

# Shining a Light on Prokaryotic Bioluminescence Pathways: A Comparative Genomics Approach

Fotini (Tina) Papazotos<sup>1,+</sup>, Mathias Renaud<sup>1,+</sup>, Nicole Wang<sup>1,+</sup>, and Debra Watkins<sup>1,+</sup>

<sup>1</sup>University of Waterloo, Dept. of Biology, Waterloo, Canada

<sup>+</sup>these authors contributed equally to this work

## Introduction

Distinct groups of prokaryotes and eukaryotes are capable of bioluminescence, a process which allows an organism to emit light through enzymatic activity.<sup>1</sup> Luciferase is the collective name for enzymes central to bioluminescence, which are capable of catalyzing visible light production, most commonly using luciferin as a substrate.<sup>1,2</sup> Chemiluminescence pathways, including bioluminescence, differ from fluorescent proteins, such as GFP, because they do not require incoming light to luminesce.<sup>3</sup> Though the specific enzymatic reactions differ, bioluminescence pathways are present in distantly related organisms across the different domains of life.<sup>4</sup> This ability is typically found in marine or host-associated environments, but free-living bioluminescent freshwater and terrestrial prokaryotes have also been isolated.<sup>4</sup>

The luciferase enzyme is commonly used in the fields of biotechnology, microscopy and in molecular biology, where it functions as a reporter gene.<sup>5</sup> In these applications, luciferase acts similarly to GFP, but only requires the substrate luciferin and ATP to be detected, instead of an external light source.<sup>6</sup> Luciferase is also used in next-generation 454 pyrosequencing, where the emitted light is an indicator of nucleotide incorporation to a growing nucleotide chain.<sup>7</sup>

Bioluminescence is not fully understood, but is thought to play roles in evading predators, detecting and luring prey, sexual selection, and quorum sensing.<sup>8</sup> Common bacterial genera that perform bioluminescence include *Vibrio*, *Aliivibrio*, *Photobacterium*, and *Photorhabdus*. The *luxCDAB(F)E* operon is the common organization of *lux* genes in most bacteria.<sup>9</sup> *luxA* and *luxB* encode the subunits of heterodimeric luciferase protein.<sup>9</sup> *luxCDE* genes code for subunits of a fatty acid reductase protein, which synthesizes aldehyde substrates that drive the bioluminescence reaction.<sup>9</sup> There are also other *lux* genes that are generally not part of the *lux* operon. *luxI* and *luxR* encode proteins that regulate expression of the *lux* operon.<sup>10</sup> *luxF*, which is found in the *lux* operon of some *Photobacterium* species, encodes flavoprotein.<sup>10</sup>

*luxG* encodes a protein which acts as a flavin reductase producing FMNH<sub>2</sub>, a substrate required for the reaction.<sup>10</sup> Other *lux* genes include *luxPQNUO* which are not found in the operon.<sup>11</sup>

*luxS* encodes the protein S-ribosylhomocysteine lyase, which plays a role in the biosynthesis of a molecule involved in signaling cell density to other members of a bacterial population.<sup>12</sup> Quorum sensing is an energy-conserving adaptation in which cells only expend the energy to express certain genes when the population density surpasses a threshold.<sup>13</sup> This gene is therefore common among both bioluminescent and non-bioluminescent bacterial species.

## Objectives

The goal of this project was to compare the genomes of prokaryotic organisms which exhibit bioluminescence or contain genes included in the pathway, and to investigate whether this parallels phylogenetic relationships between these species.

It was hypothesized that there would be homology between the pathways responsible for bioluminescence in closely related prokaryotes. It was also hypothesized that the phylogenetic relationships of the organisms would parallel the relationships between bioluminescence pathways.

## Methods

The sample of species for this project was selected from prokaryotes that have been previously identified in the literature as able to perform bioluminescence. *E. coli* was included as an out group for these analyses since it is well studied, does not perform bioluminescence, and contains only one of the *lux* genes, *luxS*. For the 22 selected species, genome assemblies were downloaded from Ensembl and NCBI Assembly. A breakdown of these species is shown in table 1.

