2019

Effects of Temperature on CRISPR/Cas System

Eddie Beckom
Augustana College, Rock Island Illinois

Dr. Lori Scott
Augustana College, Rock Island Illinois

Eddie Beckom
BIO 375
Dr. Lori R. Scott
Biology Department, Augustana College
639 38th Street, Rock Island, IL 61201

Temperature Effect on Complexity of CRISPR/Cas Systems

What is *Meiothermus ruber?*

Meiothermus ruber is a Gram-negative thermophilic rod-shaped eubacteria . The genus name derives from the Greek words ‘meion’ and ‘thermos’ meaning ‘lesser’ and ‘hot’ to indicate the thermophilic characteristics of *Meiothermus ruber.* (Nobre *et al*., 1996; Euzeby, 1997). It lives in thermal environments with an optimal temperature of 60℃. *Meiothermus ruber* belongs to the bacterial phylum Deinococcus-Thermus. The order Thermales, which is housed within the Thermus group and consists of 6 genera (Vulcanithermus, Oceanithermus, Thermus, Marinithermus, Meiothermus, Rhabdothermus), all containing genera with proteins that are thermostable. (Albuquerque and Costa, 2014). *M. ruber* is one of eight currently known species in the genus *Meiothermus* (Euzeby, 1997). As of 2017, five Meiothermus ruber genomes have been sequenced and uploaded into Genbank. Our study uses the *Meiothermus ruber* DSM 1279 genome (GenBank Name ASM2442v1) sequenced through a collaboration between the U.S. Joint Genome Institute and Leibniz-Institut DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH), which is called the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project. The many projected benefits of the GEBA project are novel gene discoveries, the identification of novel biochemical processes, and a better understanding of the processes underlying the evolutionary diversification of microbes (e.g., lateral gene transfer and gene duplication) which aids in the comparative analysis of *M. ruber* with model organisms. The complete genome of *Meiothermus ruber DSM 1279* consists of over 3 million base pairs, over 3000 protein-coding genes, 53 RNA genes and 6 CRISPR repeats; the genome has 63.4% GC content (Tindall *et al*., 2010). The phylogenetic relationship of Deinococcus-Thermus to phyla contain thermophiles is seen in Figure 1.

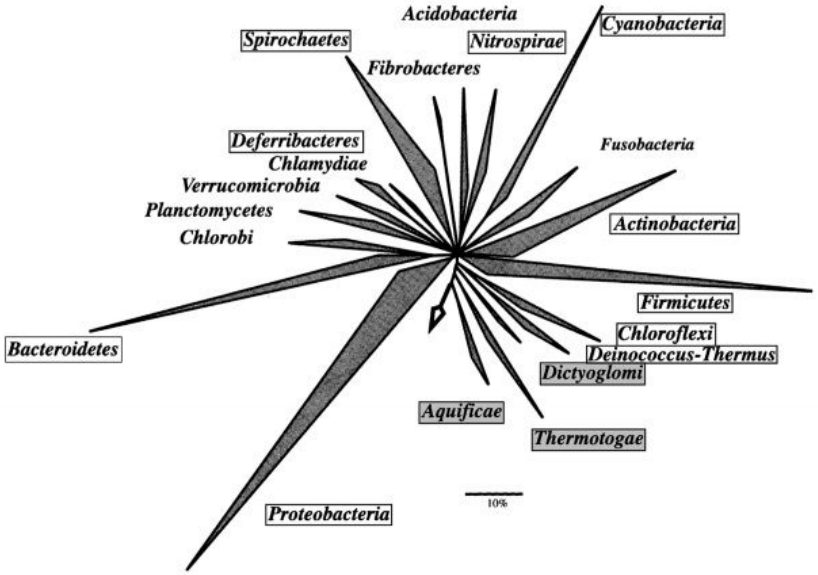

Figure 1. Phylogenetic tree showing the relationships between bacterial phyla based on W. Ludwig and H. P. Klenk's "Bergey's Manual of Systemic Bacteriology". Relationship between thermophiles based on the analysis of 16SrRNa genes with ARB parsimony tool. Of the phyla presented, Thermodesulfobacteria and Thermomicrobia are missing. (Lebedinsky *et al., 2007*)

Why study *Meiothermus ruber*?

Many physical and genome features make *M. ruber* an interesting organism to study. As a GEBA organism, little is known about the Deinococcus-Thermus phylum, with the exception of growing knowledge about members of the Thermus species. The notable characteristic is the thermophilic characteristics of *M. ruber* and how that impacts the biochemical processes that take place within the cells. Details about how thermophilic cells endure and adapt to temperature extremes in terms of the proteins and pathways that produce them are not well known. The study of CRISPR/Cas system in *M. ruber* is one of the many processes that have gone unstudied. As we learned from the GEBA project, most of what we know about biological processes in prokaryotes comes from studying a small percentage of the known microbes known as model organisms such as *Escherichia coli*. Expanding our knowledge to include poorly studied microbes from diverse branches of the tree of life is likely to identify diverse strategies to achieve many biological processes. To date, only 41 articles were pulled from the Pubmed search engine using the phrase Meiothermus ruber, while more than 366688 entries were pulled with *E. coli.* An understanding of the CRISPR/Cas system in *Meiothermus ruber* is essential in understanding the consensus between biosynthetic pathways between organisms and in what ways they may differ. Given the thermostability of proteins in thermophiles, studying the CRISPR/Cas system may expand on the knowledge of how defense mechanisms against viral infections take place in series of bacteria.

What are the parts of the CRISPR/Cas System?

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) adaptive immune systems are found in roughly 50% of bacteria and 90% of archaea (Makarova *et al*., 2015). CRISPR/Cas systems are one of the many prokaryotic defense systems present in bacterial cells (Labrie *et al.,* 2010) Similar to the immune systems of some eukaryotic organisms, such as humans, bacteria maintain a record of infections in regions of the CRISPR/Cas system sequence called CRISPR spacer-repeat arrays which allows for a quick response by the organism if a viral infection of that nature happens again. (Wright *et al*., 2016). CRISPR/Cas systems were first identified in *Escherichia coli* (Jiang & Doudna, 2015) along with an understanding of their sequence parts: a set of universal cas genes, series of cas genes varying by CRISPR type, and a spacer-array sequence. This makes *Escherichia coli* a model organism for the study of CRISPR/Cas systems in comparison to *Meiothermus ruber* and other bacterial organisms. An analysis of the *Escherichia coli* CRISPR/Cas system using Ecocyc (Keseler *et al*., 2013), a database devoted to the study of *E. coli*, and KEGG (Kanehisa *et al*., 2019) showed that the CRISPR/Cas system in *E. coli* is a singular operon with no evidence of genes for other types as shown in Figure 2 and Figure 3. This finding means that all genes shown in the KEGG analysis are found to be apart of the functional CRISPR-associated gene cascade or the universal cas gene complex.

Figure 2. Ecocyc (Keseler *et al., 2013*) analysis of the CRISPR/Cas operon in *Escherichia coli.* Shows the coding of the CRISPR/Cas system starting with signature gene *cas3* on the right. The system is coded for moving downstream, or from right to left. The system is made up of the signature gene and cas cascade genes.

Escherichia coli K-12 MG1655	KO
CRISPR-Cas system	
Universal Cas proteins	
b2755 cas1; multifunctional nuclease Cas1	K15342 cas1;
b2754 cas2; CRISPR-associated endoribonuclease Cas2	K09951 cas2;
Type I CRISPR-Cas system	
Type I signiture cas proteins	
b2761 cas3; CRISPR-associated endonuclease/helicase Cas3	K07012 cas3;
Subtype I-A factors	
Subtype I-B factors	
Subtype I-C factors	
Subtype I-D factors	
Subtype I-E factors	
b2760 casA; type I-E CRISPR system Cascade subunit CasA	K19123 casA;
b2759 casB; type I-E CRISPR system Cascade subunit CasB	K19046 casB;
b2758 casC; type I-E CRISPR system Cascade subunit CasC	K19124 casC;
b2757 casD; type I-E CRISPR system Cascade subunit CasD	K19125 casD;
b2756 casE; pre-CRISPR RNA endonuclease	K19126 casE;
Subtype I-F factors	
Subtype I-U factors	
Type II CRISPR-Cas system	
Type II signiture cas proteins	
Subtype II-A factors	
Subtype II-B factors	
Type III CRISPR-Cas system	
Type III signiture cas proteins	
Subtype III-A factors	
Subtype III-B factors	
Subtype III-U factors	

Figure 3. KEGG (Kanehisa *et al., 2019*) analysis of the CRISPR/Cas system in *Escherichia coli*. Evidence of one (1) functional operon made up of universal Cas proteins b2755 (Cas1) and b2754 (Cas2), Type 1 signature protein b2761 (Cas3), Subtype 1-E proteins b2760 (CasA), b2759 (CasB), b2758 (CasC), b2757 (CasD), b2756 (CasE).

