonaccorsi *et al* 2011) and produced 657,812 reads with an average

length of 258 bases. The resulting reads were then assembled using NextGENe software (SoftGenetics, LLC) into contigs. NextGENe assembled 604,588 of the 657,812 reads into 3,058 contigs. The contigs had an average coverage of 25x with an average length of 1,254 bases, and an N50 length of 1,859 bases. The 3,058 contigs assembled using NextGENe were the starting point for this project

From here, the computer program Geneious version 5.6 (Biomatters) was used to further assemble the contigs into larger contiguous sequences. First the assembled contigs were imported into Geneious and assembled via a *de novo* assembly into the larger contiguous sequences, which were called supercontigs. The *de novo* assembly was done using the default settings including; a minimum overlap of 25 bases, a minimum overlap identity of 80%, a maximum gap size of 2 bases, and a maximum number of gaps per read of 15%. Sequences near the ends of the contigs have lower coverage and are more likely to have errors, which then prevents assembly into larger contigs. To address this problem, the more error prone sequences at the ends of the contigs were trimmed by various lengths using Geneious. In order to increase the efficiency of the supercontig assembly, several trimming lengths were tested until the ideal length was determined that would assemble the most contigs into the fewest supercontigs. Once the ideal trimming length was determined, the contigs were trimmed and assembled into supercontigs. Next, to decrease the ambiguities in the supercontig consensus sequence, bad contigs were removed from the assembly. Bad contigs were defined as contigs that were shorter, located completely within the sequences of larger contigs, and only added ambiguity to the sequence.

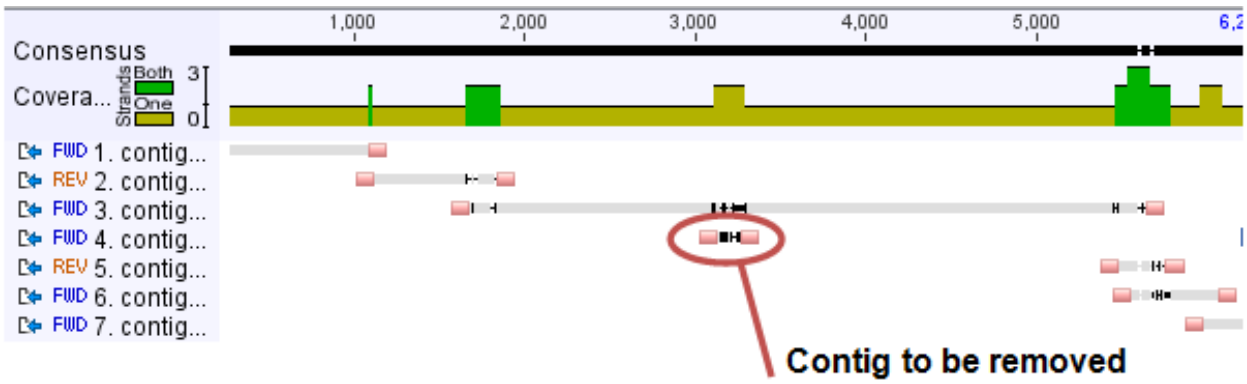


Figure 2. Bad contig removal.

Once the bad contigs were removed, the remaining ambiguities were manually corrected based on the assumption that bases closer to the ends are more likely to be wrong. This is because when the sequencing is done, the ends do not get read as frequently as the center of the reads. The supercontig consensus sequences, along with the sequences of the contigs that were not assembled, were then uploaded to the Rapid Annotation using Subsystem Technology website for initial annotation and comparison to reference genomes within the same family. Using the reference genomes, the supercontigs were then manually assembled into larger contiguous regions that were called supersupercontigs.

Figure 3. Correcting ambiguities in the consensus sequence.

RAST was used to align the genes of *Kaistella koreensis* in the same order that they are on the reference genome, *Lycomia vostokensis* also known as *Flavobacteriaceae* sp 3519-10, a closely related organism isolated from an ice core

above Lake Vostok, Antarctica. This organism was chosen as the reference genome because it is from the same family as *K. koreensis* and its genome has already been completed. The alignment could then be used to help determine which contigs should be assembled together. If two genes were next to each other on the genomes of closely related organisms and on two different contigs in the *Kaistella koreensis* genome, it would be hypothesized that they should be able to be assembled into a supersupercontig. The contig reordering tool in Artemis was used to move the hypothesized neighboring contigs next to each other. The sequences were then examined to look for overlapping segments in order to assemble the two contigs into one larger supersupercontig.

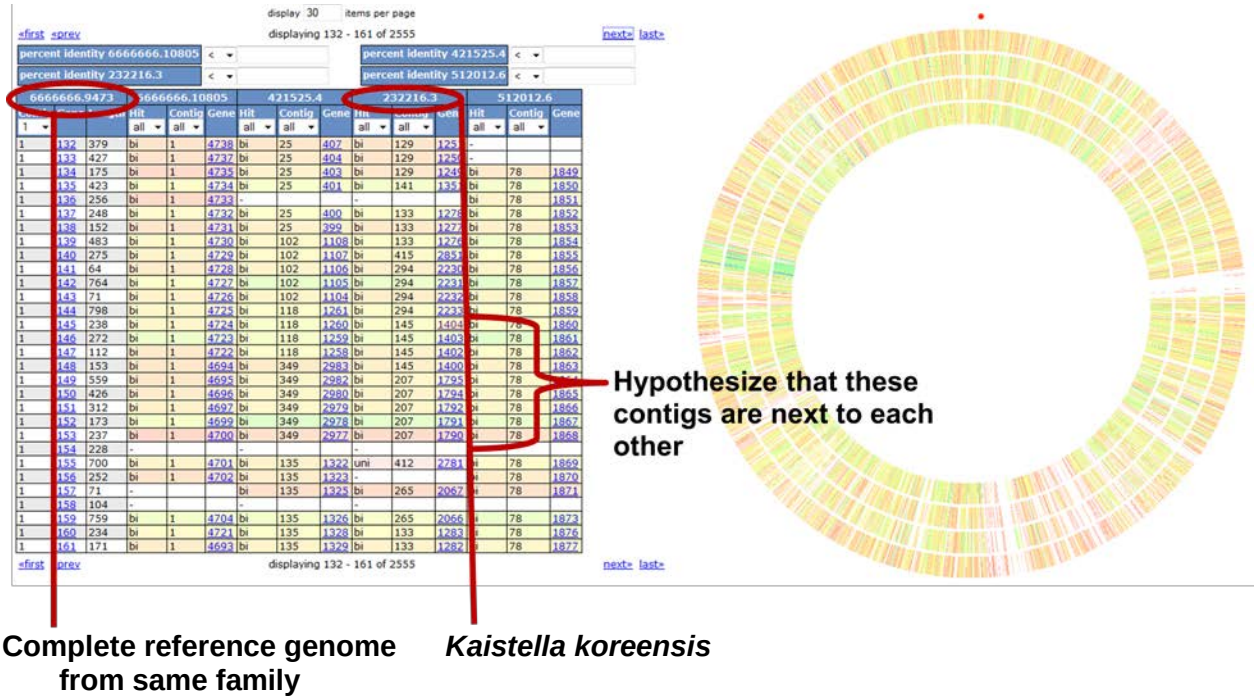


Figure 4. Sequence based comparison to reference genome using RAST.

However, not all of the supersupercontigs were able to be assembled this way due to possible errors in the sequences that prevent overlapping sequences from occurring. Also, there may be gaps between the sequences of some of the contigs,

therefore, there is no overlapping sequence and the gaps must be filled in order to assemble them. After manually assembling several of the supersupercontigs, the sequences were again uploaded into a new version of Geneious, version 6.1 (Biomatters), and another *de novo* assembly was done, again using the same default settings as before. This final assembly was then uploaded again to RAST for a final annotation of the genome and analysis via Average Nucleotide Identity.

In order to calculate the ANI, a sequence based comparison was done using *Kaistella koreensis*, as well as six other organisms within the same family; *Lycomia zaccaria, Lycomia vostokensis, Chryseobacterium gleum, Chryseobacterium haifense, Chryseobacterium* sp. CF314, and *Riemerella anatipestifer*. The sequence based comparison was done using each organism as the reference organism, and comparing the other six organisms to it. The ANI was then computed for each comparison using a Macros tool in Microsoft Excel. ANI works by comparing each individual gene and determining the average percent identity between them, and then weighting the results based on the size of the genes so that larger genes contribute more to the overall ANI than smaller genes.

In order to further compare *Kaistella koreensis* to other closely related organisms, a Biolog GenIII test as well as an API ZYM test was performed. The Biolog GenIII tests 96 different possible nutrients or inhibitory conditions, while the API ZYM test for the presence of a 20 different hydrolytic enzymes (Microbiology Laboratory Manual 2013). The Biolog GenIII test was performed by streaking cells from *Kaistella koreensis* onto a plate containing the Biolog Universal Growth medium and 5% sheep's blood (BUG + Blood). Cells from the BUG + Blood plate were then suspended in an

inoculating medium to a density between 90 and 98% transmittance on the Turbidometer. 100 µL of the inoculating medium was then added to each of the 96 wells on the GenIII plate, and the plate was then incubated at 33°C in the Omnilog instrument. The Omnilog scanned the plate every 15 minutes for 36 hours to determine in which wells the organism is actively metabolizing (Microbiology Laboratory Manual 2013). The API ZYM test was performed by culturing the bacteria in a liquid medium overnight, then centrifuging 1 mL of the culture to pellet the cells. After centrifugation, the supernatant was removed and the cells were resuspended in 0.25 mL of a sterile 0.85% NaCl solution. The resuspended cells were then added to 3 mL of 0.85% NaCl until a turbidity of 5-6 McFarland was reached. 65 µL of this suspension was then added to each of the 20 cupules on the API ZYM test strip and the strip was the incubated at 30°C for 4-8 hours. A drop of ZYM reagents A and B were then added to each cupule and the strip incubated for 5 minutes (Microbiology Laboratory Manual 2013).

A third test that was performed to compare *Kaistella koreensis* to closely related organisms was a Fatty Acid Methyl Ester analysis (FAME). The FAME analysis is done to determine the lipid composition of the plasma membrane. In order to do the FAME analysis, *K. koreensis* was first grown on R2A at 30°C. Next, 3-5 mg of log phase cells were collected and smeared into a Gas Chromatography vial (Biology 444 Laboratory Manual 2013). Then, 0.25 mL of 3.75M sodium hydroxide was added to the vial, in order to hydrolyze the ester linkages of the phospholipids and release the fatty acids from the phospholipids, and the vial was placed in a 100°C heating block for a total of 30 minutes. After letting the vial cool, 0.5 mL of a solution of 325 ml hydrochloric acid

and 275 ml methanol was added in order to methylate the fatty acids, which creates the methyl esters, and the vial was then placed in an 80°C heating block for 10 minutes (Biology 444 Laboratory Manual 2013). The vial was then removed from the heat and allowed to cool before 0.3 ml of a solution of 200 ml hexane and 200 ml methyl tert-butyl ether was added in order to neutralize the solution and remove any unesterified fatty acids. The hexane also served to extract the fatty acid methyl esters from the aqueous phase and create two layers (Biology 444 Laboratory Manual 2013). The upper organic layer, containing the fatty acid methyl esters, was extracted and analyzed on the Gas Chromatograph. The results of the FAME test for *K. koreensis* were then compared to those of *Chryseobacterium haifense, Lycomia vostokensis, Lycomia zaccaria, Chryseobacterium jejuense, Chryseobacterium shigense,* and *Chryseobacterium soli*.

Results

The Ion-Torrent sequencing resulted in 657,812 reads with an average read length of 258 bases and an average coverage of 25x. The initial assembly using the NextGENe assembly program assembled the reads into 3,058 contigs. After trying trimming lengths of 15, 35, 50, and 60 bases trimmed from each end, the ideal trimming length was determined to be 50 bases trimmed from each end, as seen in Figure 5 below. The Geneious assembly using the trimmed NextGENe contigs resulted in 482 supercontigs assembled, and 367 of the NextGENe contigs were not assembled into supercontigs. The supercontigs had a length range from 187 bases to 46,968 bases, with an average length of 6,264 bases and an N50 length of 9,724 bases. The N50 length is a weighted median such that 50% of all the bases in all of the sequences are contained within contigs that are equal to or greater than the length of the N50 value.