An initial phylogenetic tree of the species was created using the online tool phyloT, which uses NCBI taxonomy to generate trees.<sup>22</sup> The Newick tree file was visualized using Archaeopteryx, a java-based tool for visualizing phylogenetic trees.<sup>23</sup>

Each genome assembly was processed through the Prokka annotation pipeline. Prokka takes in assembled bacterial genomes in FASTA format and runs five different annotation programs on the sequences.<sup>24</sup> The main program of interest to this study was Prodigal, which annotates protein coding

**Table 1.** Sample of Prokaryotic Species

Species Name	Accession Number	Length	N50	Environment
<i>Aliivibrio fischeri</i>	ASM1180v1	4,273,718	2,897,536	Marine, Symbiont <sup>14</sup>
<i>Aliivibrio logei</i>	ASM169105v1	4,638,003	302,500	Symbiont <sup>14</sup>
<i>Aliivibrio salmonicida</i>	ASM19649v1	4,655,660	3,325,165	Symbiont <sup>15</sup>
<i>Aliivibrio wodanis</i>	AWOD1	4,635,126	3,003,353	Symbiont <sup>4</sup>
<i>Escherichia coli</i> (K12)	ASM584v2	4,641,652	4,641,652	Symbiont <sup>16</sup>
<i>Photobacterium aquimaris</i>	ASM167607v1	4,382,175	177,782	Marine <sup>4</sup>
<i>Photobacterium damsela</i>	ASM167614v1	4,457,113	176,840	Symbiont <sup>4</sup>
<i>Photobacterium kishitanii</i>	CFSAN029432_01.0	4,709,666	174,214	Symbiont <sup>14</sup>
<i>Photobacterium leiognathi</i>	CFSAN029437_01.0	4,607,317	87,333	Marine, Symbiont <sup>14</sup>
<i>Photobacterium phosphoreum</i>	CFSAN029438_01.0	4,475,676	83,930	Symbiont <sup>14</sup>
<i>Candidatus Photodesmus katoptron</i>	ASM47868v1	1,015,921	555,626	Marine, symbiont <sup>17</sup>
<i>Photorhabdus asymbiotica</i>	ASM19647v1	5,094,138	5,064,808	Symbiont <sup>4</sup>
<i>Photorhabdus luminescens</i>	GCA_000767775.1	5,596,724	186,732	Symbiont <sup>4</sup>
<i>Shewanella woodyi</i>	ASM1952v1	5,935,403	5,935,403	Marine, Symbiont <sup>4</sup>
<i>Vibrio azureus</i>	ASM46716v1	4,716,275	60,692	Marine <sup>18</sup>
<i>Vibrio campbellii</i>	ASM77265v1	5,652,224	947,033	Marine <sup>19</sup>
<i>Vibrio cholerae</i>	GFC_11	4,012,216	80,785	Freshwater, Marine <sup>14</sup>
<i>Vibrio harveyi</i>	ASM83407v1	5,864,537	162,233	Marine, Symbiont <sup>14</sup>
<i>Vibrio mediterranea</i>	ASM235813v1	5,945,836	246,425	Marine <sup>20</sup>
<i>Vibrio orientalis</i>	ASM17623v1	4,698,244	2,361,996	Marine <sup>14</sup>
<i>Vibrio sagamiensis</i>	ASM40042v1	4,557,580	50,295	Marine <sup>21</sup>
<i>Vibrio splendidus</i>	ASM130621v1	5,895,444	96,546	Marine <sup>15</sup>

Prokaryotic species used for comparative genome analysis of bioluminescence pathway. The quality of each individual genome assembly was determined by comparing genome length to the N50 value.

sequences (CDSs).<sup>25</sup> The protein multi-FASTA output file (.faa) generated by Prodigal was used as the input for further analysis. This file contains each of the predicted CDSs. Additionally, the Genbank files (.gbk) were used for synteny analysis in Mauve in order to incorporate annotations into the aligned genomes.

The protein multi-FASTA file for each species was run through the pfastGO pipeline. This pipeline is primarily for annotating bacterial protein sequences with GO terms. The input is a protein multi-FASTA file, which is used to perform BLASTp against Swiss-Prot to find homologs to the CDSs. PfamScan is also employed to search for conserved protein domains that may not have known homologs. The results from these two searches are used to transfer homology, GO terms, and KEGG IDs to each of the CDSs.<sup>26</sup> Over 9 million homologs across the 22 genomes were narrowed down to just under 5000 by constraining

to genes named “lux” found within the same genus as the query species.