The CRISPR operon structure begin with a signature protein, a necessary protein component for the production of an effective CRISPR/Cas operon, called *cas3. Cas3* downstream of following CRISPR associated cas genes: *casA*, *casB*, *casC*, *casD*, *casE* as seen in Figure 1. This sequence of downstream genes make up the cas cascade with *cas1* and *cas2* downstream from the final gene of the operon. *Cas1* and *cas2* are the universal Cas proteins present in all CRISPR/Cas operons. A classification of CRISPR/Cas operon types (Makarova *et al*., 2011) compared to the KEGG and Ecocyc results demonstrates that *E. coli* has a Type I-E CRISPR/Cas operon as represented in Figure 3.

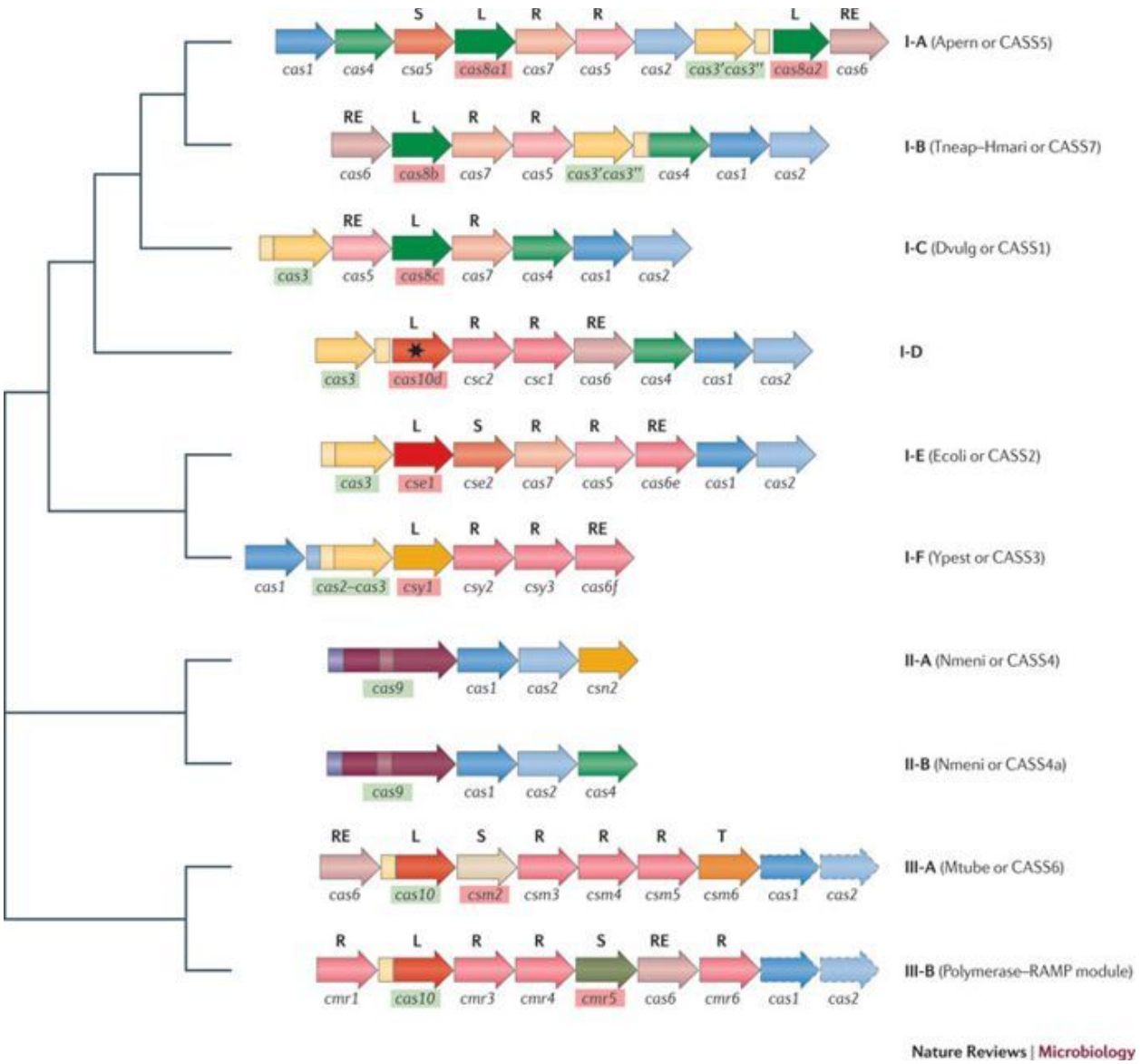

Figure 4. A classification of CRISPR/Cas type based on the presence of universal cas genes and CRISPR-associated genes that must be present in order to have a function CRISPR/Cas system. (Makarova *et al., 2011*)

How do bacteria adapt to temperature?

It is shown that *Escherichia coli* has a simple Type I-E CRISPR/Cas system as shown in Figure 3. *E. coli* is also a mesophile meaning that it grows an optimal temperature between 20-45℃. Bacteria in general grow at a wide range of temperatures: psychrophile below 20℃, mesophiles between 20-45℃, thermophiles 45-80℃, and hyperthermophiles above 80℃. (Goldstein 2007) The protein stability in these species of bacteria is referred to as the thermostability of the proteins. The interactions between amino acids during the folding is subject to the temperature of the bacterial environment and gives rise to functionally folded proteins or misfolded proteins based on the temperature effect. A bacterial protein in an organism outside of its temperature range will likely misfold due to the impact the temperature has on the interactions between the amino acids within the protein domains. An organismal adaptation to temperature-based impacts on interactions is for produced proteins to incorporate amino acids that are similar enough that the function of the protein remains but that the protein is more fitted for that specific temperature range.. Previous studies look at the interaction of thermophilic proteins and concludes that interactions in the protein domain to maintain thermostability are based on the hydrophobic effect of the environment and the enthalpic contributions amongst the protein domain (Berezovsky *et al*., 2005). The hydrophobic effect is the tendency for nonpolar molecules to interact in an aqueous environment. At room temperatures the hydrophobic effect is relatively low, but is directly related to the environmental temperatures. Proteins in various temperature environments interact in a continuously changing manner based on the effect. Another study looking at the thermostability of proteins in bacteria list the greater number of G+C content in a sequence as the region for stability in thermophilic bacteria proteins. (Lebedinsky *et al*., 2007) In comparing the G+C % content of thermophiles to mesophile, they approximately maintain 65 mol% to 55 mol%, respectively.

One group of thermophilic bacteria is the phylum Deinococcus-Thermus which consist of *Meiothermus ruber*, an organism known to have a CRISPR/Cas system. Lebedinsky *et al* shows the phylogenetic relationship between the Firmicutes and Deinococcus-Thermus which leads to the hypothesis that organisms within both phylum have a CRISPR/Cas system given *M. ruber* does. A study on the CRISPR/Cas system in *Streptococcus thermophilus,* an organism within the Firmicutes phylum, details the complexity of the *S. thermophilus* CRISPR/Cas system. (Horvath & Barrangou, 2010) *S. thermophilus* is known to have four individual operons with their own unique spacer-repeat CRISPR arrays. A previous study explored the classification and diversity of CRISPR/Cas systems based on the conclusion that the high energy cost of maintaining and expressing several CRISPR/Cas genes must offer major advantages on to the cell. (Garrett *et al*., 2011) The presence of multiple spacer-repeat arrays in a CRISPR/Cas systems likely confers the conclusion that parts of the system work independently of each other. This means that phase 1, comprised of *cas1* and *cas2*, works independently of the function CRISPR/Cas operon, comprised of a universal cas gene and CRISPR-associated cascade. With the presence of CRISPR/Cas systems in 50% of bacteria and 90% of archaea (Makarova *et al*., 2015) and some organisms,such as *Streptococcus thermophilus,* having multiple CRISPR/Cas operons, types, and arrays leads to question on the complexity of CRISPR/Cas systems. The goal of this study will be to analyze the CRISPR/Cas system in a model organism *Escherichia coli* to other mesophiles,

thermophiles, and psychrophiles to assess the complexity of the CRISPR/Cas operons across the classifications to maintain thermostability and effective CRISPR/Cas systems.