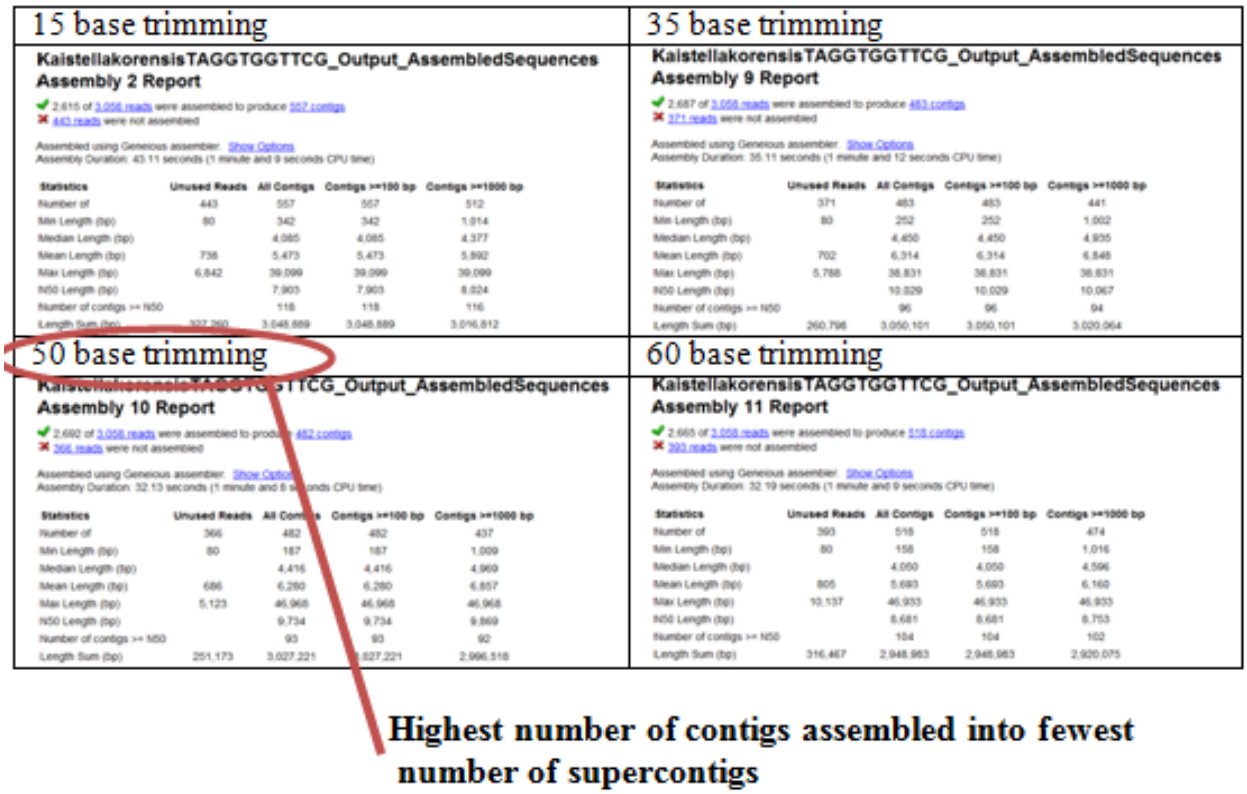


Figure 5. Trials of four different trimming lengths using Geneious in order to determine an ideal length.

From the initial annotation using RAST, the genome size of *Kaistella koreensis* was 3,225,452 base pairs. Within the genome there were 3,685 genes, 1,785 of which coded for known products. The remaining genes coded for hypothetical proteins. As a result of the supersupercontig assembly, 26 supersupercontigs were assembled using 71 contigs and supercontigs. This brought the total number of separate sequences from 691 that were uploaded to RAST, down to 620 sequences. The final assembly using the new version of Geneious, version 6.1, resulted in 83 of the contigs being assembled into 41 larger contiguous sequences, and 537 contigs were not further assembled. The average length of these contigs was 9,381 bases, and the N50 length was 18,079 bases. This gave a final assembly consisting of 578 contigs. In the final annotation using RAST, the genome size of *Kaistella koreensis* is 3,199,137 base pairs. 3,677 genes were identified, of which 40 are RNA genes and 3,637 are protein coding

genes. Of the 3,677 genes, 1,784 of the genes code for known products, while the other 1,893 genes code for hypothetical proteins. A screenshot of the RAST output for the final annotation can be seen if Figure 6 below. The list of subsystems found by RAST and the genes present can be found in Figure 7 below.

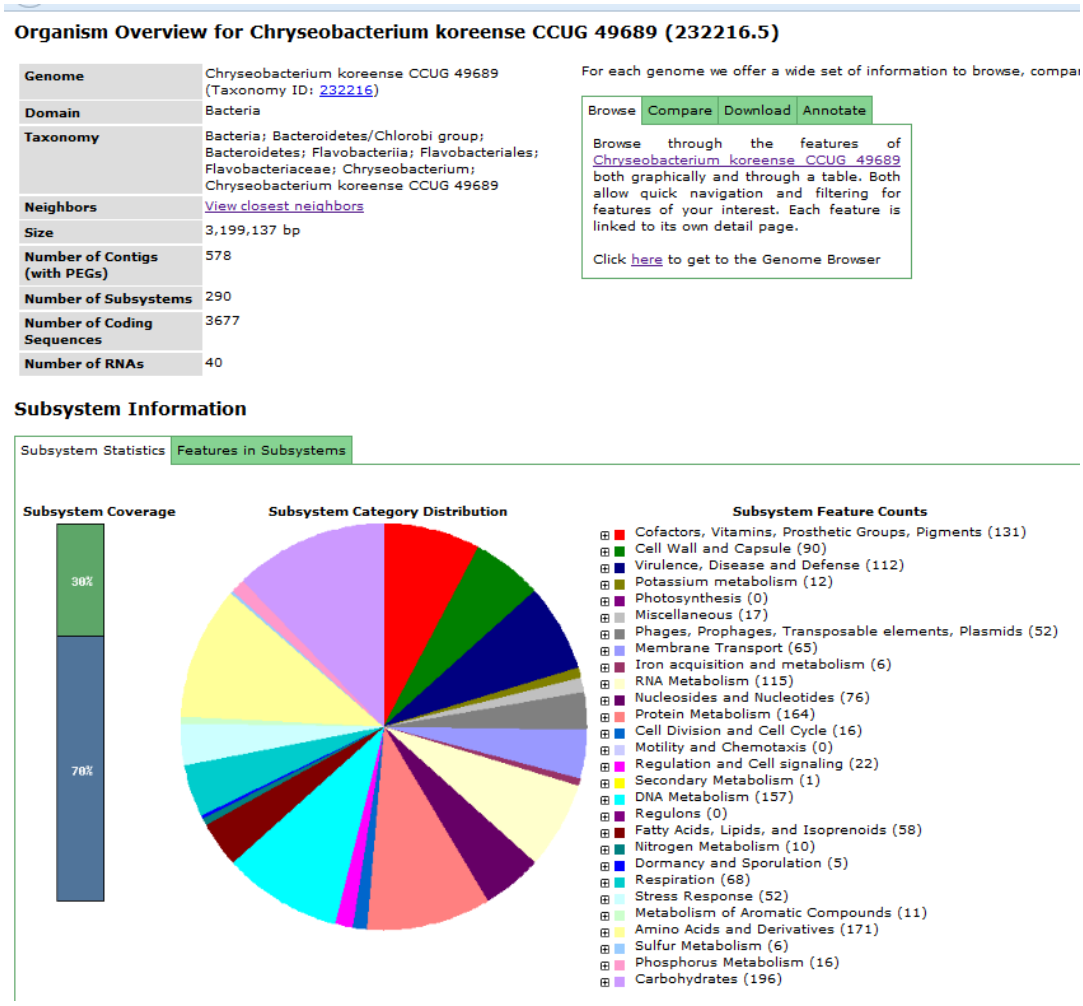


Figure 6. Screenshot of the final annotation using RAST.

- Category
  - Subcategory
    - Subsystem (number of genes)
- Cofactors, Vitamins, Prosthetic Groups, Pigments (131)
  - Biotin (1)
    - Biotin biosynthesis Experimental (1)
  - Cofactors, Vitamins, Prosthetic Groups, Pigments - no subcategory (0)
  - Quinone cofactors (7)
    - Menaquinone and Phylloquinone Biosynthesis (7)
  - Tetrapyrroles (11)
    - Heme and Siroheme Biosynthesis (11)
  - Riboflavin, FMN, FAD (31)
    - Riboflavin, FMN and FAD metabolism in plants (15)
    - riboflavin to FAD (5)
    - Riboflavin, FMN and FAD metabolism (11)
  - Fe-S clusters (0)
  - Mycofactocin (0)
  - Pyridoxine (8)
    - Pyridoxin (Vitamin B6) Biosynthesis (8)
  - NAD and NADP (11)
    - NAD and NADP cofactor biosynthesis global (11)
  - Coenzyme B (0)
  - Folate and pterines (45)
    - Folate biosynthesis cluster (11)
    - Folate Biosynthesis (17)
    - 5-FCL-like protein (17)
  - Lipoic acid (4)
    - Lipoic acid metabolism (4)
  - Coenzyme F420 (0)
  - Coenzyme M (0)
  - Coenzyme A (13)
    - Coenzyme A Biosynthesis cluster (5)
    - Coenzyme A Biosynthesis (8)
- Cell Wall and Capsule (90)
  - Capsular and extracellular polysacchrides (21)
    - dTDP-rhamnose synthesis (7)
    - Rhamnose containing glycans (9)
    - Sialic Acid Metabolism (5)
  - Gram-Negative cell wall components (15)
    - KDO2-Lipid A biosynthesis (14)
    - Lipoprotein sorting system (1)
  - Cell Wall and Capsule - no subcategory (53)
    - Murein Hydrolases (5)
    - Peptidoglycan Biosynthesis (27)
    - UDP-N-acetylmuramate from Fructose-6-phosphate Biosynthesis (5)
    - YjeE (2)
    - Recycling of Peptidoglycan Amino Sugars (1)
    - Recycling of Peptidoglycan Amino Acids (3)
    - Peptidoglycan biosynthesis--gjo (10)
  - Gram-Positive cell wall components (1)
    - Teichuronic acid biosynthesis (1)
  - Cell wall of Mycobacteria (0)
- Virulence, Disease and Defense (112)

- Adhesion (0)
- Toxins and superantigens (0)
- Bacteriocins, ribosomally synthesized antibacterial peptides (0)
- Resistance to antibiotics and toxic compounds (96)
  - Copper homeostasis (6)
  - Cobalt-zinc-cadmium resistance (69)
  - Resistance to fluoroquinolones (4)
  - Copper homeostasis: copper tolerance (1)
  - Arsenic resistance (6)
  - Multidrug Resistance, Tripartite Systems Found in Gram Negative Bacteria (5)
  - Beta-lactamase (4)
  - Zinc resistance (1)
- Virulence, Disease and Defense - no subcategory (0)
- Detection (0)
- Invasion and intracellular resistance (16)
  - Mycobacterium virulence operon involved in protein synthesis (SSU ribosomal proteins) (5)
  - Mycobacterium virulence operon involved in DNA transcription (2)
  - Mycobacterium virulence operon possibly involved in quinolinate biosynthesis (3)
  - Mycobacterium virulence operon involved in lipid metabolism (2)
  - Mycobacterium virulence operon involved in protein synthesis (LSU ribosomal proteins) (4)

Potassium metabolism (12)

- Potassium metabolism - no subcategory (12)
  - Potassium homeostasis (12)

Photosynthesis (0)

- Light-harvesting complexes (0)
- Photosynthesis - no subcategory (0)
- Electron transport and photophosphorylation (0)

Miscellaneous (17)

- Conversion of Succinyl-CoA to Propionyl-CoA (0)
- Plant-Prokaryote DOE project (12)
  - Iron-sulfur cluster assembly (12)
- Miscellaneous - no subcategory (5)
  - Phosphoglycerate mutase protein family (1)
  - DedA family of inner membrane proteins (1)
  - Muconate lactonizing enzyme family (2)
  - Broadly distributed proteins not in subsystems (1)

Phages, Prophages, Transposable elements, Plasmids (52)

- Phage family-specific subsystems (0)
- Transposable elements (51)
  - Conjugative transposon, Bacteroidales (51)
- Phages, Prophages (0)
- Phages, Prophages, Transposable elements, Plasmids - no subcategory (1)
  - Integrons (1)
- Pathogenicity islands (0)
- Gene Transfer Agent (GTA) (0)
- Plasmid related functions (0)

Membrane Transport (65)

- Protein secretion system, Type II (0)

- ABC transporters (2)
  - ABC transporter oligopeptide (TC 3.A.1.5.1) (2)
- Protein secretion system, Type VII (Chaperone/Usher pathway, CU) (0)
- Protein translocation across cytoplasmic membrane (4)
  - Bacterial signal recognition particle (SRP) (2)
  - Twin-arginine translocation system (2)
- Protein secretion system, Type V (0)
- Protein secretion system, Type I (0)
- Cation transporters (11)
  - Magnesium transport (5)
  - Copper Transport System (5)
  - Copper transport and blue copper proteins (1)
- Protein secretion system, Type III (0)
- Uni- Sym- and Antiporters (6)
  - Proton-dependent Peptide Transporters (3)
  - NhaA, NhaD and Sodium-dependent phosphate transporters (3)
- Membrane Transport - no subcategory (42)
  - Ton and Tol transport systems (42)
- TRAP transporters (0)
- Sugar Phosphotransferase Systems, PTS (0)
- Protein secretion system, Type VI (0)
- Protein secretion system, Type VIII (Extracellular nucleation/precipitation pathway, ENP) (0)
- Protein and nucleoprotein secretion system, Type IV (0)

Iron acquisition and metabolism (6)

- Siderophores (0)
- Iron acquisition and metabolism - no subcategory (6)
  - Hemin transport system (6)
- Iron transpot (0)

RNA Metabolism (115)

- RNA processing and modification (93)
  - RNA pseudouridine syntheses (6)
  - tRNA nucleotidyltransferase (1)
  - Methylthiotransferases (3)
  - Ribonucleases in Bacillus (2)
  - tRNAmodification position 34 (7)
  - RNA processing and degradation, bacterial (4)
  - RNA methylation (7)
  - ATP-dependent RNA helicases, bacterial (2)
  - 16S rRNA modification within P site of ribosome (7)
  - tRNA modification Bacteria (31)
  - mnm5U34 biosynthesis bacteria (5)
  - Queuosine-Archaeosine Biosynthesis (8)
  - Ribonuclease H (2)
  - tRNA processing (8)
- Transcription (19)
  - Transcription initiation, bacterial sigma factors (4)
  - RNA polymerase bacterial (3)
  - Transcription factors bacterial (9)
  - Rrf2 family transcriptional regulators (3)
- RNA Metabolism - no subcategory (3)
  - Group II intron-associated genes (3)

Nucleosides and Nucleotides (76)

- Pyrimidines (27)

- pyrimidine conversions (17)
- De Novo Pyrimidine Synthesis (10)

Purines (36)
- De Novo Purine Biosynthesis (15)
- Purine conversions (21)

Nucleosides and Nucleotides - no subcategory (6)
- Ribonucleotide reduction (3)
- Adenosyl nucleosidases (3)