To look at differences in the complete sets of GO terms between the species, eggNog-mapper was used on the genome assemblies. This web-based tool provides the frequency of each GO term in a genome.<sup>27</sup> The two heat maps for the *lux* gene and GO term tables were generated in R using the heatmap.2 function from the gplots package.

Lastly, a synteny analysis was done using Mauve. Mauve allows for alignment of multiple genomes, which was used to compare differences in *lux* operon structure between the genera.<sup>28</sup> All scripts and data used for this project can be found at [github.com/MathiasRenaud/BIOL469-Project](https://github.com/MathiasRenaud/BIOL469-Project).

## Results & Discussion

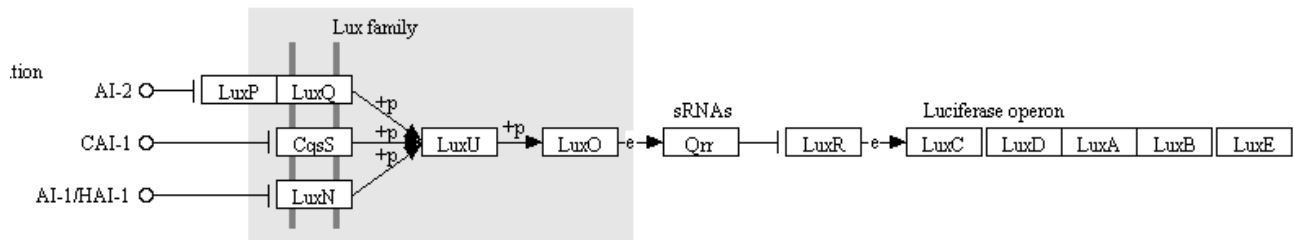
Literature searches were performed to identify genes involved in the bioluminescence pathways of both prokaryotic and eukaryotic species. A preliminary comparison of the amino acid sequences of eukaryotic and prokaryotic luciferase proteins was then conducted using Blastp searches. These Blastp results, shown in table 2, yielded low sequence identity and high E-values, indicating that bioluminescence pathways of prokaryotes and eukaryotes are likely not homologous, and possibly arose by convergent evolution. Due to the low similarity observed between prokaryotic and eukaryotic luciferase proteins, only bacterial bioluminescence genes were analyzed in this project.

**Table 2.** Comparison of Eukaryote and Prokaryote Luciferases

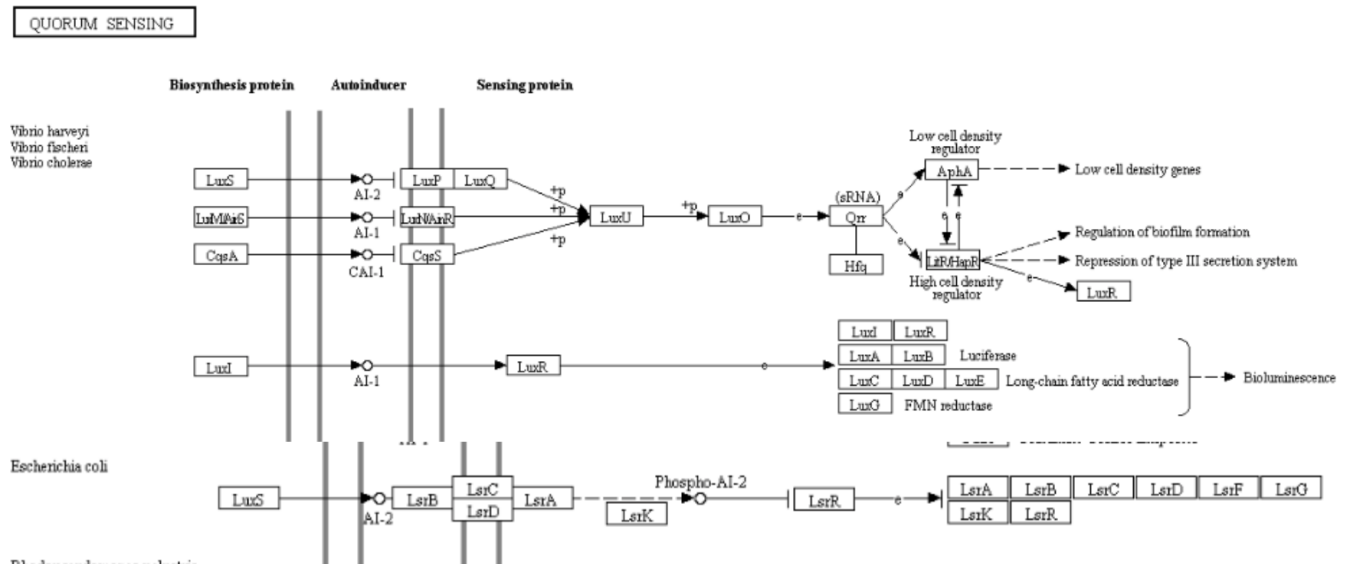
Eukaryotic Protein	Bacterial Protein	Sequence Identity	E value
Fire Fly Luciferase (Q27758)	<i>V.harveyi</i> ( <i>LuxA</i> ) Luciferase alpha subunit (B2XS31)	38%	8.3
Dinoflagellate Luciferase (O77206)	<i>Photobacterium phosphoreum</i> Beta luciferase subunit (Q52099)	26%	2
<i>Photinus pyralis</i> Luciferase (A3KBZ8)	<i>Vibrio campbellii</i> ( <i>LuxB</i> ) Luciferase subunit B (A8JPA2)	26%	0.69