Purpose

The purpose of this study will be to compare the CRISPR/Cas system in a model organism *Escherichia coli* to other mesophiles, thermophiles, and psychrophiles to assess the complexity of the CRISPR/Cas systems across these bacterial classifications. We are interested in studying an environmental condition such as temperature might influence the complexity of the CRISPR-Cas system.

Methods

Dr. Wegman-Geedey provided her expertise in identifying suitable prokaryotes for this study. A literature search was performed to identify other suitable organisms for this project. In addition to using *E. coli* K12 MG1655 and *Meiothermus ruber* DSM1279, the following organisms were chosen for this project: *Salmonella enterica subsp. enterica serovar, Yersinia pseudotuberculosis, Xanthomonas albilineans, Fusobacterium hwasookii, Hymenobacter nivis, Psychrobacter, Listeria weinhenstephansis, Thermus aquaticus sp. G, Streptococcus thermophilus, Thermodesulfobium narugense, and Geobacillus stearothermophilus*. These species were chosen because they represent a diversity of temperature growth optimum (See Table 1) and their genomes are sequenced and available through GenBank. *Xanthomonas albilineans* (Willerslev *et. al*., 2004) and *Fusobacterium hwasookii* (Simon, 1977) are capable of growth across multiple temperature classifications. *X. albilineans* demonstrates a temperature growth range that classifies it as both mesophilic and psychrophilic, while *Fusobacterium hwasookii* can grow across the full spectrum.

The bioinformatics tool KEGG (Kanehisa *et al., 2019*) was used to predict the components and organization of the CRISPR-Cas system in each of the chosen species. The KEGG Brite Hierarchy ko02048 categorized genes into different CRISPR-Cas system types and subtypes. It also identified specific *cas* genes as "universal" (*cas1* and *cas2*; found in all CRISPR-Cas systems) and as "signature" genes (cas3 for the type IE; found in type-specific systems). Genes that are components of the different subtypes are also predicted. In this study, a CRISPR-Cas system type was identified by the presence of genes that encode both the "universal" and "signature" proteins. In addition, a functional/complete CRISPR-Cas operon was defined as possessing all the components of a specific subtype, based on KEGG predictions. In general, genes in GenBank that are positioned in sequential order are given sequential locus tags. However, recent renumbering of locus tags by GenBank doesn't follow this rule. Consequently, if the genes for a particular CRISPR-Cas subtype are present but do not appear to be sequentially positioned on the chromosome, then we confirmed their position using IMG/M chromosome map. Also, if an organism is not predicted to have a CRISPR-Cas system, then this was also confirmed using IMG/M.

Results

As described in the literature (see Figure 3), the mesophilic *E. coli* has a single CRISPR-Cas operon, Type IE. This is supported by a KEGG (Kanehisa *et al., 2019*) analysis and the study reviewed in the Jiang & Dounda article.

As seen in Figure 5, A KEGG analysis of *Meiothermus ruber* shows that the CRISPR/Cas system is more complex with three (3) function operons and a series of cas genes, CRISPR-associated genes, as described in Figure 5. Of the three functional operons, one is a Type 3-A operon, one a Type 3-B, and the other a Type 1-E operon similar to that of *E. coli.* Cas genes for *M. ruber* contain Cas4 and Cas6 genes present across subtypes. The presence of the multiple universal cas genes with respective cas genes is evidence for the presence of similar genes coding for the different functional CRISPR/Cas operons.

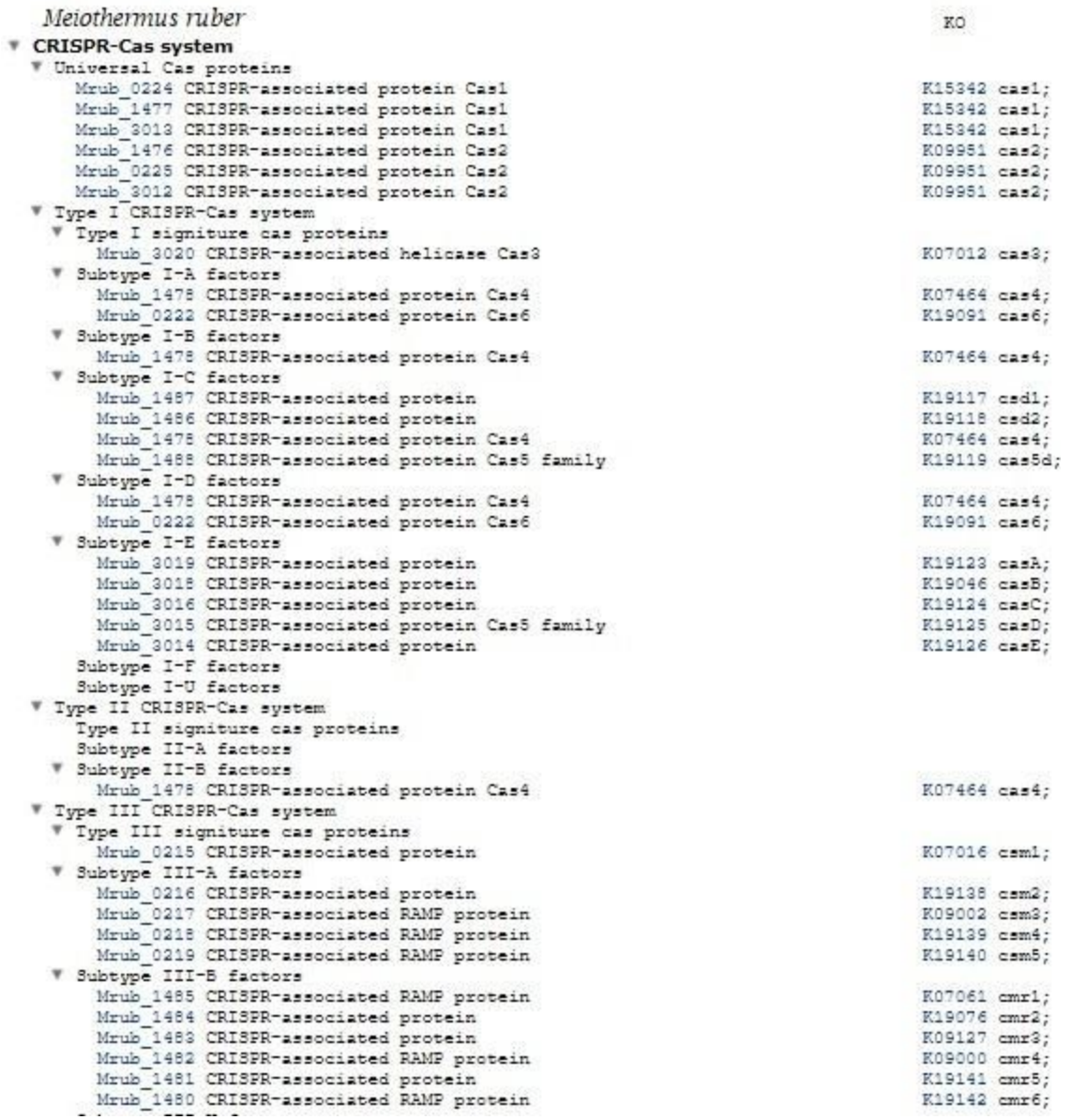

Figure 5. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Meiothermus ruber*. Evidence of four (4) functional operons. Operon 1: made up of universal Cas genes Mrub_0224 (*cas1*) and Mrub_0225 (*cas2*), Type 3 signature gene Mrub_0215 (*csm1*), Subtype 3-A genes Mrub_0216 (*csm2*), Mrub_0217 (*csm3*), Mrub_0218 (*csm4*), Mrub_0219 (*csm5*). Operon 2: universal cas genes Mrub_3013 (*cas1*) and Mrub_3012 (*cas2*), Type 1 signature gene Mrub_3020 (*cas3*), Subtype 1-E genes Mrub_3019 (*casA*), Mrub_3018 (*casB*), Mrub_3016 (*casC*), Mrub_3015 (*casD*), Mrub_3014 (*casE*). Operon 3: universal cas genes Mrub_0224 (*cas1)* and Mrub_0225 (*cas2*), Type 3 signature gene Mrub_0215 (*csm1),* Subtype 3-B genes Mrub_1485 (*cmr1*), Mrub_1484 (*cmr2*), Mrub_1483 (*cmr3*), Mrub_1482 (*cmr4*), Mrub_1481 (*cmr5*), Mrub_1480 (*cmr6*). Operon 4: universal cas genes Mrub_3013 (*cas1*) and Mrub_3012 (*cas2*), Type 1 signature gene Mrub_3020 (*cas3*), Subtype genes cas genes Mrub_1478 (*cas4*),

Mrub_1487 (*csd1*), Mrub_1486 (*csd2*), Mrub_1488 (*cas5d*). Cas genes Mrub_1477 (*cas1*), Mrub_1476 (*cas2*), Mrub_0222 (*cas6*).