Detoxification (7)
- Nucleoside triphosphate pyrophosphohydrolase MazG (1)
- Nudix proteins (nucleoside triphosphate hydrolases) (2)
- Housecleaning nucleoside triphosphate pyrophosphatases (4)

Protein Metabolism (164)

Protein folding (10)
- GroEL GroES (2)
- Protein chaperones (6)
- Periplasmic disulfide interchange (1)
- Peptidyl-prolyl cis-trans isomerase (1)

Selenoproteins (2)
- Selenoprotein O (2)

Protein biosynthesis (100)
- tRNA aminoacylation, Val (1)
- tRNA aminoacylation, Met (2)
- tRNA aminoacylation, Ile (1)
- tRNA aminoacylation, Arg (1)
- Ribosome SSU bacterial (23)
- tRNA aminoacylation, Gly (1)
- Ribosome activity modulation (1)
- tRNA aminoacylation, Ala (1)
- tRNA aminoacylation, Trp (1)
- Ribosome LSU bacterial (34)
- Programmed frameshift (2)
- tRNA aminoacylation, Cys (1)
- Translation termination factors bacterial (9)
- tRNA aminoacylation, His (1)
- tRNA aminoacylation, Asp and Asn (2)
- Translation elongation factors bacterial (7)
- tRNA aminoacylation, Lys (2)
- tRNA aminoacylation, Thr (1)
- Translation elongation factor G family (2)
- tRNA aminoacylation, Glu and Gln (2)
- tRNA aminoacylation, Ser (1)
- tRNA aminoacylation, Tyr (1)
- tRNA aminoacylation, Leu (1)
- tRNA aminoacylation, Phe (2)

Protein processing and modification (27)
- Protein-L-isoaspartate O-methyltransferase (1)
- Ribosomal protein S12p Asp methylthiotransferase (2)
- Peptide methionine sulfoxide reductase (2)
- N-linked Glycosylation in Bacteria (7)
- Lipoprotein Biosynthesis (3)
- Modification of eukaryotic initiation factor 5A (2)
- Signal peptidase (3)
- G3E family of P-loop GTPases (metallocenter biosynthesis) (7)

Protein degradation (25)

Aminopeptidases (EC 3.4.11.-) (1)
Protein degradation (8)
Metallocarboxypeptidases (EC 3.4.17.-) (3)
Dipeptidases (EC 3.4.13.-) (2)
Serine endopeptidase (EC 3.4.21.-) (1)
Proteolysis in bacteria, ATP-dependent (8)
Omega peptidases (EC 3.4.19.-) (2)
Cell Division and Cell Cycle (16)
Checkpoint control (0)
Cell Division and Cell Cycle - no subcategory (16)
Bacterial Cytoskeleton (16)

Motility and Chemotaxis (0)

Magnetotaxis (0)
Motility and Chemotaxis - no subcategory (0)
Flagellar motility in Prokaryota (0)
Social motility and nonflagellar swimming in bacteria (0)

Regulation and Cell signaling (22)

Regulation and Cell signaling - no subcategory (12)
cAMP signaling in bacteria (6)
LysR-family proteins in Salmonella enterica Typhimurium (1)
LysR-family proteins in Escherichia coli (1)
Stringent Response, (p)ppGpp metabolism (4)
Signal transduction in Eukaryotes (0)
Quorum sensing and biofilm formation (0)
Proteolytic pathway (0)
Regulation of virulence (0)
Programmed Cell Death and Toxin-antitoxin Systems (10)
Toxin-antitoxin replicon stabilization systems (10)

Secondary Metabolism (1)

Secondary Metabolism - no subcategory (0)
Lipid-derived mediators (0)
Plant Octadecanoids (0)
Bacterial cytostatics, differentiation factors and antibiotics (0)
Biosynthesis of phenylpropanoids (0)
Hydrocarbons (0)
Aromatic amino acids and derivatives (0)
UV-absorbing secondary metabolites (0)
Plant Alkaloids (1)
Alkaloid biosynthesis from L-lysine (1)
Biologically active compounds in metazoan cell defence and differentiation (0)
Plant Hormones (0)

DNA Metabolism (157)

DNA repair (51)
Uracil-DNA glycosylase (1)
DNA repair, bacterial MutL-MutS system (3)
DNA repair, UvrABC system (4)
DNA repair, bacterial photolyase (1)
DNA repair system including RecA, MutS and a hypothetical protein (2)
DNA repair, bacterial (18)
DNA repair, bacterial RecFOR pathway (12)
DNA Repair Base Excision (8)

- DNA repair, bacterial UvrD and related helicases (2)
- CRISPs (2)
  - CRISPRs (2)
- DNA Metabolism - no subcategory (48)
  - Type I Restriction-Modification (22)
  - Restriction-Modification System (22)
  - YcfH (2)
  - DNA structural proteins, bacterial (2)
- DNA replication (43)
  - DNA topoisomerases, Type I, ATP-independent (8)
  - DNA-replication (31)
  - DNA topoisomerases, Type II, ATP-dependent (4)
- DNA recombination (3)
  - RuvABC plus a hypothetical (3)
- DNA uptake, competence (10)
  - DNA processing cluster (10)

Regulons (0)

- Atomic Regulons (0)

Fatty Acids, Lipids, and Isoprenoids (58)

- Phospholipids (12)
  - Glycerolipid and Glycerophospholipid Metabolism in Bacteria (12)
- Triacylglycerols (0)
- Fatty acids (25)
  - Fatty Acid Biosynthesis FASII (16)
  - Fatty acid metabolism cluster (9)
- Fatty Acids, Lipids, and Isoprenoids - no subcategory (13)
  - Polyhydroxybutyrate metabolism (13)
- Isoprenoids (8)
  - Myxoxanthophyll biosynthesis in Cyanobacteria (1)
  - Mevalonate Branch of Isoprenoid Biosynthesis (7)

Nitrogen Metabolism (10)

- Nitrogen Metabolism - no subcategory (4)
  - Nitrosative stress (2)
  - Ammonia assimilation (2)
- Denitrification (6)
  - Denitrification (6)

Dormancy and Sporulation (5)

- Spore DNA protection (0)
- Dormancy and Sporulation - no subcategory (5)
  - Spore Core Dehydration (1)
  - Persister Cells (3)
  - Sporulation-associated proteins with broader functions (1)

Respiration (68)

- Biotin (1)
- ATP synthases (9)
  - F0F1-type ATP synthase (9)
- Electron accepting reactions (11)
  - Terminal cytochrome C oxidases (9)
  - Anaerobic respiratory reductases (2)
- Electron donating reactions (39)

- Respiratory Complex I (17)
- Respiratory dehydrogenases 1 (2)
- Succinate dehydrogenase (3)
- NADH ubiquinone oxidoreductase (17)

Reverse electron transport (0)
Respiration - no subcategory (9)
- Biogenesis of cbb3-type cytochrome c oxidases (4)
- Biogenesis of c-type cytochromes (2)
- Soluble cytochromes and functionally related electron carriers (3)

Sodium Ion-Coupled Energetics (0)

Stress Response (52)

Osmotic stress (5)
- Osmoregulation (5)

Dessication stress (0)
Acid stress (0)
Oxidative stress (24)
- Protection from Reactive Oxygen Species (3)
- Oxidative stress (15)
- Glutathione: Non-redox reactions (1)
- Redox-dependent regulation of nucleus processes (4)
- Glutathione: Redox cycle (1)

Cold shock (0)
Heat shock (12)
- Heat shock dnaK gene cluster extended (12)

Detoxification (9)
- D-tyrosyl-tRNA(Tyr) deacylase (1)
- Uptake of selenate and selenite (1)

Stress Response - no subcategory (4)
- SigmaB stress responce regulation (1)
- Dimethylarginine metabolism (2)
- Hfl operon (1)

Periplasmic Stress (5)
- Periplasmic Stress Response (5)

Metabolism of Aromatic Compounds (11)
Peripheral pathways for catabolism of aromatic compounds (1)
- Quinate degradation (1)

Anaerobic degradation of aromatic compounds (0)
Metabolism of central aromatic intermediates (8)
- Catechol branch of beta-ketoadipate pathway (3)
- Salicylate and gentisate catabolism (2)
- Homogentisate pathway of aromatic compound degradation (3)

Metabolism of Aromatic Compounds - no subcategory (2)
- Gentisare degradation (1)
- Aromatic Amin Catabolism (1)

Amino Acids and Derivatives (171)

Glutamine, glutamate, aspartate, asparagine; ammonia assimilation (19)
- Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis (13)
- Glutamate dehydrogenases (1)
- Glutamine synthetases (1)
- Glutamate and Aspartate uptake in Bacteria (4)

Histidine Metabolism (4)
- Histidine Degradation (4)

Arginine; urea cycle, polyamines (12)
- Polyamine Metabolism (2)

- Arginine and Ornithine Degradation (5)
- Cyanophycin Metabolism (5)

Lysine, threonine, methionine, and cysteine (28)
- Threonine and Homoserine Biosynthesis (6)
- Threonine degradation (4)
- Lysine Biosynthesis DAP Pathway, GJO scratch (9)
- Lysine Biosynthesis DAP Pathway (9)

Amino Acids and Derivatives - no subcategory (0)

Branched-chain amino acids (44)
- Isoleucine degradation (17)
- Leucine Degradation and HMG-CoA Metabolism (15)
- Valine degradation (12)

Polyamines (0)

Aromatic amino acids and derivatives (26)
- Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate) (7)
- Chorismate Synthesis (9)
- Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate and more. (8)
- Phenylalanine and Tyrosine Branches from Chorismate (2)

Proline and 4-hydroxyproline (1)
- Proline, 4-hydroxyproline uptake and utilization (1)

Alanine, serine, and glycine (37)
- Glycine Biosynthesis (6)
- Alanine biosynthesis (5)
- Serine Biosynthesis (6)
- Glycine cleavage system (4)
- Glycine and Serine Utilization (16)

Sulfur Metabolism (6)

Inorganic sulfur assimilation (0)

Sulfur Metabolism - no subcategory (6)
- Thioredoxin-disulfide reductase (6)

Organic sulfur assimilation (0)

Phosphorus Metabolism (16)

Phosphorus Metabolism - no subcategory (16)
- Phosphate metabolism (11)
- Polyphosphate (2)
- Alkylphosphonate utilization (3)

Carbohydrates (196)

Central carbohydrate metabolism (87)
- Methylglyoxal Metabolism (7)
- Pyruvate metabolism II: acetyl-CoA, acetogenesis from pyruvate (10)
- Pyruvate Alanine Serine Interconversions (4)
- Glycolysis and Gluconeogenesis (14)
- Entner-Doudoroff Pathway (11)
- Dehydrogenase complexes (13)
- TCA Cycle (15)
- Pentose phosphate pathway (8)
- Pyruvate metabolism I: anaplerotic reactions, PEP (5)

Aminosugars (0)

Di- and oligosaccharides (6)
- Maltose and Maltodextrin Utilization (6)

Glycoside hydrolases (0)

One-carbon Metabolism (37)
- Serine-glyoxylate cycle (31)
- One-carbon metabolism by tetrahydropterines (6)

Organic acids (0)
Fermentation (23)
- Butanol Biosynthesis (10)
- Acetyl-CoA fermentation to Butyrate (13)

CO2 fixation (0)
Sugar alcohols (0)
Carbohydrates - no subcategory (0)
Polysaccharides (15)
- Glycogen metabolism (7)
- Cellulosome (8)

Monosaccharides (28)
- Mannose Metabolism (6)
- D-ribose utilization (2)
- Deoxyribose and Deoxynucleoside Catabolism (6)
- D-Galacturonate and D-Glucuronate Utilization (14)

Figure 7. Table showing subsystems in which RAST identified genes, classified first by category, second by subcategory, and third by subsystem with the number of genes in each category, subcategory, and subsystem in parentheses.

The results of the Biolog GenIII test for *Kaistella koreensis* can be seen in Figure 8 below, positive results are indicated by the presence of a purple dye in the well. *Kaistella koreensis* only tested positive for 9 of the tests and weakly positive for 5 others. *K. koreensis* tested positive for the use of the nutrients; gelatin, glutamic acid, tellurite, acetoacetic acid, and Na-butyrate and was not inhibited by the inhibitory conditions of pH at 6, 1% NaCl, and 1% Na-lactate. *K. koreensis* tested weakly positive for the usage glycyl-L-proline, L-aspartic acid, and acetic acid, and also tested weakly positive for the reduction of tetrazolium violet and tetrazolium blue. The results of the Biolog GenIII test for *K. koreensis* were then compared to the results of Biolog tests for *Lycomia zaccaria*, *Lycomia vostokensis*, and *Chryseobacterium haifense* as well as three representatives of the *Chryseobacterium* genus, *Chryseobacterium jejuense, Chryseobacterium shigense,* and *Chryseobacterium soli* in order to look for distinguishing characteristics of the genus *Kaistella* that would differentiate it from the genera *Lycomia* and *Chryseobacterium*. The results of this comparison can be seen in

Figure 9. *K. koreensis* differed from all of the other organisms in that it only very weakly utilized maltose and tween, and that it was not inhibited at all by 1%Na-lactate. *K. koreensis* was also different in that it was not inhibited by the presence of tellurite or Na-butyrate; however, it was inhibited by aztreonam, unlike the rest of the organisms being compared. What is also notable by looking at the results for *K. koreensis* taken from the same plate at two different times is that in some cases *K. koreensis* was negative for a test after 22 hours, but positive for the same test after 41 hours. *K. koreensis* was not able to utilize, or utilized only weakly, glycyl-L-proline, aspartic acid, galacturonic acid, acetic acid, and Na-butyrate after 22 hours, but was able to utilize all of these after 41 hours.