Preliminary comparison of prokaryotic and eukaryotic luciferase genes. Conducted using Blastp, to find sequence identity and E values between representative prokaryotic and eukaryotic luciferases.

22 bacteria were chosen to be analyzed based on a literature search of prokaryotes that exhibit bioluminescence. Several common genera were found to carry out the light-producing pathway, including *Vibrio*, *Aliivibrio*, *Photobacterium*, and *Photorhabdus*. Two pathways were found to be associated with luciferases upon a KEGG Pathway database search: a two-component regulatory pathway (figure 1), and a



**Figure 1.** The Kegg Pathway of two-component system by the *lux* family.<sup>11</sup>



**Figure 2.** The KEGG Pathway of Quorum Sensing by the *lux* family.<sup>11</sup>

quorum sensing pathway (figure 2).<sup>29</sup> Both pathways were further investigated and it was found that the two-component system involving the luciferase family plays a role in regulation of the quorum sensing pathways.<sup>30</sup>

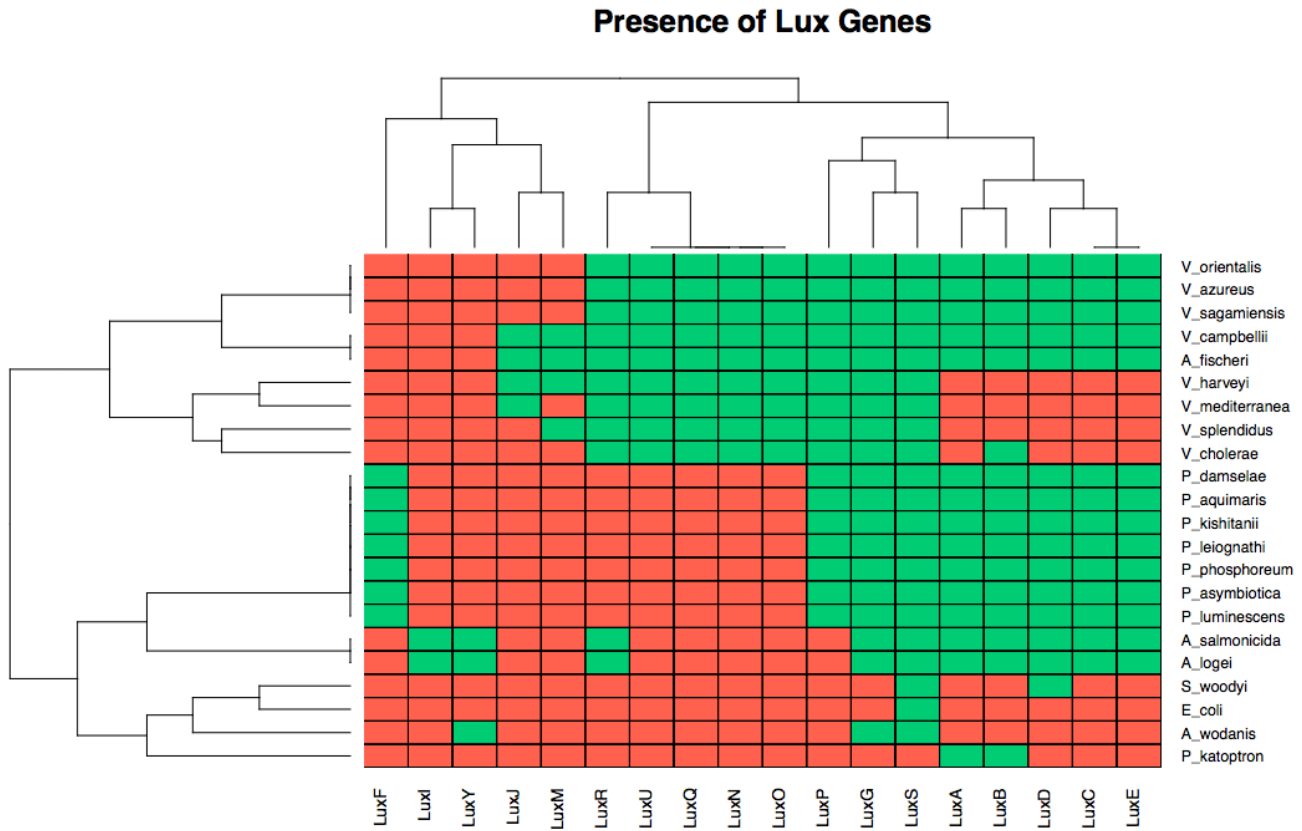
To examine relationships between bioluminescent species, a phylogenetic tree of the 22 chosen bacterial genomes was computed based on NCBI taxonomy. This tree was then used to compare against the functional relationships between these same species. Figure 3 shows the phylogenetic tree based on NCBI taxonomy. In this phylogeny-based tree, *S. woodyi* was an outgroup.