X. albilineans has a temperature growth range that classifies it as mesophilic and psychrophilic. KEGG analysis of the organism (Figure 6) shows that its CRISPR/Cas system contains a functional Type 1-C operon, functional Type 1-F operon, and cas gene Cas4 that's present across subtypes. The presence of a Type 1 operon is similar to *E. coli* , which is also mesophilic.

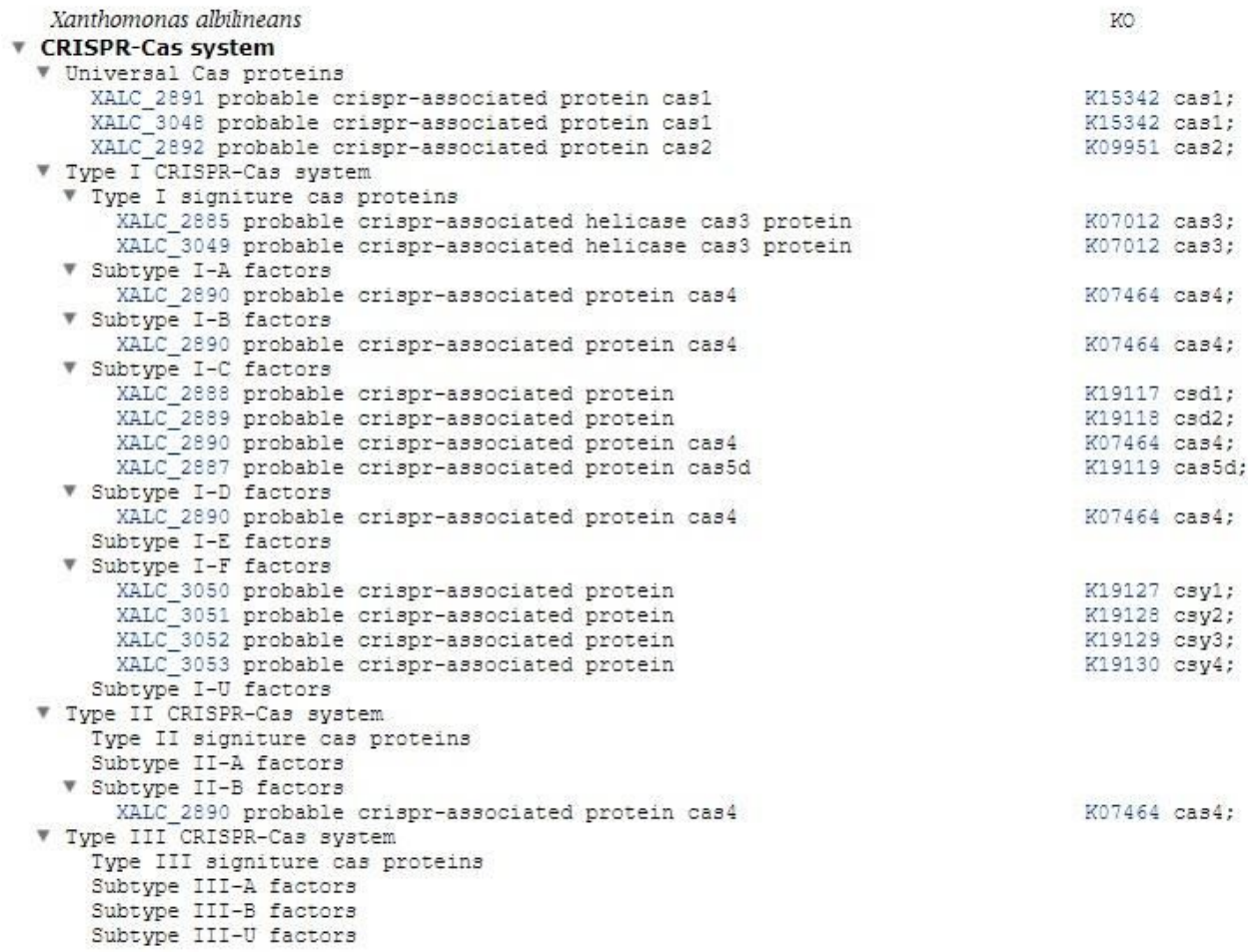

Figure 6. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Xanthomonas albilineans*. Evidence of two (2) functional operons. First made up universal Cas genes XALC_2891 (*cas1*) and XALC_2892 (*cas2*), Type 1 signature gene XALC_2885 (*cas3*), and Subtype 1-C genes XALC_2888 (*csd1*), XALC_2889 (*csd2*), XALC_2890 (*cas4*), XALC_2887 (*cas5d*). Second made up of universal Cas gene XALC_3048 (*cas1*), Type 1 signature gene XALC_3049 (*cas3*), and Subtype 1-F genes XALC_3050 (*csy1*), XALC_3051 (*csy2*), XALC_3052 (*csy3*), XALC_3053 (*csy4*).

KEGG analysis of *F. hwasookii* (Figure 7), classified in all three temperature classifications, shows evidence of a functional Type 1-A operon, functional Type 3-A operon, and cas genes *cas4* and *cas6* that are apparent across subtypes. Presence of a Type 1 operon is similar to *E. coli* and further evidence of relationship, while presence of a Type 3 operon and repeating *cas4* and *cas6* genes is similar to the thermophilic *M. ruber* (Figure 5).

Fusobacterium hwasookii	KO
CRISPR-Cas system	
Universal Cas proteins	
RN87_08950 type I-B CRISPR-associated endonuclease Cas1	K15342 cas1;
RN87_05765 CRISPR-associated protein Cas1	K15342 cas1;
RN87_08945 CRISPR-associated protein Cas2	K09951 cas2;
RN87_05770 CRISPR-associated protein Cas2	K09951 cas2;
Type I CRISPR-Cas system	
Type I signiture cas proteins	
RN87_08960 CRISPR-associated helicase Cas3	K07012 cas3;
Subtype I-A factors	
RN87_08970 type I-B CRISPR-associated protein Cas7/Cst2/DevR	K19075 cst2;
RN87_08955 CRISPR-associated protein Cas4	K07464 cas4;
RN87_08965 CRISPR-associated protein Cas5	K19090 cas5t;
RN87_08980 CRISPR-associated endoribonuclease Cas6	K19091 cas6;
Subtype I-B factors	
RN87_08955 CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-C factors	
RN87_08955 CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-D factors	
RN87_08955 CRISPR-associated protein Cas4	K07464 cas4;
RN87_08980 CRISPR-associated endoribonuclease Cas6	K19091 cas6;
Subtype I-E factors	
Subtype I-F factors	
Subtype I-U factors	
Type II CRISPR-Cas system	
Type II signiture cas proteins	
Subtype II-A factors	
Subtype II-B factors	
RN87_08955 CRISPR-associated protein Cas4	K07464 cas4;
Type III CRISPR-Cas system	
Type III signiture cas proteins	
RN87_05730 type III-A CRISPR-associated protein Cas10/Csm1	K07016 csm1;
Subtype III-A factors	
RN87_05735 type III-A CRISPR-associated protein Csm2	K19138 csm2;
RN87_05740 type III-A CRISPR-associated RAMP protein Csm3	K09002 csm3;
RN87_05745 type III-A CRISPR-associated RAMP protein Csm4	K19139 csm4;
RN87_05750 type III-A CRISPR-associated RAMP protein Csm5	K19140 csm5;
Subtype III-B factors	
Subtype III-U factors	

Figure 7. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Fusobacterium hwasookii*. Evidence of 2 (two) functional operons. First made up of universal cas genes RN87_08950 (*cas1*) and RN87_08945 (*cas2*), Type 1 signature gene RN87_08960 (*cas3*), Subtype 1-A genes RN87_08970 (*cst2*), RN87_08955 (*cas4*), RN87_08965 (*cas5t*), RN87_08980 (*cas6*). Second made up of universal cas genes RN87_05765 (*cas1*) and RN87_05770 (*cas2*), Type 3 signature gene RN87_05730 (*csm1*), Subtype 3-A genes RN87_05735 (*csm2*), RN87_05740 (*csm3*), RN87_05745 (*csm4*), RN87_05750 (*csm5*).