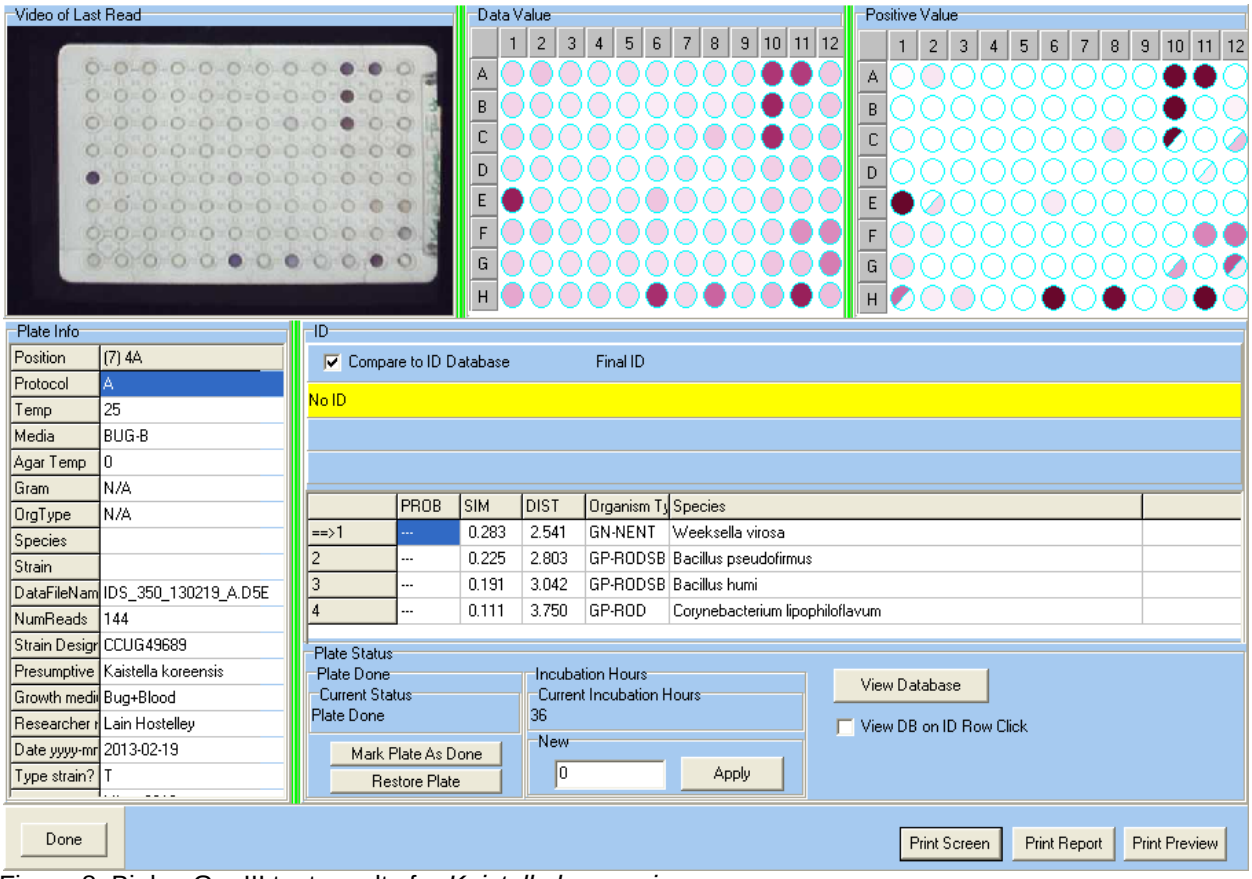


Figure 8. Biolog GenIII test results for *Kaistella koreensis*.

	abbrev	Kkor	Kkor	Kkor	Lzac	Lvos	Lhaif	Cjeju	Cshig	Csoli			abbrev	Kkor	Kkor	Kkor	Lzac	Lvos	Lhaif	Cjeju	Cshig	Csoli
	strain	CCUG 49689	CCUG 49689	CCUG 49689	JJC	3915-10	DSM 19056	DSM 19399	DSM 17126	DSM 19298			strain	CCUG 49689	CCUG 49689	CCUG 49689	JJC	3915-10	DSM 19056	DSM 19399	DSM 17126	DSM 19298
	<-- test Genus species	Kaistella koreensis 22 hr	Kaistella koreensis 41hr	Kaistella koreensis	Lycomia zaccaria	Lycomia vostokensis	Lycomia haifense	Chryseobacterium jejuense	Chryseobacterium shigense	Chryseobacterium soli			<-- test Genus species	Kaistella koreensis 22 hr	Kaistella koreensis 41hr	Kaistella koreensis	Lycomia zaccaria	Lycomia vostokensis	Lycomia haifense	Chryseobacterium jejuense	Chryseobacterium shigense	Chryseobacterium soli
A01	neg control	25	14	18.7	22	25.5	22.6	14.8	22.3	17.25		E01	gelatin	100	100	100.0	100.0	45	92.2	99.7	79.5	82.25
A02	dextrin	41	38	33.7	50.8	36.5	28.4	43.3	67.3	60.25		E02	glycyl-L-proline	16	100	42.3	65.5	93.5	95.2	84.5	87.0	45.5
A03	D-maltose	14	26	18.3	70.5	89.5	89.2	69.5	93.5	62		E03	L-alanine	9	15	10.0	10.75	43	6.8	49.7	17.5	6.25
A04	D-trehalose	13	12	11.3	11.25	**50.5**	6.8	63.8	83.8	35.75		E04	L-arginine	11	18	12.3	16	60	8.6	35.7	18.8	7.25
A05	D-cellobiose	12	16	13.3	**13.25**	**64.5**	12.8	6.0	10.0	7.25		E05	L-aspartic acid	39	100	48.7	68.0	21	39.4	72.0	82.3	13.5
A06	gentiobiose	10	12	10.7	**76.25**	66	9.6	86.3	94.5	81		E06	L-glutamic acid	100	100	73.7	100.0	69.0	98.4	77.5	99.0	65.25
A07	sucrose	9	9	9.7	8	29.5	37.2	6.0	86.8	28.75		E07	L-histidine	9	24	14.0	20	16.5	14.8	28.7	7.0	7.75
A08	D-turanose	12	22	14.7	7.75	4	4.6	3.5	7.3	5		E08	L-pyroglutamic acid	9	16	11.0	21	36	9.2	29.2	8.8	5.75
A09	stachyose	11	14	12.7	9.5	**23**	4.6	2.3	5.8	6		E09	L-serine	9	12	9.0	14.8	53	9	67.5	54.3	20
A10	pos control	99	99	98.7	99.8	100.0	99.0	94.8	98.8	99		E10	lincomycin	10	24	15.3	13.25	11.5	10.8	91.8	11.3	12
A11	pH 6	100	99	98.3	99.3	**82**	98.6	93.8	98.3	98.75		E11	guanidine HCl	13	21	14.3	12.25	11.5	11.8	42.0	10.5	13
A12	pH 5	17	11	14.7	16.25	8.5	16.8	42.0	20.0	42.25		E12	niaproof 4	14	13	13.3	16	9.5	15.6	11.2	17.5	20.5
B01	D-raffinose	17	10	14.3	18.25	**15**	21.4	12.0	35.0	22.5		F01	pectin	17	29	22.7	100.0	75.5	97.0	52.8	98.8	92
B02	α-D-lactose	13	10	12.3	8.5	**20.5**	**16.8**	3.8	11.3	8		F02	D-galacturonic acid	20	66	35.3	100.0	100.0	27.4	67.0	99.5	98.75
B03	D-melibiose	9	14	11.3	8	35.5	**4.6**	1.8	7.8	5		F03	L-galacturonic acid lactone	11	14	12.7	9.5	1	5.8	6.8	8.0	8.75
B04	β-methyl-D-glucoside	9	12	11.3	7.75	**32.5**	**4.6**	1.7	5.8	4.5		F04	D-gluconic acid	10	16	11.7	22.5	57.5	16.2	16.8	12.8	8
B05	D-salicin	13	12	10.7	7.25	**7**	4.8	2.3	53.8	6.25		F05	D-glucuronic acidi	9	29	17.7	18.75	93.5	22.8	41.3	12.0	10.25
B06	N-acetyl-D-glucosamine	9	12	9.0	8.25	**54.5**	4.6	5.8	6.0	4.5		F06	glucuronamide	13	50	25.3	25.5	93.0	23.2	42.3	10.8	12.5
B07	N-acetyl-β-D-mannosamine	9	17	11.7	10.5	**48.5**	4.6	11.0	7.3	5		F07	mucic acid	9	14	10.0	19.25	5.5	17	32.8	13.8	8.75
B08	N-acetyl-D-galactosamine	9	14	9.7	11.25	47.5	7	5.2	7.3	5		F08	quinic acid	9	15	10.7	13.25	77	10.8	14.8	6.8	6
B09	N-acetyl neuraminic acid	9	11	9.7	8	10	4.8	3.0	5.5	5.75		F09	D-saccharic acid	9	14	10.7	17.5	31.5	9.6	14.5	7.3	6.5
B10	1% NaCl	99	100	99.7	80.3	**89**	58.0	86.5	61.5	58		F10	vancomycin	10	15	11.7	11.5	10	**10.8**	9.7	11.5	12.25
B11	4% NaCl	23	11	15.7	12.75	8	20.8	11.0	14.5	18.25		F11	tetrazolium violet	43	50	46.0	39.25	52	**45.6**	99.7	60.0	67.5
B12	8% NaCl	36	13	22.7	29	9	22.8	8.2	21.3	27.25		F12	tetrazolium blue	56	65	57.0	43.5	32	92.8	100.0	100.0	100
C01	α-D-glucose	18	24	19.7	49.25	**29**	43.6	67.7	89.0	62.75		G01	p-hydroxy-phenylacetic acid	14	12	16.3	17.5	4	13.2	9.8	13.8	15.5
C02	D-mannose	10	19	14.7	7.25	60.5	22.6	66.3	85.5	63.75		G02	methyl pyruvate	10	15	11.7	32.0	1	7.2	9.7	49.0	24
C03	D-fructose	9	17	11.3	10	**59.5**	59.4	63.2	94.0	81.75		G03	D-lactic acid methyl ester	10	16	13.3	34.25	**48.5**	16	29.5	19.0	9
C04	D-galactose	11	16	13.3	12.5	**49.5**	10.6	21.0	13.5	9.75		G04	L-lactic acid	11	15	12.7	13.5	**45.5**	10.4	7.7	9.5	7.75
C05	3-methyl glucose	11	17	11.7	10.75	**32.5**	9.8	4.3	11.0	6.25		G05	citric acid	9	16	12.7	**19**	1	**12.4**	77.2	8.0	5.75
C06	D-fucose	9	19	13.0	11.5	**62**	9.2	27.5	10.5	6.75		G06	α-keto-glutaric acid	9	15	11.7	**17.5**	42.5	**11.2**	15.2	26.3	6.5
C07	L-fucose	9	21	13.0	14	**65**	12	41.7	11.3	8.25		G07	D-malic acid	9	14	11.7	16.25	1	4.6	8.8	4.8	5.5
C08	L-rhamnose	9	19	17.0	11.25	**52.5**	8.4	16.7	7.8	6.75		G08	L-malic acid	9	15	11.0	13.5	**56.5**	9	11.7	6.5	6
C09	inosine	10	13	11.0	10.75	17.5	5	10.5	8.8	7		G09	bromo-succinic acid	9	7	7.3	7.25	1	4.6	3.8	5.0	5.75
C10	1% Na-lactate	98	100	99.0	48.5	9	9.8	77.2	47.0	62		G10	nalidixic acid	11	27	17.3	18.75	18	14.8	33.3	22.0	19.5
C11	fusidic acid	15	19	15.0	13	10.5	11	9.5	13.5	15		G11	LiCl	13	17	14.0	16	8	13.8	9.0	14.5	15.5
C12	D-serine	13	16	14.0	11	9	9.6	85.0	14.0	34		G12	K-tellurite	99	97	85.3	81.5	32	87.6	14.3	22.0	35.25
D01	D-sorbitol	15	9	12.7	13	10	10.4	3.3	14.0	13.25		H01	tween-40	25	14	30.3	67.0	1	97.6	75.0	96.5	84.75
D02	D-mannitol	9	12	11.3	12.25	17.5	6.2	6.2	52.8	7		H02	γ-amino-butyric acid	10	13	14.3	20.5	51	12	13.8	14.8	15.25
D03	D-arabitol	9	11	9.7	8	19	4.8	3.2	7.3	5.5		H03	α-hydroxy-butyric acid	12	14	16.3	17.25	2	14	10.2	13.8	14.25
D04	myo-inositol	10	10	9.7	9	26	6	3.3	8.0	5.5		H04	β-hydroxy-D,L-butyric acid	10	15	13.7	17.25	17.5	14.2	11.3	12.8	11.25
D05	glycerol	9	12	9.0	**9.5**	17.5	5	44.5	69.3	19.25		H05	α-keto-butyric acid	9	11	12.0	10.75	4.5	7.6	7.3	6.5	7.5
D06	D-glucose-6-PO4	10	24	14.3	53.8	56.5	83.6	32.2	13.8	7.5		H06	acetoacetic acid	100	100	100.0	34.3	86.0	43.6	88.3	64.8	25.5
D07	D-fructose-6-PO4	9	27	15.3	23.25	96.5	16.4	48.7	15.0	24.25		H07	propionic acid	9	9	11.3	11	1.5	14.2	9.8	7.3	7.5
D08	D-aspartic acid	9	7	7.7	**7.25**	6	4.6	60.2	4.8	4.25		H08	acetic acid	10	46	50.7	99.8	98.0	77.8	94.5	99.0	30.5
D09	D-serine	9	27	15.0	32.5	1	7.2	45.5	5.0	5.75		H09	formic acid	9	15	13.3	15.75	8.5	8.4	67.2	6.8	11
D10	troleandomycin	9	23	14.3	9.5	11	9.2	79.2	10.0	9.75		H10	aztreonam	22	46	30.0	99.5	62.5	97.6	93.3	98.8	99
D11	rifamycin SV	10	20	13.7	10.75	10.5	10	48.0	10.0	12.25		H11	Na-butyrate	13	98	70.3	20.5	10.5	20.2	36.3	19.0	20.25
D12	minocycline	13	22	16.7	14.25	11.5	14	24.5	12.5	17		H12	Na bromate	17	14	17.3	21.75	8	22.8	13.8	24.5	22

Figure 9. Comparison of Biolog GenIII test results of a previous trial of *Kaistella koreensis* with results taken from the same plate at 22 hours incubation and 41 hours incubation, *Kaistella koreensis* from this study*, Lycomia zaccaria, Lycomia vostokensis, Chryseobacterium haifense, Chryseobacterium jejuense, Chryseobacterium shigense,* and *Chryseobacterium soli*.