The *lux* gene heatmap in figure 4 shows the presence or absence of each *lux* gene in each species. The trees displayed in this figure were constructed based on similarities in *lux* gene sets. Tree A clusters the *lux* genes commonly present together in a genome. It is observed that the genes of the *luxCDABE* operon are commonly found together, and that *lux N*, *O*, *Q*, *R*, *U* are all commonly found in the same bacterial genomes. The groupings of the *lux* genes imply their functional importance to each other for



**Figure 3.** Phylogenetic tree derived from NCBI taxonomy. *E. coli* was included in the analysis as an outgroup, but was actually determined to cluster with *Photorhabdus*. The phyloT tree building program was used to develop the tree based on NCBI taxonomy.

performance of the pathways. In tree B, the species clustered into two very distinct branches, based on their complement of *lux* genes. One branch is composed almost entirely of the *Vibrio* species, while the other branch includes all other analyzed genera. The crucial difference between the two main branches of the tree is the presence of *lux N*, *O*, *Q*, *R*, *U* in the *Vibrio* genus. These genes are responsible for regulating expression of the *luxCDABE* operon through a two-component system, as outlined in figure 1. This indicates that *Vibrio* has evolved unique mechanisms for regulating bioluminescence. One *Aliivibrio* species, *A. fischeri*, was found to contain the *lux N*, *O*, *Q*, *R*, *U* genes, and therefore cluster with the *Vibrio* species, rather than with its own genus. Until 2007, *A. fischeri* was considered to be a member of the *Vibrio* genus, due to similarity between physical characteristics, although classification by 16S rRNA places it within the *Aliivibrio*.<sup>31</sup> This disagreement in classification is interesting, and further research is