The remaining psychrophilic organisms - *Hymenobacter nivis* (Figure 8)*, Psychrobacter sp. G* (Figure 9,10)*, Listeria weihenstephanensis* (Figure12)-- were determined to have either one or no functional CRISPR-Cas operons. The notable piece of evidence from the KEGG analysis is that all the psychrophiles had present a repeating cas gene, *cas4*, similar to *X. albilineans* and *F. hwasookii*. The repeating *cas4* gene among all psychrophilic provides some evidence for a relationship within the classification related to the cas gene. Specific to the analysis of *Psychrobacter sp. G*, argument of a functional operon can made based on locus tag cluster similarity that exist if the need for *cas2* is excluded. Looking at the KEGG analysis of *Psychrobacter immobilis DSM 7229*, an organism in the same genus as *Psychrobacter sp. G*, you can see the presence of both universal cas genes *cas1 and cas2*. Given the relationship between the two *Psychrobacter* organisms, it is possible that the KEGG analysis *P. sp. G* is simply missing the *cas2* gene and a functional system can actually be predicted. The same argument is however void for *H. nivis* as it lacks a majority of the genes for each CRISPR/Cas system subtype. *H. nivis* also couldn't be found in IMG/M database to confirm or refute the KEGG analysis.

Hymenobacter nivis	KO
CRISPR-Cas system	
Universal Cas proteins	
DDQ68_04835 subtype I-B CRISPR-associated endonuclease Cas1	K15342 cas1;
DDQ68_04830 cas2; CRISPR-associated endonuclease Cas2	K09951 cas2;
Type I CRISPR-Cas system	
Type I signiture cas proteins	
DDQ68_04855 cas3; CRISPR-associated helicase Cas3'	K07012 cas3;
DDQ68_04875 CRISPR-associated endonuclease Cas3''	K07012 cas3;
Subtype I-A factors	
DDQ68_04840 cas4; CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-B factors	
DDQ68_04840 cas4; CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-C factors	
DDQ68_04840 cas4; CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-D factors	
DDQ68_04840 cas4; CRISPR-associated protein Cas4	K07464 cas4;
Subtype I-E factors	
Subtype I-F factors	
Subtype I-U factors	
Type II CRISPR-Cas system	
Type II signiture cas proteins	
Subtype II-A factors	
Subtype II-B factors	
DDQ68_04840 cas4; CRISPR-associated protein Cas4	K07464 cas4;
Type III CRISPR-Cas system	
Type III signiture cas proteins	
Subtype III-A factors	
Subtype III-B factors	
Subtype III-U factors	

Figure 8. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system of *Hymenobacter nivis*. Evidence of zero (0) functional operons. Presence of universal Cas genes DDQ68_04835 (*cas1*) and DDQ68_04830 (*cas2*), Type 1 signature genes DDQ68_04855 (*cas3*) and DDQ68_04875 (*cas3*), cas gene DDQ68_04840 (*cas4*).

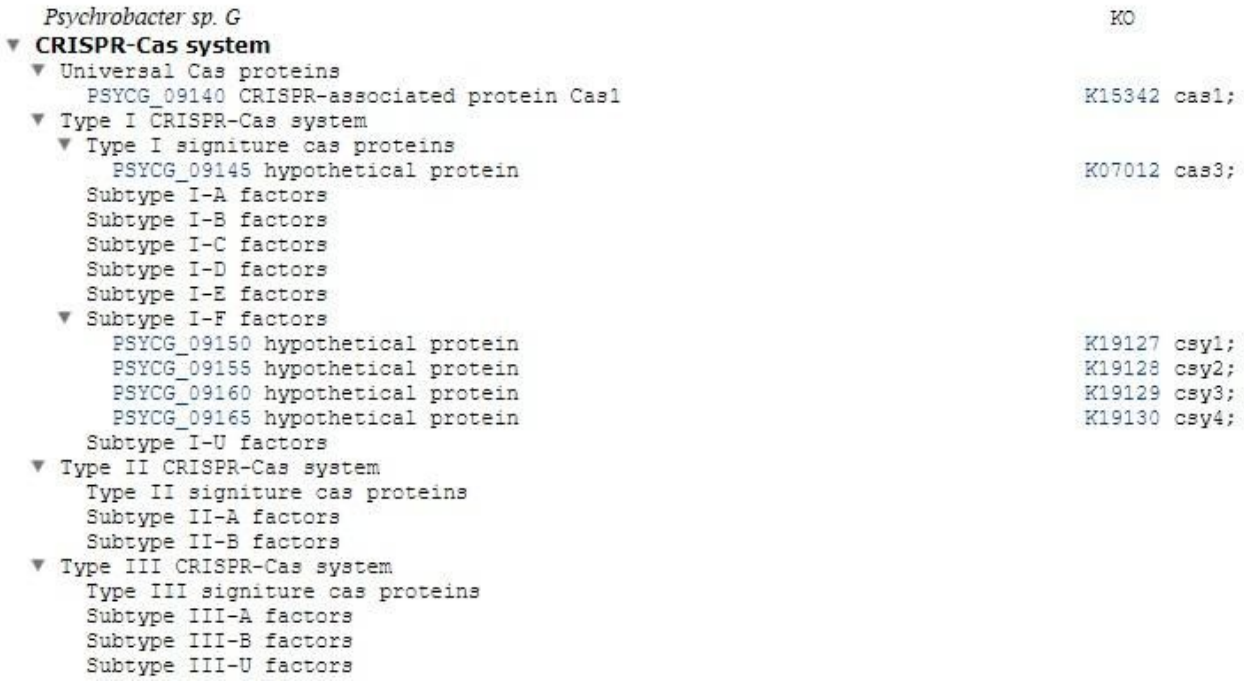

Figure 9. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Psychrobacter sp. G*. Evidence of zero (0) functional operons. Presence of universal cas gene PSYCG_09140 (*cas1*), Type 1 signature gene PSYCG_09145 (*cas3*), Subtype 1-F genes PSYCG_09150 (*csy1*), PSYCG_09155 (*csy2*), PSYCG_09160 (*csy3*), PSYCG_09165 (*csy4*).

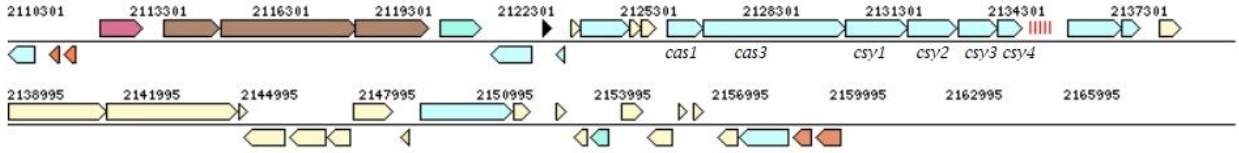

Figure 10. IMG/M (Markowitz, 2012) analysis of *Psychrobacter sp. G* confirming the presence of *cas1*, *cas3*, *csy1*, *csy2*, *csy3*, *csy4* in an operon downstream from CRISPR spacer-repeat array. The operon lacks the presence of *cas2*.

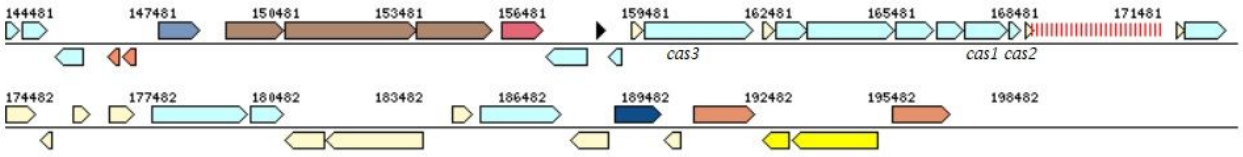

Figure 11. IMG/M (Markowitz) analysis of *Psychrobacter immobilis DSM 7229*, an organism apart of the same genus as *Psychrobacter sp. G*. Analysis shows the presence of signature gene *cas3* and universal proteins *cas1* and *cas2*.

Figure 12. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Listeria weihenstephanensis*. Evidence of one (1) functional operon made up of universal cas genes UE46_03610 (*cas1*) and UE46_03615 (*cas2*), Type 1 signature genes UE46_03600 (*cas3*), Subtype 1-A genes UE46_03585 (*cst1*), UE46_03590 (*cst2*), UE46_03605 (*cas4*), UE46_03595 (*cas5t*), UE46_03580 (*cas6*).