The results for the API ZYM test for the presence of hydrolytic enzymes in *Kaistella koreensis* can be seen in Figure 10 below. *K. koreensis* tested positive for the presence of 10 of the 20 enzymes that were being tested for including; alkaline phosphatase, esterase, esterase lipase, leucine arylamidase, valine arylamidase, cystine arylamidase, chymotrypsin, acid phosphatase, napthol-AS-BI-phosphohydrolase, and glucosidase. The results of *K.* koreensis were then compared

to the results for *Lycomia zaccaria, Lycomia vostokensis, Chryseobacterium haifense, Chryseobacterium jejuense, Chryseobacterium vrystaatense,* and *Chryseobacterium oranimense;* these results can be seen in Figure 11 below. The most notable result of this comparison is the presence of the enzyme chymotrypsin, which was not found to be present in any of the closest related organisms to *K. koreensis.*

Figure 10. API ZYM test results for the presence of hydrolytic enzymes in *Kaistella koreensis*. The presence of a color indicates a positive result. The test for the presence of chymotrypsin is in well 10.

abbrev	Kkor	Lzac	Lvos	Lhaif	Cjeju	Cvrys	Coran
strain	CCUG 49689	JJC	3915-10	DSM 19056	DSM 19399	LMG 22846	LMG 24030
<-- test Genus species	Kaistella koreensis	Lycomia zaccaria	Lycomia vostokensis	Lycomia haifense	Chryseobacterium jejuense	Chryseobacterium vrystaatense	Chryseobacterium oranimense
control	-	-	w	w	-	-	-
alkaline phosphatase	+	+	+	+	+	+	+
C4 esterase	+	+	+	+	+	+	+
C8 esterase lipase	+	+	+	+	+	+	+
C14 lipase	-	w	w	w	-	-	-
leucine arylamidase	+	+	+	+	+	+	+
valine arylamidase	+	w	w	+	+	+	+
cystine arylamidase	+	w	w	+	-	-	-
trypsin	-	-	-	-	+	-	-
α-chymotrypsin	+	-	-	-	-	-	-
acid phosphatase	+	+	+	+	+	+	+
naphthol-AS-BI-phosphohydrolase	+	+	+	+	+	+	+
α-galactosidase	-	-	-	w	-	-	-
β-galactosidase	-	-	-	+	-	-	-
β-glucuronidase	-	-	-	w	-	-	-
α-glucosidase	+	w	w	w	+	-	+
β-glucosidase	-	-	-	-	-	-	+
N-acetyl-β-glucosaminidase	-	-	-	-	w	w	+
α-mannosidase	-	-	-	-	-	w	-
α-fucosidase	-	-	-	-	-	-	-

Figure 11. API ZYM test results for the presence of hydrolytic enzymes in *Kaistella koreensis* and comparison to *Lycomia zaccaria, Lycomia vostokensis, Chryseobacterium haifense, Chryseobacterium jejuense, Chryseobacterium vrystaatense,* and *Chryseobacterium oranimense*.

The results of the Gas Chromatograph analysis of the Fatty Acid Methyl Esters can be seen in Figure 12 below. The predominant fatty acid in the lipid membrane of *Kaistella koreensis* was the 15:0 iso fatty acids, and the second most common fatty acid was the 15:0 anteiso fatty acids. The 15: iso fatty acid was also the predominant fatty acid in each of the other four organisms to which *K. koreensis* was compared except for

Lycomia vostokensis which had the 15:0 anteiso as the predominant fatty acid. *K. koreensis* differed in fatty acid composition from all four of the other organisms being compared in that it had a much higher percent total of 13:0 iso, 14:0 iso, 16:1 w9c, and 17:1 anteiso w9c fatty acids, and a much lower percent total of 17:1 iso w9c, and 17:0 iso 3OH fatty acids. The three members of the *Chryseobacterium* differed from the other organisms in that they had none of the 15:0 anteiso fatty acids present in their plasma membranes, while 15:0 anteiso fatty acids make up a significant portion of each of the other organisms' membranes.

Equivalent Chain Length (ECL)	Fatty Acid Name	K. koreensis	C. haifense	L. vostokensis	L. zaccaria	C. jejuense	C. shigense	C. soli
12.6189	13:0 iso	6.61	2.21	1.06	2.6	1.33	1.54	1.12
13.6261	14:0 iso	4.43	--	tr	--	--	1.58	--
14.6313	15:0 iso	30.23	35.95	23.92	35.99	37.06	42.16	38.06
14.7251	15:0 anteiso	16.86	15.4	28.8	13.22	tr	tr	2.86
15.4844	16:1 iso H	1.55	tr	tr	tr	--	--	--
15.6358	16:0 iso	2.3	tr	2.88	tr	--	tr	tr
15.7933	16:1 w9c	8.54	--	--	--	--	--	--
15.8802	16:1 w6c/16:1 w7c	2.73	4.78	1.84	2.59	11.43	8.47	11.98
16.002	16:0	2.19	tr	--	--	1.51	1.57	1.59
16.1656	15:0 iso 3OH	1.57	3.18	1.03	2.94	3.25	2.75	2.57
16.259	15:0 2OH	--	tr	1.43	tr	--	--	tr
16.637	17:0 iso	--	tr	tr	tr	1.28	tr	1.14
16.4464	17:1 iso w9c	5.66	21.6	14	22.31	20.22	18.43	18.96
16.5575	17:1 anteiso w9c	5.77	--	tr	--	--	--	--
17.1797	16:0 iso 3OH	3.41	1.3	3.51	1.88	tr	tr	1.22
17.9984	18:0	1.52	--	--	--	--	--	--
18.1907	17:0 iso 3OH	3	9.34	10.63	11.46	19.34	17.91	17.01
18.297	17:0 2OH	--	1.76	5.86	2	tr	tr	tr

Figure 12. FAME analysis results for *Kaistella* koreensis and comparison to results of *Chryseobacterium haifense, Lycomia vostokensis, Lycomia zaccaria, Chryseobacterium jejuense, Chryseobacterium shigense,* and *Chryseobacterium soli*. The values represent the percent total that each fatty acid makes up the plasma membrane.
tr=trace amount present

The results of the Average Nucleotide Identity computations can be seen in Figure 13 below. *Kaistella koreensis* exhibited a 78.6% and 75.3% ANI to organisms belonging to the genus *Lycomia*; *L. zaccaria* and *L. vostokensis* respectively. *K. koreensis* also exhibited an ANI of 77.6%, 70.2%, and 69.7% to organisms belonging to the genus *Chryseobacterium*; *C. haifense, C. gleum,* and C. sp. CF314 respectively. Additionally, *K. koreensis* showed an ANI of 67.9% to *Riemerella anatipestifer*. It is also notable that the organism *Chryseobacterium haifense* exhibited an ANI of 83.8% and 79.7% to members of the genus *Lycomia*, while it exhibited an ANI of 72.0% and 72.1% to members of the genus *Chryseobacterium*. Finally, the two members analyzed from the genus *Lycomia* showed an ANI to each other of 80.7% and the two members analyzed from the genus *Chryseobacterium* showed an ANI to each other of 81.1%.

% Average Nucleotide Identity							
	Kaistella koreensis	Lycomia zaccaria	Lycomia vostokensis	Chryseo. haifense	Chryseo. gleum	Chryseo. Sp.CF314	Riemerella anatipestifer
K. koreensis							
L. zaccaria	78.6%						
L. vostokensis	75.3%	80.7%					
C. haifense	77.6%	83.8%	79.7%				
C. gleum	70.2%	71.9%	68.7%	72.0%			
C. sp. CF314	69.7%	72.4%	68.6%	72.1%	81.1%		
R. anatipestifer	67.9%	70.3%	68.0%	69.1%	66.4%	67.35%	

Figure 13. Average Nucleotide Identities of 7 members of the *Flavobacteriaceae* family.

The annotated genome also allowed for a partial metabolic reconstruction of a few of the metabolic pathways that *Kaistella koreensis* utilizes. These reconstructed metabolic pathways can be seen in Figures 16 through 22 at the end. *K. koreensis* utilizes the glycolysis pathway in order to break down glucose into pyruvate as seen in Figure 16. It also uses the Tri Carboxylic Acid cycle by which the pyruvate is completely oxidized in order to produce NADH and $FADH_2$, which can then be used to make energy in the form of ATP, as seen in Figure 17. The pentose phosphate pathway, Figure 18, is used in order to produce ribose, which can then be used for DNA synthesis. Fermentation, Figure 19, is the process by which the pyruvate is able to be broken down in the absence of the electron transport chain. Finally, *K. koreensis* utilizes the purine and pyrimidine metabolic pathways, Figures 20-22, in order to synthesize the nucleotide bases for DNA synthesis.

Discussion

There are several differences between the genome of *Kaistella koreensis* and that of other members of the genus *Chryseobacterium* that have had their genomes sequenced, although there are currently only two, *Chryseobacterium gleum* and *Chryseobacterium* sp. CF314. The first notable difference is the size of the genome itself. The size of the *K. koreensis* genome is approximately 3.2 Mbp, which is much smaller than the size of other *Chryseobacterium* species with genomes available in RAST for comparison, which are closer to the range of 4.5 – 5.5 Mbp. A second notable difference is that *K. koreensis* has many fewer genes than the other two *Chryseobacteria*; *K. koreensis* has 3,677 genes while the other *Chryseobacteria* both have over 4,000 genes. Additionally, *K. koreensis* has many fewer RNA's than the

other *Chryseobacteria*. *K. koreensis* is closer to the genus Lycomia in terms of genome size and number of genes; however, *K. koreensis* has a slightly larger genome and more genes than the members of the *Lycomia* genus. This comparison can be seen in Figure 14 below. These differences suggest that *K. koreensis* is more closely related to the *Lycomia* genus than it is to the *Chryseobacterium* genus and support the assertion that *K. koreensis* should not have been reclassified into the *Chryseobacterium* genus. *K. koreensis* appears to be somewhere in between the *Chryseobacterium* and *Lycomia* genera, it is smaller in genome size and number of genes than the *Chryseobacterium* genus, and larger in genome size and number of genes than the *Lycomia* genus. This relationship can also be seen in the phylogenetic tree, made using 16S rRNA sequences, shown in Figure 15 below. *K. koreensis* branches off between the *Chryseobacterium* and the *Lycomia* genera, although it branches slightly closer to the *Lycomia* genus.

	k. koreensis	L. zaccaria	L. vostokensis	C. haifense	C. gleum	C. sp. CF314
Genome size (bp)	3,199,137	2,859,501	2,768,102	2,822,542	5,570,594	4,488,346
Number of genes	3,677	2,860	2,555	3,248	5,163	4,178
Number of RNA's	40	38	48	37	76	54

Figure 14. Comparison of genome size, number of genes, and number of RNA's of *K. koreensis, L. zaccaria, L. vostokensis, C. haifense, C. gleum,* and *C.* sp. CF314.

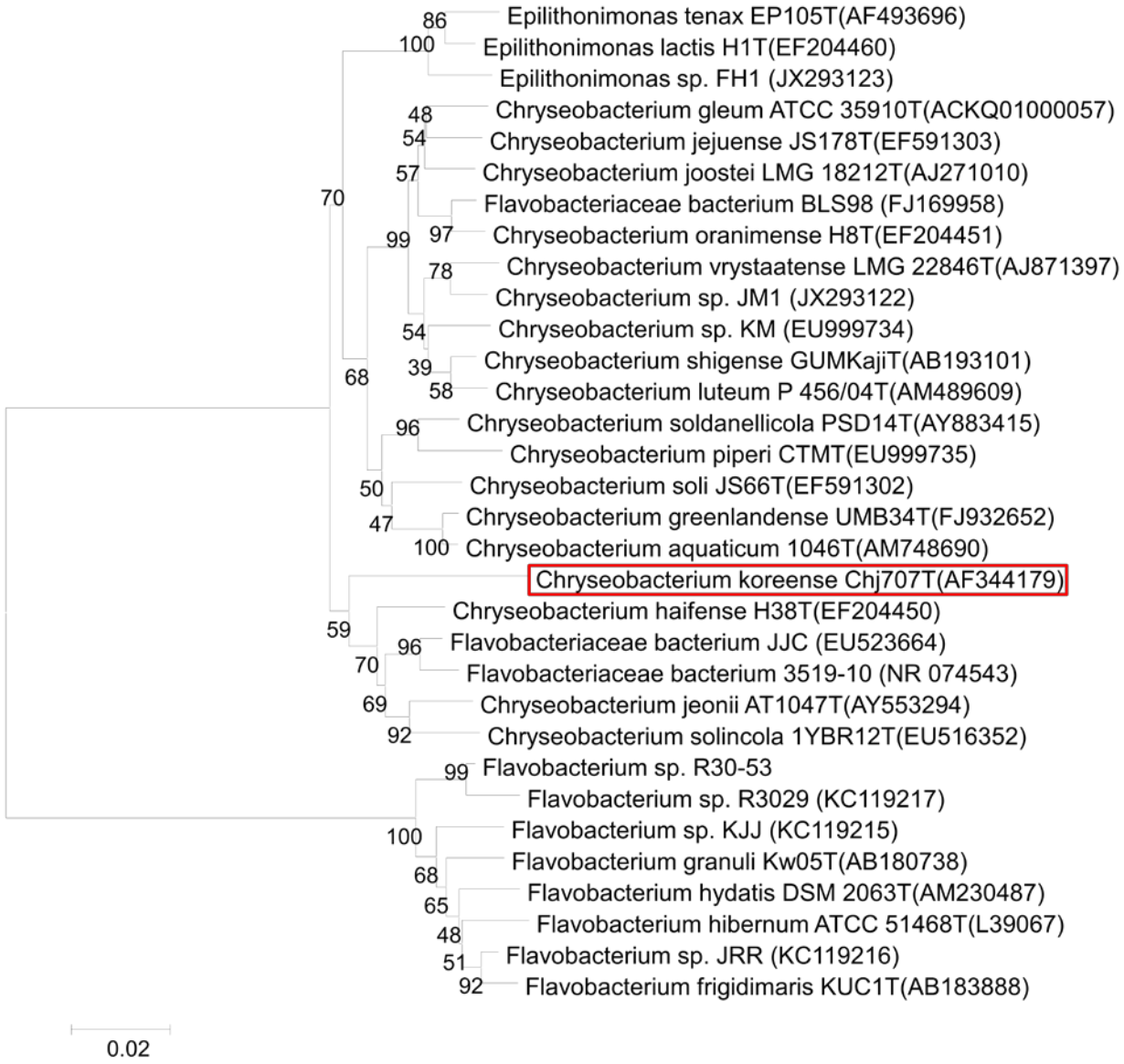


Figure 15. Phylogenetic tree made using 16S rRNA sequences showing the relationship between *Kaistella koreensis*, highlighted, and members of the *Chryseobacterium* and *Lycomia* genera. *Flavobacteriaceae bacterium JJC* is also known as *Lycomia zaccaria*, and *Flavobacteriaceae bacterium* 3519-10 is also known as *Lycomia vostokensis*.

The Biolog GenIII test is a very useful test for the characterization of bacteria because it allows 96 different tests to be done simultaneously and thus reveal a lot of information about the organism in a relatively short period of time. The comparison of the Biolog results of *Kaistella koreensis* to *Lycomia zaccaria, Lycomia vostokensis, Chryseobacterium haifense, Chryseobacterium jejuense, Chryseobacterium shigense,*

and *Chryseobacterium soli* show that there are a number of differences between *K. koreensis* and both the *Lycomia* and *Chryseobacterium* genera. These differences suggest that *K. koreensis* has a number of phenotypic differences from other members of the *Flavobacteriaceae* family, which may make it different enough that it belongs in its own genus. This would also support the original assertion that *K. koreensis* should not have been reclassified as *Chryseobacterium koreense* and that it should be reclassified back to its original genus. The results of the Biolog test are also consistent with what genes are present in the organism's genome for breaking down different nutrient sources. It is also useful to see how the results changed from the same plate after a longer incubation time. This suggests that the genes for utilizing nutrient sources such as glycyl-L-proline, aspartic acid, galacturonic acid, and acetic acid were not initially being expressed, but after a period of exposure to these nutrients the organism was able to induce the genes needed for the utilization of these nutrients. *K. koreensis* was able to use some of the sources immediately because some of the genes for utilizing these sources were probably already being induced as a result of the growth conditions on the plate the organism was grown on prior to performing the test. The genome annotation shows what the organism is capable of using, while the phenotypic tests show what the organism is actually using at that specific time, often as a result of the conditions it was previously growing in. For example, the annotated genome shows a number of genes present for the utilization of maltose, mannose, galacturonic acid, and glucoronic acid. However, the Biolog results showed only very weak utilization of all four of these nutrient sources. This would suggest that *K. koreensis* is capable of using these sources, but it was not utilizing them at that particular time. It is also notable that

K. koreensis was able to utilize both gelatin and glycyl-proline. This makes sense because gelatin is made from bone and connective tissue that is made up of collagen. The structure of collagen is a triple helix, the polypeptide chain that makes up the strands of the helix consists of glycine and proline residues as the small side chain of glycine allows for the strands to be close together and the ring structure of the proline side chain bound to the backbone allows for the chain to make turns. If the organism is able to utilize gelatin, which is made of glycine and proline, it makes sense that it would also be able to use glycyl-proline.

The API ZYM test also allows a number of tests to be performed simultaneously, as it tests for the presence of 20 different hydrolytic enzymes in an organism at the same time. The most notable result of this test for *Kaistella koreensis* is the presence of the enzyme chymotrypsin. Chymotrypsin is a serine protease, meaning it breaks down proteins and polypeptides by cleaving peptide bonds. Specifically, chymotrypsin cleaves peptide bonds after hydrophobic amino acids including; tyrosine, tryptophan, and phenylalanine. The presence of this enzyme in *K. koreensis* is notable due to its absence in the other closely related organisms. *K. koreensis* is the only organism within the *Lycomia, Chryseobacterium, or Kaistella* genera that yielded a positive result for this enzyme. *K. koreensis* is also the only member of the genus *Kaistella*. This difference distinguishes the *Kaistella* genus from the other genera in the *Flavobacteriaceae* family and would be a defining characteristic of this genus as *K. koreensis* is the type strain for the genus. All future members of the *Kaistella* genus would also have to have chymotrypsin present. However, the API ZYM test measures the ability of the organism to breakdown the substrate, not the presence of that specific enzyme. There is no gene

specifically for chymotrypsin in the *K. koreensis* genome, there is, however, a gene for a serine endopeptidase that would be responsible for breaking down the substrate that yielded a positive result for chymotrypsin. Whether or not this serine endopeptidase is chymostrypsin is unknown.

The Fatty Acid Methyl Ester analysis can aid in distinguishing differences between organisms because the lipid composition of the membrane can affect the ability of the organism to transport molecules across, and within, the membrane. The amount of saturated and unsaturated fatty acids present in the membrane can affect the fluidity of the membrane, making it either more or less fluid and, therefore, more or less easy for molecules to move about within the membrane. The results of the FAME analysis for *Kaistella koreensis* and the comparison to *Chryseobacterium haifense, Lycomia vostokensis, Lycomia zaccaria, Chryseobacterium jejuense, Chryseobacterium shigense,* and *Chryseobacterium soli* showed several differences in composition between *K. koreensis* and the other four organisms. *K. koreensis* tends to have a greater number of shorter, saturated fatty acids than the other organisms. This would cause the plasma membrane of *K. koreensis* be more compact and less fluid than those organisms closely related to it. This is because saturated fatty acids do not have any kinks or bends that are present in unsaturated fatty acids, so they would be able to pack together more tightly. *K. koreensis* also differed from the members of the *Chryseobacterium* in that members of the *Chryseobacterium* genus have little to none of the 15:0 anteiso fatty acids present, while the 15:0 anteiso fatty acids made up a significant portion, almost 17%, of the plasma membrane of *K. koreensis*. The differences that are seen between *K. koreensis* and the members of the genera

Lycomia and *Chryseobacterium* also support the hypothesis that *Kaistella koreensis* does not belong in either the *Chryseobacterium* or *Lycomia* genera; instead, it belongs in its own separate genus, *Kaistella*, as it was originally classified. The results of the FAME experiment are consistent with what is published on *K. koreensis*, although the exact percent total of the fatty acids are slightly different in some cases. The most notable differences between the percent total fatty acids determined in this experiment and what has been published is in the two predominant fatty acids present. The predominant fatty acid 15:0 iso, was found to be higher in the published article on this organism, 52% in literature versus approximately 30% in this experiment, while the second most predominant fatty acid, 15:0 anteiso, was found to be lower in what has been published, 11% in literature versus approximately 17% in this experiment (Kim *et al* 2004).

The results of the Average Nucleotide Identity computations are in agreement with the hypothetical division between different genera. Both members of the *Lycomia* genus that were compared showed an ANI to each other of 80.7%, while *Chryseobacterium gleum* and *Chryseobacterium* sp. CF314 had an ANI of 81.1%. Furthermore, *Reimerella anatipestifer*, which did not share a genus with any of the organisms being compared, had an ANI in the low 70% and lower when compared to each of the other organisms. This suggests that an ANI of 80% to 95%, which defines different species, does indicate that the organisms belong to the same genus, while an ANI in the low 70% or lower indicates that the organisms being compared belong to different genera. The ANI results for *Kaistella koreensis* show an ANI of 70.2% and 69.7% to *Chryseobacterium* gleum and *Chryseobacterium* sp. CF314 respectively. This

indicates that *K. koreensis* definitely does not belong to the genus *Chryseobacterium* and should never have been reclassified as *Chryseobacterium koreense*. However, the ANI when compared to the members of the *Lycomia* genus are 78.6% and 75.3%, which falls right in the unclear zone for determining different genera. The ANI of 78.6% to *Lycomia zaccaria* suggests that perhaps *K. koreensis* does belong to the *Lycomia* genus while the ANI of 75.3% to *Lycomia vostokensis* suggests that maybe *K. koreensis* does not belong to *Lycomia* genus. As a result, the exact genus to which *K. koreensis* cannot be determined by the ANI computations alone.

Additionally, the ANI results for *Chryseobacterium haifense* show an ANI of 72.0% and 72.1% when compared to the other two organisms in the *Chryseobacterium* genus that were analyzed. This suggests that *C. haifense* does not belong to the *Chryseobacterium* genus. Instead, the ANI of 83.8% and 79.7% to the members of the *Lycomia* genus suggests that *C. haifense* actually belongs to the genus *Lycomia.*

In conclusion, the ANI test is useful for comparing sequence similarities between genomes. However, there are currently no clear cutoff points that would distinguish ANI between different genera. This leaves the results of the ANI mostly ambiguous especially for the mid to upper 70's range. Therefore, the ANI results must be supplemented by other phenotypic tests, such as the Biolog GenII, API ZYM, and FAME analysis in order to support the results. In this case the phenotypic tests supported the results of the ANI that *Kaistella koreensis* does not belong in the *Chryseobacterium* or *Lycomia* genera and should not have been reclassified as *Chryseobacterium koreense*. Instead, *Chryseobacterium koreense* should be reclassified back to *Kaistella koreensis* as it was initially classified.

References

Achberger, A.M., Brox, T.I., Skidmore, M.L., and Christner, B.C. (2011). Expression and partial characterization of an ice-binding protein from a bacterium isolated at a depth of 3,519 m in the Vostok ice core, Antarctica. *Frontiers in Microbiology* 2:1-8.

Auch, A.F., von Jan, M., Klenk, H.P., Goker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Standards in Genomic Sciences* 2:117-34.

Aziz, R., Bartels, D., Best, D., DeJongh, M., Disz, T., Edwards, R., *et al*. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75.

Baker, M. (2012). De novo genome assembly: what every biologist should know. *Nature Methods* 9, 333-7.

Beer, H. Hugo, C.J., Jooste, P.J., Willems, A., Vancanneyt, M., Coenye, T., and Vandamme, P.A.R. (2005). Chryseobacterium vrystaatense sp. nov., isolated from raw chicken in a chicken-processing plant. *International Journal of Systematic and Evolutionary Microbiology* 55:2149-53.

Buonaccorsi, V., Boyle, M., Grove, D., Praul, C., Sakk, E., Stuart, A., Tobin, T., Hosler, J., Carney, S., Engle, M., Overton, B., Newman, J., Pizzorno, M., Powell, J., and Trun, N. (2011). GCAT-SEEKquence: Genome Consortium for Active Teaching of Undergraduates through Increased Faculty Access to Next-Generation Sequencing Data. *CBE Life Sci Educ.* 10(4): 342-345.

Carver, T., Berriman, M., Tivey, A., Patel, C., Böhme, U., Barrell, B., Parkhill, J., and Rajandream, M. (2008). Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24:2672-6.

Geneious version 6.1 created by Biomatters. Available from http://www.geneious.com.

Hantsis-Zacharov, E., and Halpern, M. (2007). Chryseobacterium haifense sp. nov., a psychrotolerant bacterium isolated from raw milk. *International Journal of Systematic and Evolutionary Microbiology* 57:2344-8.

Hantsis-Zacharov, E., Shakéd, T., Senderovich, Y., and Halpern, M. (2008). Chryseobacterium oranimense sp. nov., a psychrotolerant, proteolytic and lipolytic bacterium isolated from raw cow's milk. *International Journal of Systematic and Evolutionary Microbiology* 58:2635-9.

Henson, J., Tischler, G., and Ning, Z. (2012). Next-generation sequencing and large genome assemblies. *Pharmacogenomics* 13:901-15.

Holmes, B., Owen, R.J., Steigerwalt, A.G., and Brenner, D.J. (1984). Flavobacterium gleum, a New Species Found in Human Clinical Specimens. *International Journal of Systematic Bacteriology* 34:21-5.

Kämpfer, P., Vaneechoutte, M., Lodders, N., De Baere, T., *et al*. (2009). Description of Chryseobacterium anthropi sp. nov. to accommodate clinical isolates biochemically similar to Kaistella koreensis and Chryseobacterium haifense, proposal to reclassify Kaistella koreensis as Chryseobacterium koreense comb. nov. and emended description of the genus Chryseobacterium. *International Journal of Systematic and Evolutionary Microbiology* 59:2421-8.

Kim, M., Im, W., Shin, Y., Lim, J., Kim, S., Lee, B., Park, M., Lee, K., and Lee, S. (2004). Kaistella koreensis gen. nov., sp. nov., a novel member of the Chryseobacterium–Bergeyella–Riemerella branch. *International Journal of Systematic and Evolutionary Microbiology* 54:2319-24.