KEGG analysis of the organisms within the thermophilic classification predicted the presence of complex CRISPR/Cas systems based on the presence of either multiple functional operons or repeating *cas4, cas6* genes similar to *M. ruber* with *Streptococcus thermophilus* being the exception the trend. Of the organisms following the trend -- *M. ruber* (Figure 5), *F. hwasookii* (Figure 8), *T. aquaticus* (Figure 13), *T. narugense* (Figure 14), *G. stearothermophilus* (Figure 15) -- have a seemingly functional Type 1 and Type 3 operon with *G. stearothermophilus* missing a type 3 signature gene. While *G. stearothermophilus* lacks a type 3 signature gene, the Subtype 3-B gene cluster contains locus tags similar to the universal cas genes and Subtype 1-B functional operon cluster. Given the similarity between locus tag clusters, evidence for a functional Type 3 operon in *G. stearothermophilus* is provided. Including all the thermophilic organisms, excluding *S. thermophilus*, the repetition of *cas4*, *cas6* genes is evidence of relationship based on these cas genes that exist between thermophilic CRISPR/Cas systems.

Repetition of *cas4* gene across thermophiles and psychrophiles provides evidence that the cas gene is involved in the complexity or maintenance of thermostability the CRISPR/Cas system.

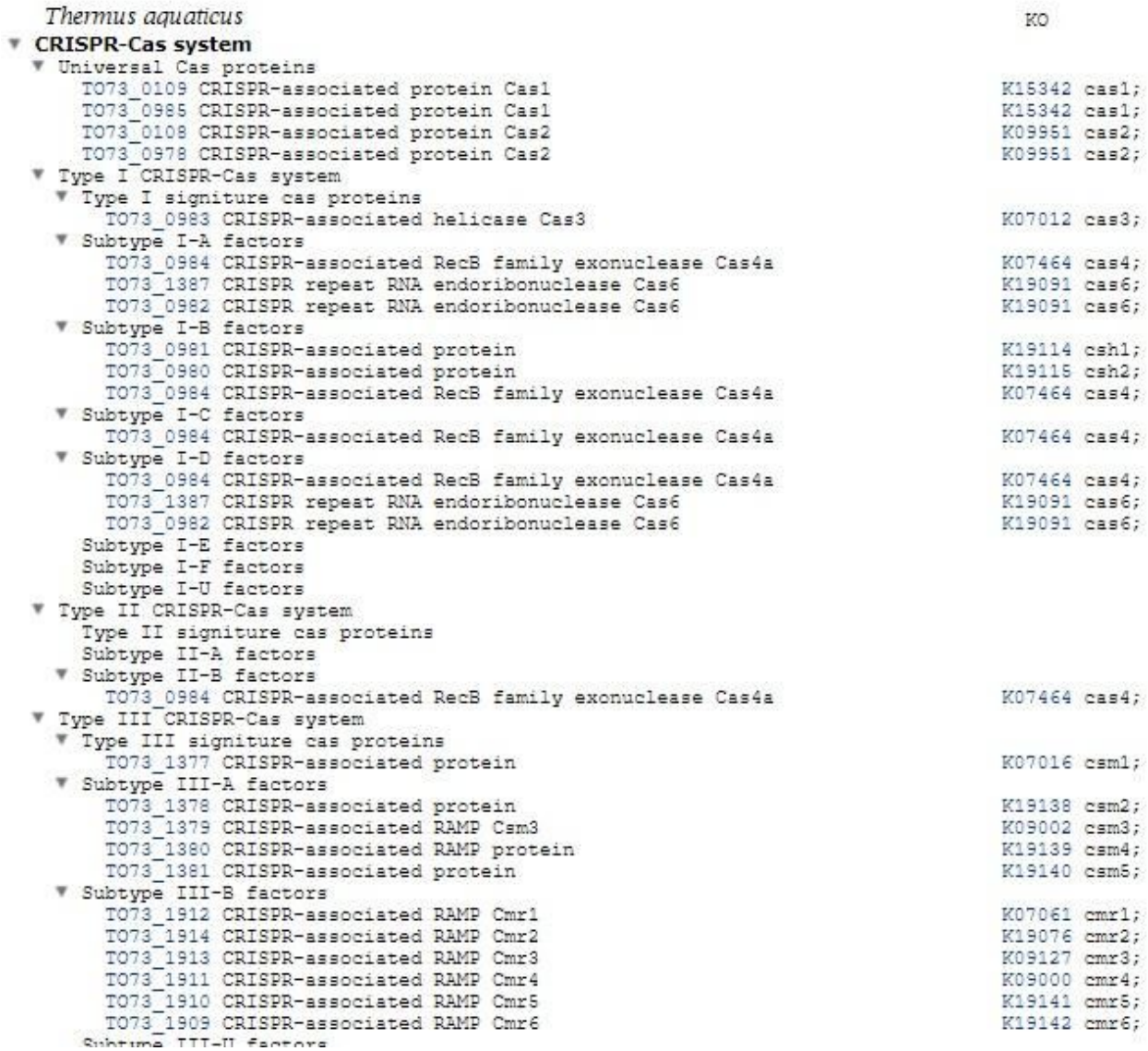

Figure 13. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Thermus aquaticus*. Evidence of three (3) functional operons. First made up of universal cas genes TO73_0109 (*cas1*) and TO73_0108 (*cas2*), Type 3 signature gene TO73_1377(*csm1*), Subtype 3-A genes TO73_1378 (*csm2*), TO79_1379 (*csm3*), TO73_1380 (*csm4*), TO73_81 (*csm5*). Second made up of universal cas genes TO73_0985 (*cas1*) and TO73_0978 (*cas2*), Type 1 signature gene TO73_0983 (*cas3*), Subtype 1-B genes TO73_0981 (*csh1*), TO73_0980 (*csh2*), TO73_0984 (*cas4*). Third made up of universal cas genes TO73_0109 (*cas1*) and TO73_0108 (*cas2*), Type 3 signature gene TO73_1377(*csm1*), Subtype 3-B genes TO73_1912 (*cmr1*), TO73_1914 (*cmr2*), TO73_1913 (*cmr3*), TO73_1911 (*cmr4*), TO73_1910 (*cmr5*), TO73_1909 (*cmr6*). Cas genes TO73_1387 (*cas6*), TO73_0982 (*cas6*).

Figure 14. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Thermodesulfobium narugusense*. Evidence of two (2) functional operons. First made up of universal cas genes Thena_0820 (*cas1*) and Thena_0821 (*cas2*), Type 1 signature gene Thena_0818 (*cas3*), Subtype 1-B genes Thena_0815 (*csh1*), Thena_0816 (*csh2*), Thena_0819 (*cas4*), Thena_0817 (*cas5h*). Second made up of universal cas genes Thena_1727 (*cas1*) and Thena_1726 (*cas2*), Type 3 signature gene Thena_1733 (*csm1*), Subtype 3-A genes Thena_1731 (*csm3*), Thena_1730 (*csm4*), Thena_1729 (*csm5*).