Klenk, H. P., and Göker, M. (2010). En route to a genome-based classification of Archaea and Bacteria. *Systematic and Applied Microbiology* 33:175-82.

Life Technologies Corporation. *Ion Torrent Technology How Does It Work?* 2013. http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Semiconductor-Sequencing/Semiconductor-Sequencing-Technology/Ion-Torrent-Technology-How-Does-It-Work.html.

Newman JD. Biology 444 Laboratory Manual, Spring 2013, pp. 1-8.

Newman JD. Microbiology Laboratory Manual, Spring 2013, pp. 43-5, 52-3.

Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *PNAS* 106:19126-31.

Shimomura, K., Kaji, S., and Hiraishi, A. (2005). Chryseobacterium shigense sp. nov., a yellow-pigmented, aerobic bacterium isolated from a lactic acid beverage. *International Journal of Systematic and Evolutionary Microbiology* 55:1903-6.

Weon, H.Y., Kim, B.Y., Yoo, S.H., Kwon, S.W., Stackebrandt, E., and Go, S.J. (2008). Chryseobacterium soli sp. nov., and Chryseobacterium jejuense sp. nov., isolated from soil samples from Jeju, Korea. *International Journal of Systematic and Evolutionary Microbiology* 58:470-3.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Kunin, V., Goodwin, L., Wu, M., Tindall, B. J., Hooper, S. D., Pati, A., Lykidis, A., Spring, S., Anderson, I. J., D'haeseleer, P., Zemla, A., Singer, M., Lapidus, A., Nolan, M., Copeland, A., Han, C., Chen, F., Cheng, J., Lucas, S., Kerfeld, C., Lang, E., Gronow, S., Chain, P., Bruce, D., Rubin, E. M., Kyrpides, N. C., Klenk, H., and Eisen, J. A. (2009). A phylogeny-driven genomic encyclopedia of Bacteria and Archaea. *Nature* 462:1056-60.

Partial Metabolic Reconstruction

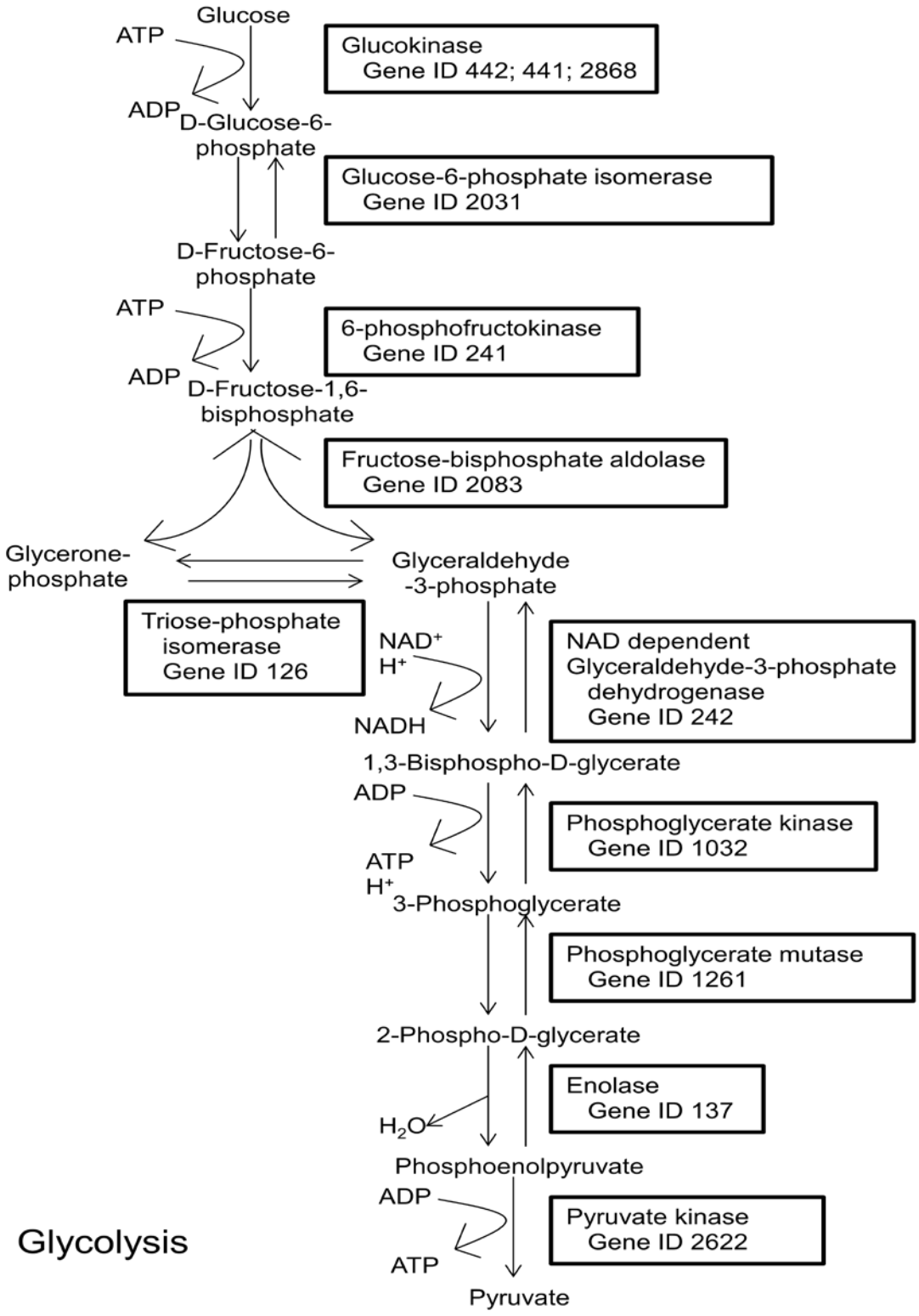


Figure 16. Glycolysis, metabolic pathway by which *K. koreensis* breaks down glucose into pyruvate.

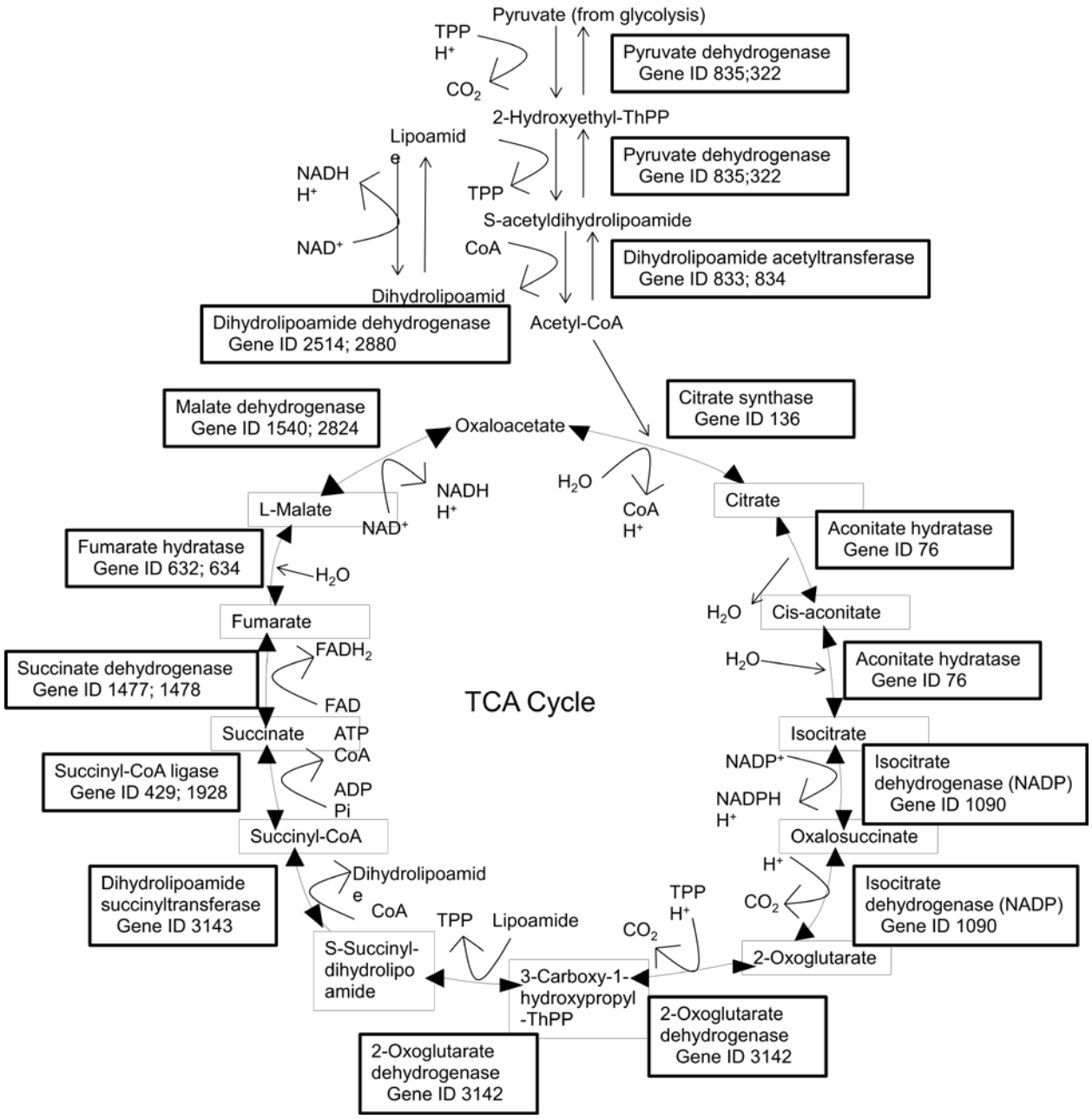


Figure 17. Tri Carboxylic Acid Cycle, metabolic pathway by which *K. koreensis* completely oxidizes pyruvate in order to produce NADH and $FADH_2$ for energy production.

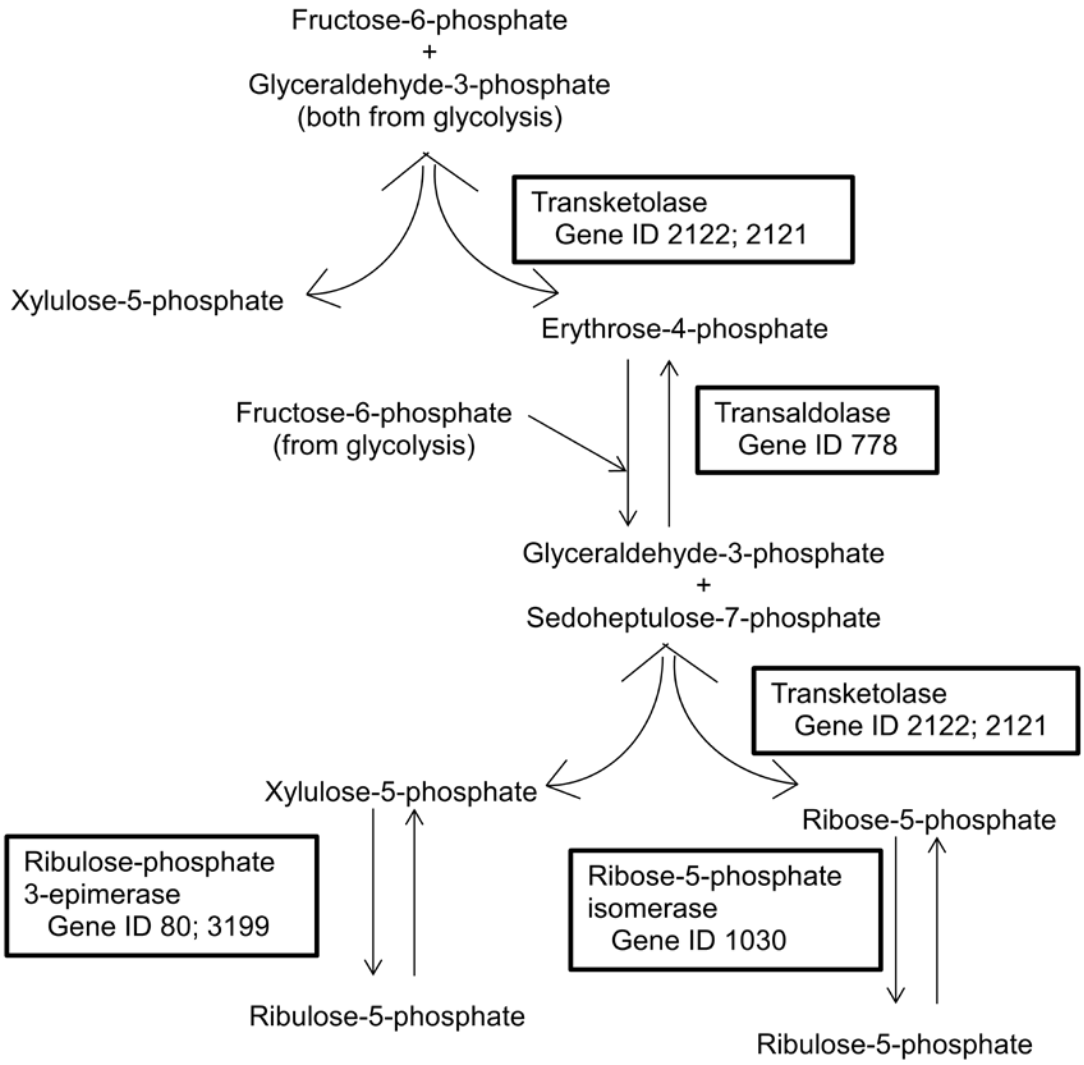


Figure 18. Pentose Phosphate Pathway, metabolic pathway by which *K. koreensis* makes ribose for DNA synthesis.