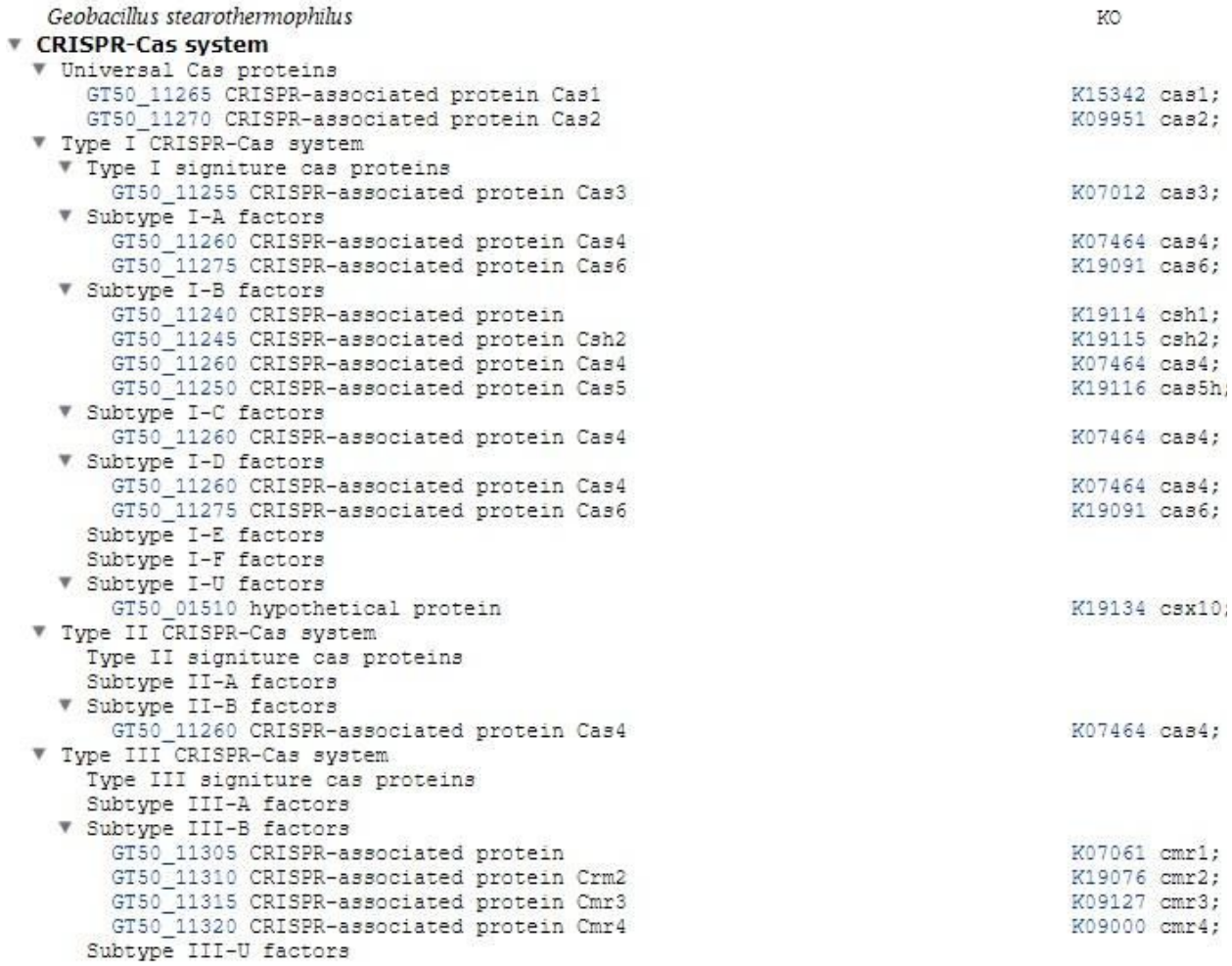

Figure 15. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Geobacillus stearothermophilus*. Evidence of one (1) functional operon made up of universal cas genes GT50_11265 (*cas1*) and GT50_11270 (*cas2*), Type 1 signature gene GT50_11255 (*cas3*), Subtype 1-B genes GT50_11240 (*csh1*), GT50_11245 (*csh2*), GT50_11260 (*cas4*), GT50_11250 (*cas5h*). Cas genes GT50_11275 (*cas6*), GT50_11305 (*cmr1*), GT50_11310 (*cmr2*), GT50_11315 (*cmr3*), GT50_11320 (*cmr4*).

KEGG analysis of *Streptococcus thermophilus* (Figure 16) in comparison to KEGG data of other thermophilic organisms studied demonstrates major differences and lack of correspondence to apparent trends. Analysis showed no evidence of a functional operon but presence of universal cas genes, signature genes for Type 2 and Type 3, and complete set of genes for Subtype 3-A. Analyzing the locus tag cluster similarities the type 3 signature gene and respective subtype genes have a corresponding universal cas gene, *cas2*. The type 2 signature gene corresponds to universal cas genes *cas1*, *cas2* based on locus tag clustering method. Presence of corresponding Type 3-A nonfunctional operon is seemingly evidence that expands upon relationship amongst CRISPR/Cas system in thermophilic bacteria based on presence of Type 3 operons.

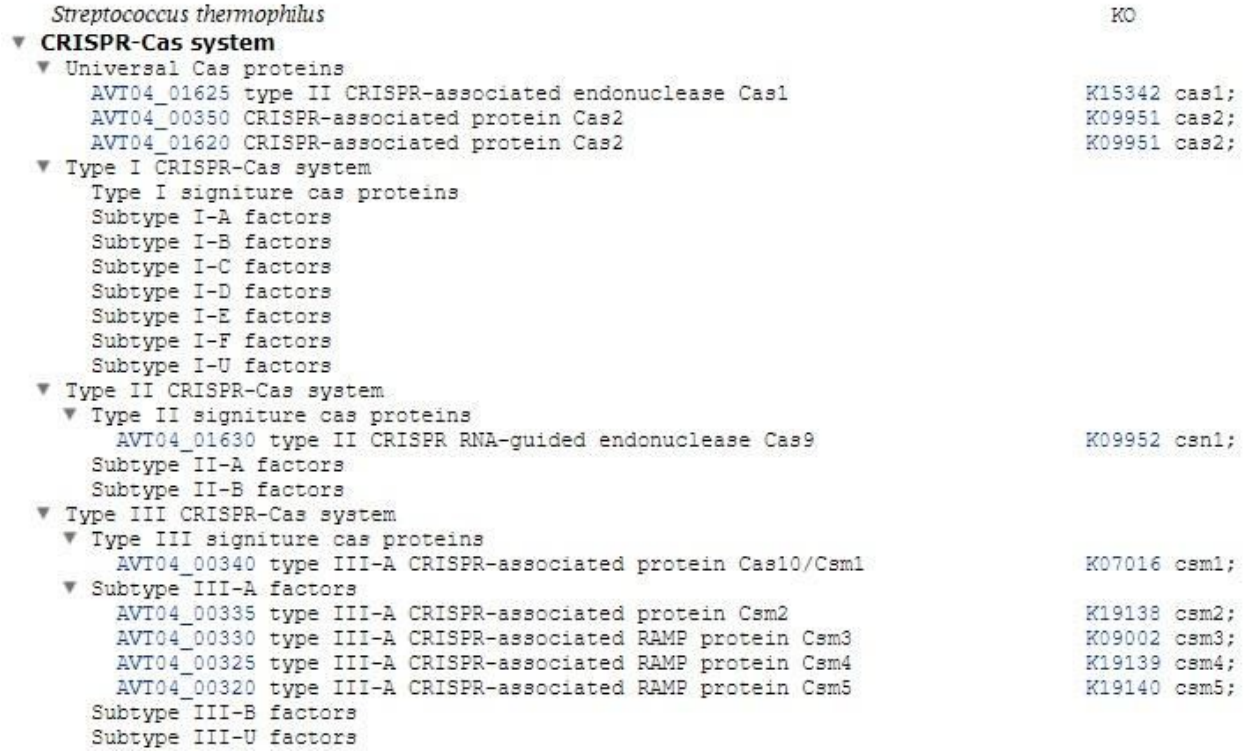

Figure 16. KEGG (Kanehisa *et al., 2019*) analysis of CRISPR/Cas system in *Streptococcus thermophilus*. Evidence of (1) functional operons made up of universal cas genes AVT04_01625 (*cas1*), AVT04_01620 (*cas2*), Type 3 signature gene AVT04_00340 (*csm1*), Subtype 3-A genes AVT04_00335 (*csm2*), AVT04_00330 (*csm3*), AVT04_00325 (*csm4*), AVT04_00320 (*csm5*). Cas genes AVT04_00350 (*cas2*), Type 2 signature gene AVT04_01630 (*cas9*).

Table 1 summarizes the data from all the organisms based on temperature classification ($N_{meosphiles}$=5, $N_{Psychrophiles}$= 5, $N_{Thermophiles}$=6).

Organism Name	Temperature range#	Predicted Number of CRISPR/Cas Operons	Predicted Number of CRISPR arrays
Hymenobacter nivis	Psychro	0	na
Psychrobacter sp. G	Psychro	1	
Psychrobacter immobilis DSM 7229**		1	1
Listeria weihenstephanensis	Psychro	1	6
Escherichia coli	Meso	1	2
Salmonella enterica subsp. enterica serovar	Meso	1	2
Yersinia pseudotuberculosis	Meso	1	2
Meiothermus ruber DSM1279	Thermo	4	6
Thermus aquaticus Y51MC23	Thermo	3	7
Streptococcus thermophilus LMD-9	Thermo	1	3
Thermodesulfobium narugense	Thermo	2	2
Geobacillus stearothermophilus	Thermo	1	1
*Xanthomonas albilineans**	Meso-Psychro	2	6
*Fusobacterium hwasookii**	Meso-Psychro-Thermo	2	3

Table 1. A comparison of the number of putative CRISPR/Cas operons and CRISPR arrays in a select group of psychrophilic, mesophilic and thermophilic prokaryotes.

#Bacteria in general grow at a wide range of temperatures: psychrophile below 20℃, mesophiles between 20-45℃, thermophiles 45-80℃, and hyperthermophiles above 80℃. (Goldstein 2007)
*Organisms demonstrating a wider temperature range for characterized growth
**Aditional organisms were added to the list based on the absence of related organisms not being found in IMG/M database. These organisms are related to the chosen organisms and thus their IMG/M (Markowitz, 2012) analysis is relevant to the the chosen organisms.

Discussion

The KEGG data, complementary IMG/M data, and temperature classifications reveal that several relationships may exist within and between classifications that alludes to the complexity in CRISPR/Cas systems that arise in psychrophiles and thermophiles compared to mesophiles such as model organism *Escherichia coli.* CRISPR/Cas systems in thermophilic bacteria demonstrate a trend of containing a components of a functional Type 3 operon. Mesophilic and and some psychrophilic bacteria ,such as *Listeria weihenstephanensis*, have components that correspond to a functional Type 1 operon. Differing from mesophiles, the CRISPR/Cas systems in psychrophiles and thermophiles seem to be related by the presence of *cas4* which is presence across subtypes of Type 1 and Type 2. Thermophiles, in addition to *cas4*, demonstrate the repetition of *cas6* which follows similar trends of presence in subtypes of Type 1 and Type 2. All studied organisms, except *S. thermophilus* and *Hymenobacter nivis*, have at least components of or a functional Type 1 operon. This data presents the possibility of evolutionary significance linked to Type 1 CRISPR/Cas systems as being shared across organisms The presence of Type 1 signature genes and cas genes also suggest the possibility of this relationship . Given the data collected however, the conclusion that complexity arises outside of thermophilic conditions can be made with a limited certainty until data is collected for a larger pool of organisms that fall within the the specified temperature classifications. In addition to conclusions based on the bioinformatic data, conclusions of functionality despite absence of specific genes can be made as the function of the operon may not be completely dependent on the presence of that specific gene (Richter *et. al*., 2012) and the biological principle that a previous study explored relating the classification and diversity of CRISPR/Cas systems to the conclusion that the high energy cost of maintaining and expressing several CRISPR/Cas genes must offer major advantages on to the cell. (Garrett *et al*., 2011). This means that the absence of a specific cas gene from prediction outputted by KEGG may be due to the lack identification of that gene somewhere else in the genome of an organisms. Analysis of *Psychrobacter sp. G* with KEGG (Figure 9) and IMG/M (Figure 10) suggest this when compared to the IMG/M analysis of *Psychrobacter immobilis DSM 7229* (Figure 11) as *cas2* is absent in the first but predicted to be present in the latter. A limitation of the methodology used, as mentioned earlier, is the renumbering of the locus tags for the genes analyzed by GenBank, a database KEGG and IMG/M pulls from. Due to this limitation, the method of predicting number of operons and functional systems based on locus tag similarities is solely applicable to analysis of the subtypes. In analyzing the entire functionality of the system, the method of identifying a complete set of universal cas genes, a type-specific signature gene, and a complete set of subtype genes without correspondence between the locus tags yielded adequate results that were reaffirmed by IMG/M. Given the results, this methodology was validated by the predicted number of CRISPR arrays which suggest that the phases of the CRISPR/Cas systems work independently of each other; *cas1 and cas2* form a functional protein*,* cas cascade genes of a subtype form a functional protein.

Literature Cited

Willerslev, E., Hansen, A.J., Ronn, R., Brand, T.B., Barnes, I., Wiuf, C., Gilichinsky, D., Mitchell, D., Cooper, A., 2004. Long-term persistence of bacterial DNA. Curr. Biol., R9eR10.

Simon PC. The effect of temperature on growth and survival of Fusobacterium necrophorum isolated from bovine liver abscesses. *Can J Comp Med*. 1977;41(2):169-73.

Brady MF, Bhimji SS. Yersinia Pseudotuberculosis. [Updated 2018 Nov 14]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2018 Jan-.https://www.ncbi.nlm.nih.gov/books/NBK430717/

Giannella RA. Salmonella. In: Baron S, editor. Medical Microbiology. 4th edition. Galveston (TX): University of Texas Medical Branch at Galveston; 1996. Chapter 21.https://www.ncbi.nlm.nih.gov/books/NBK8435/

Makarova, Kira S and Eugene V Koonin. "Annotation and Classification of CRISPR-Cas Systems" *Methods in molecular biology (Clifton, N.J.)* vol. 1311 (2015): 47-75.
Richter, Corinna et al. "Function and regulation of clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR associated (Cas) systems" *Viruses* vol. 4,10 2291-311. 19 Oct. 2012, doi:10.3390/v4102291

Jeanthon, C.; S. L'Haridon, V. Cueff, A. Banta, A. L. Reysenbach and D. Prieur. "Thermodesulfobacterium hydrogeniphilum sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a deep-sea hydrothermal vent at Guaymas Basin, and emendation of the genus Thermodesulfobacterium." International Journal of Systematic and Evolutionary Microbiology, Vol 52, 765-772, Copyright © 2002 by Society for General Microbiology.

Tindall, B. J., Sikorski, J., Lucas, S., Goltsman, E., Copeland, A., Glavina Del Rio, T., … Lapidus, A. (2010). Complete genome sequence of Meiothermus ruber type strain (21T). Standards in Genomic Sciences, 3(1), 26–36. http://doi.org/10.4056/sigs.1032748

Garrett, R. A., & Shah, S. A. (2011). CRISPR/Cas and Cmr modules, mobility and evolution of adaptive immune systems. *Research in Microbiology,162*(1), 27-38. doi:10.1016/j.resmic.2010.09.001

Euzéby JP. List of bacterial names with standing in nomenclature: A folder available on the Internet. Int J Syst Bacteriol 1997; 47:590-592. PubMed doi:10.1099/00207713-47-2-590

Nobre MF, Trüper HG, Da Costa MS. Transfer of Thermus ruber (Loginova et al. 1984), Thermus silvanus (Tenreiro et al. 1995), and Thermus chliarophilus (Tenreiro et al. 1995) to Meiothermus gen. nov. as Meiothermus ruber comb. nov., Meiothermus silvanus comb. nov., and Meiothermus chliarophilus comb. nov., respectively, and emendation of the genus Thermus. Int J Syst Bacteriol 1996; 46:604-606. doi:10.1099/00207713- 46-2-604

Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., . . . Koonin, E. V. (2011, May 09). Evolution and classification of the CRISPR–Cas systems. Retrieved from https://www.nature.com/articles/nrmicro2577

Dupuis, M., Villion, M., Magadán, A. H., & Moineau, S. (2013, July 02). CRISPR-Cas and restriction–modification systems are compatible and increase phage resistance. Retrieved from https://www.nature.com/articles/ncomms3087

Terashima, M., Ohashi, K., Takasuka, T. E., Kojima, H., & Fukui, M. (2018, November 15). Antarctic heterotrophic bacterium Hymenobacter nivis P3T displays light‐enhanced growth and expresses putative photoactive proteins. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12702

Lebedinksy, A. V., Chernyh, N. A., & Bonch-Osmolovskaya, E. A. (2007). Phylogenetic Systematics of Microorganisms Inhabiting Thermal Envirnments [Abstract]. *BIOCHEMISTRY (Moscow),72*(12), 1299-1311. Retrieved January 12, 2019.

Jiang, F., & Doudna, J. A. (2015). The Structural Biology of CRISPR-Cas Systems. *Curr Opin Struct Biol,30*, 100-111. Retrieved February 5, 2019.

Horvath, & Barrangou. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *SCIENCE,327*, 167-169. Retrieved February 5, 2019.

Wright, A., Nuñez, J., & Doudna, J. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell,164*(1-2), 29-44. doi:10.1016/j.cell.2015.12.035

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R.P., Paulsen, I., and

Karp, P.D. 2013. EcoCyc: fusing model organism databases with systems biology *Nucleic Acids Research* 41:D605-612.

Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, et al. 2012. IMG: The integrated microbial genomes database and comparative analysis system. Nucleic Acids Research 40(D1):D115-22. Available from: http://nar.oxfordjournals.org/content/40/D1/D115.full

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M.; New approach for understanding genome variations in KEGG. Nucleic Acids Res. 47, D590-D595 (2019).