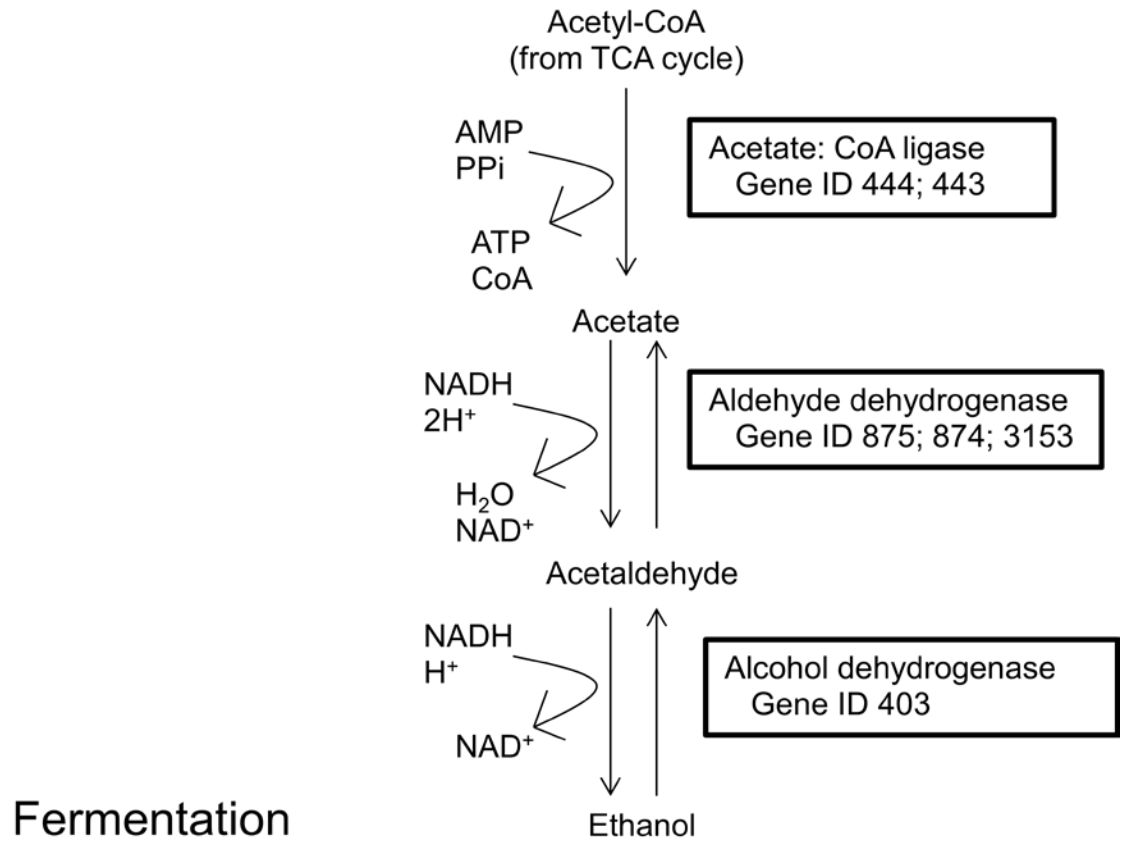


Figure 19. Fermentation, metabolic pathway by which *K. koreensis* breaks down pyruvate under anaerobic conditions.

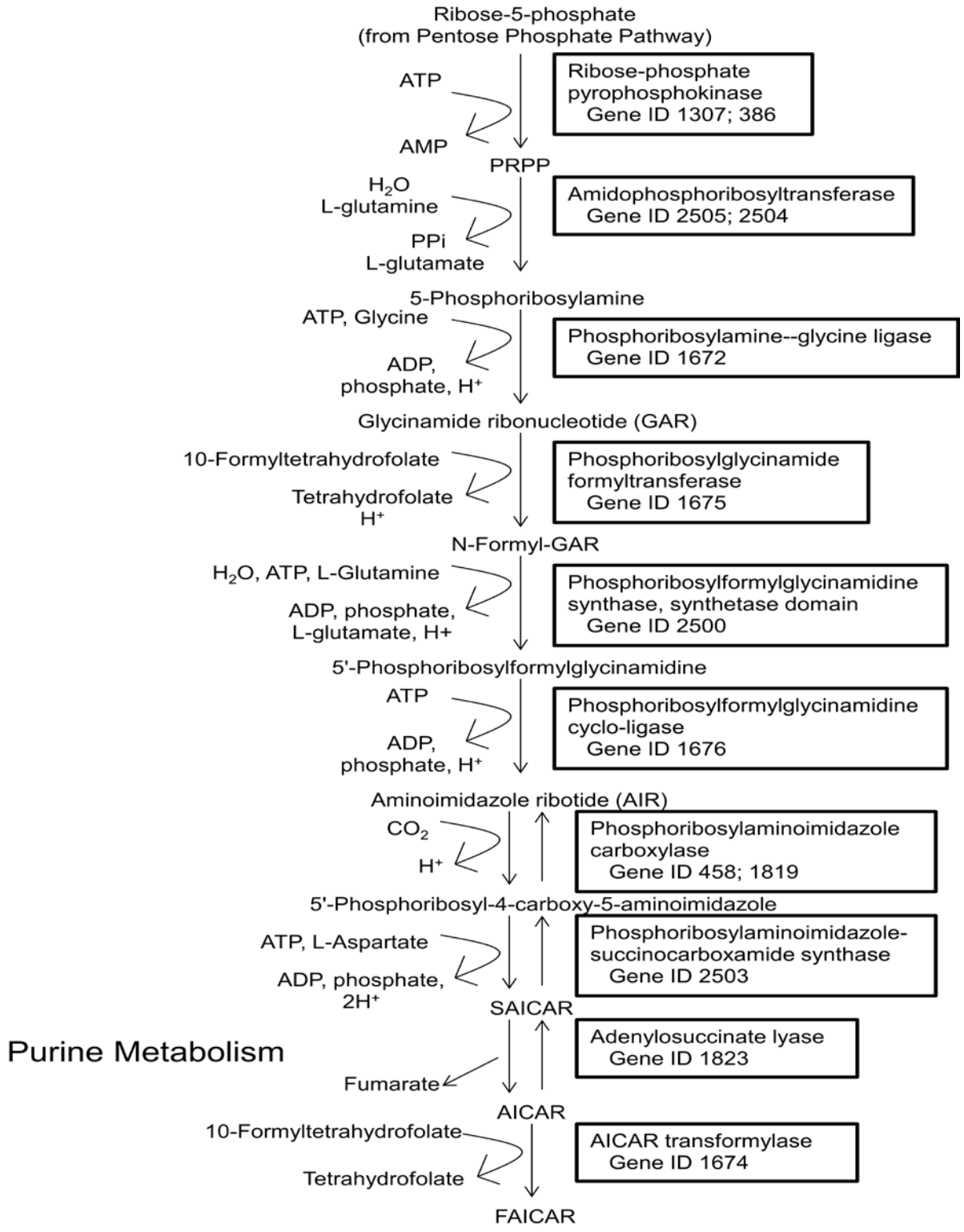
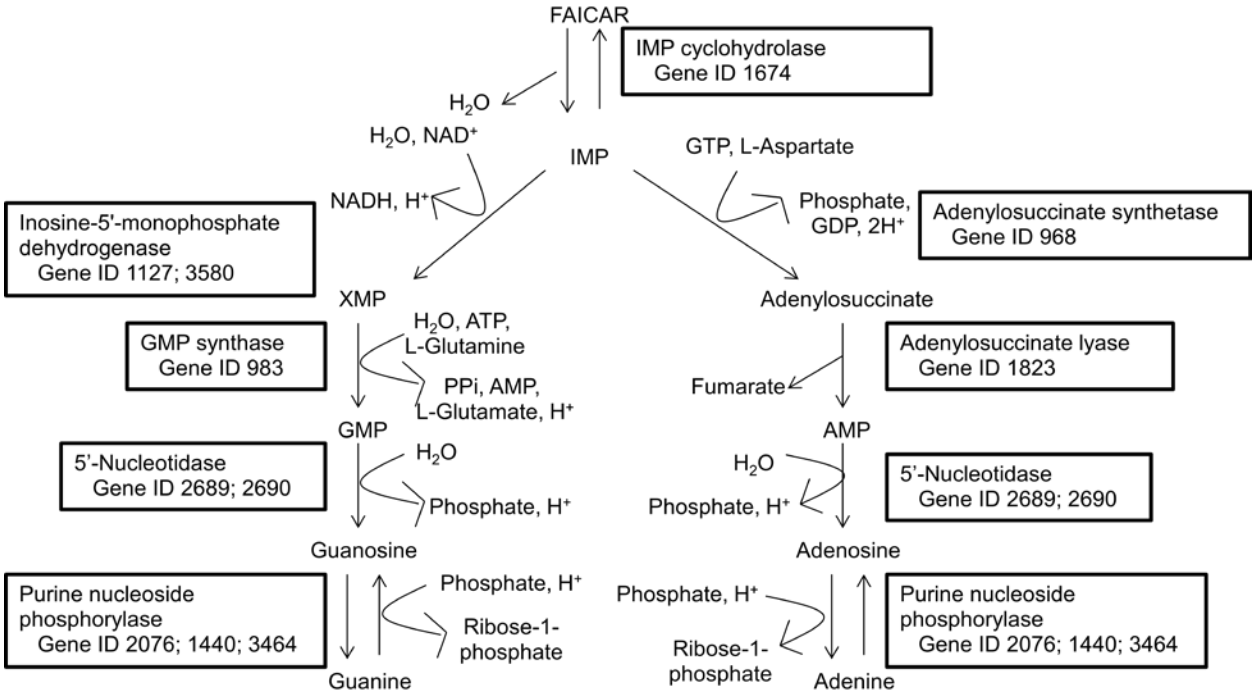


Figure 20. Purine Metabolism, metabolic pathway by which *K. koreensis* makes adenine and guanine for DNA synthesis.



Purine Metabolism Continued

Figure 21. Purine Metabolism continued, metabolic pathway by which *K. koreensis* makes adenine and guanine for DNA synthesis.

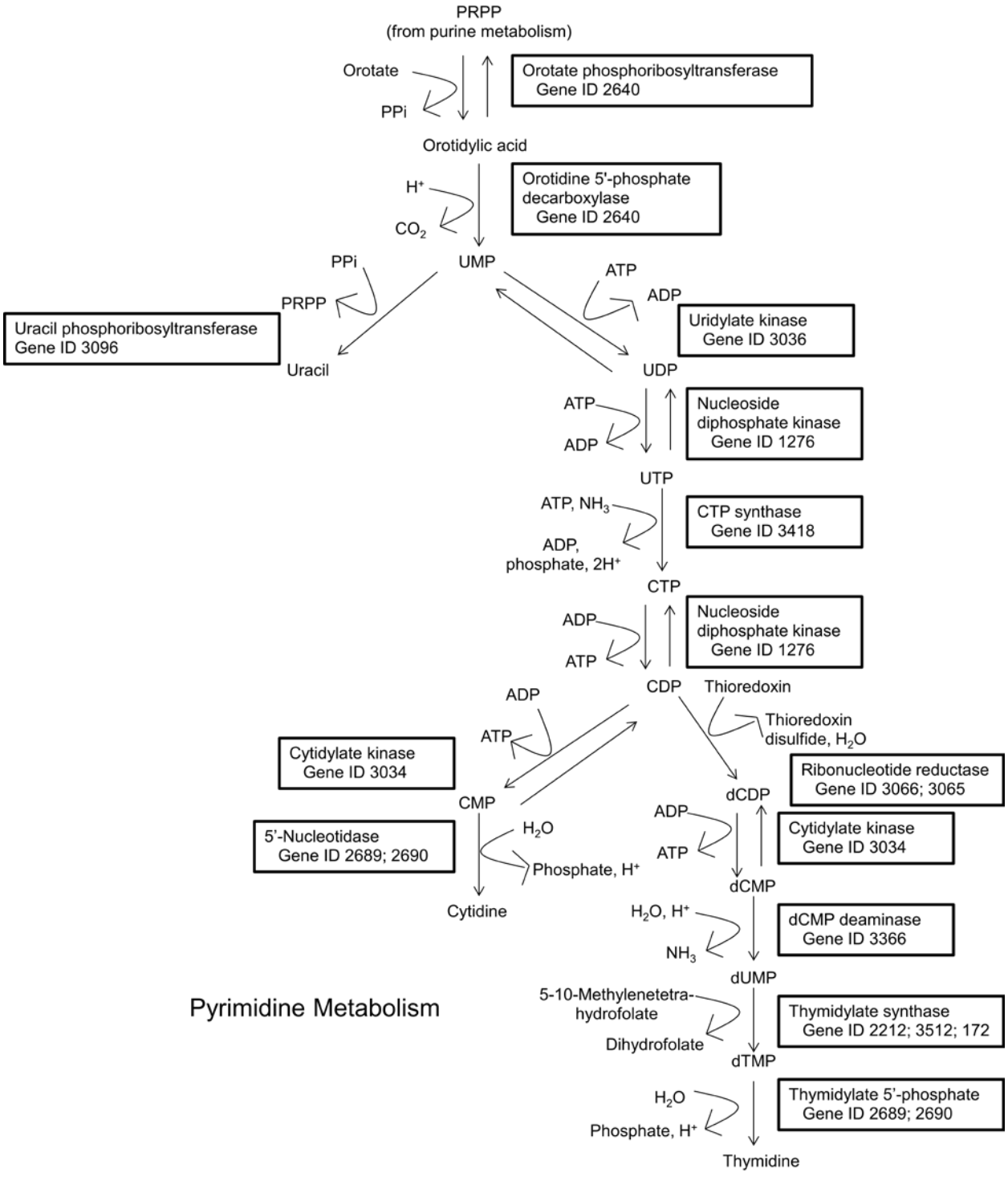


Figure 22. Pyrimidine Metabolism, metabolic pathway by which *K. koreensis* makes uracil, cytosine, and thymine for DNA synthesis.