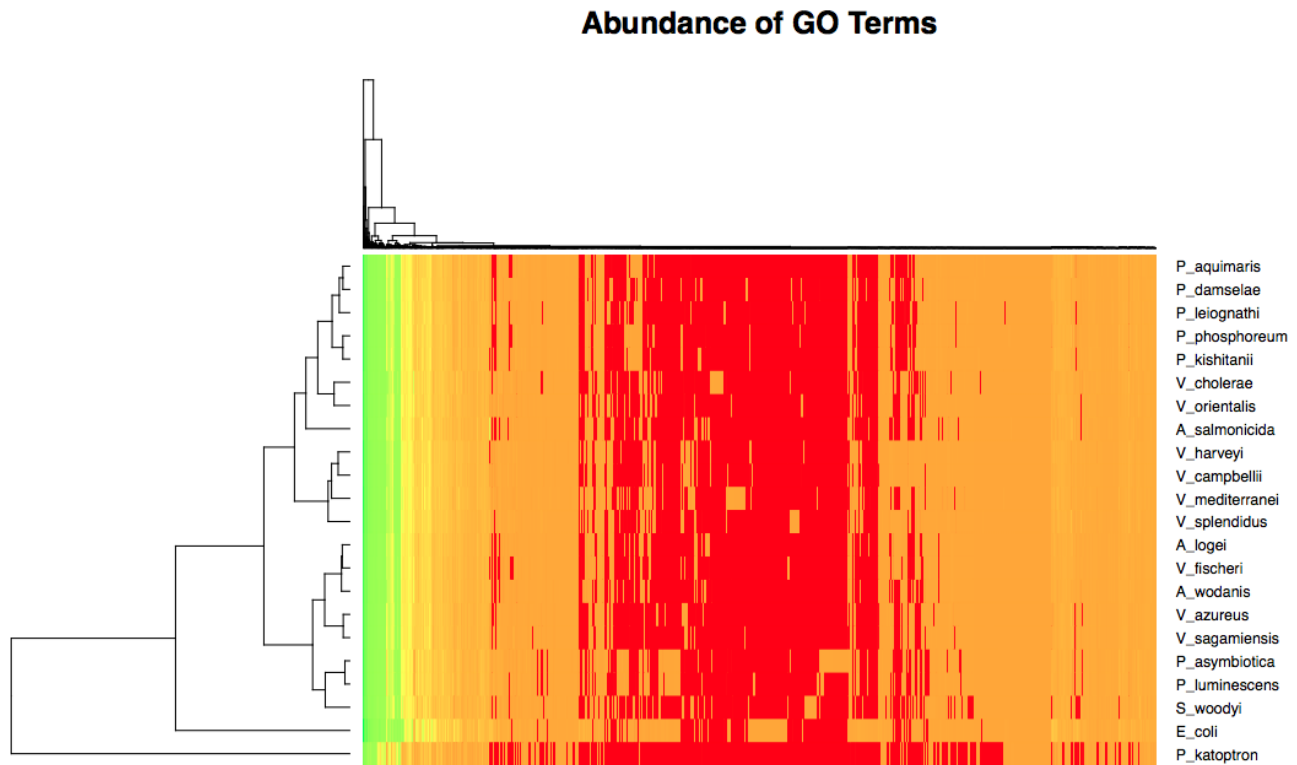


**Figure 4.** Plotting the presence of *lux* genes in bacteria. For each bacterial species, any *lux* gene present in their genome is coloured green, and the *lux* genes absent are coloured red. The 22 bacteria are also clustered into a tree based on their *lux* gene sets (B).

necessary to clarify the evolutionary history of this species, as well as evaluate horizontal gene transfer as a potential mechanism of gaining these *Vibrio*-specific genes.

The tree derived from NCBI taxonomy, showed *Shewanella* and *Photorhabdus* species to be the most distantly related to the *Photobacterium*, *Aliivibrio*, and *Vibrio* genera, which are members of the *Vibrionaceae* family.<sup>32</sup> Research by Urbanczyk et al. presents the hypothesis that *Shewanella* and *Photorhabdus* acquired the *lux* genes from *Aliivibrio* through horizontal transfer. If this hypothesis is correct, it would explain how these distantly related genera evolved the ability to bioluminesce.<sup>32</sup>

Figure 5 was constructed using annotations from eggNOG-mapper, and is based on similarities between GO term sets of each prokaryotic species. In this tree, *Candidatus photodesmus katoptron* appears to be the most distantly related. This differs from the organization of the tree based on NCBI taxonomy, in which *S. woodyi* was most distantly related. *Candidatus photodesmus katoptron* is a luminous symbiont of flashlight fish, and many of the functions required for sustaining life may be provided by the host.<sup>17</sup> The nature of



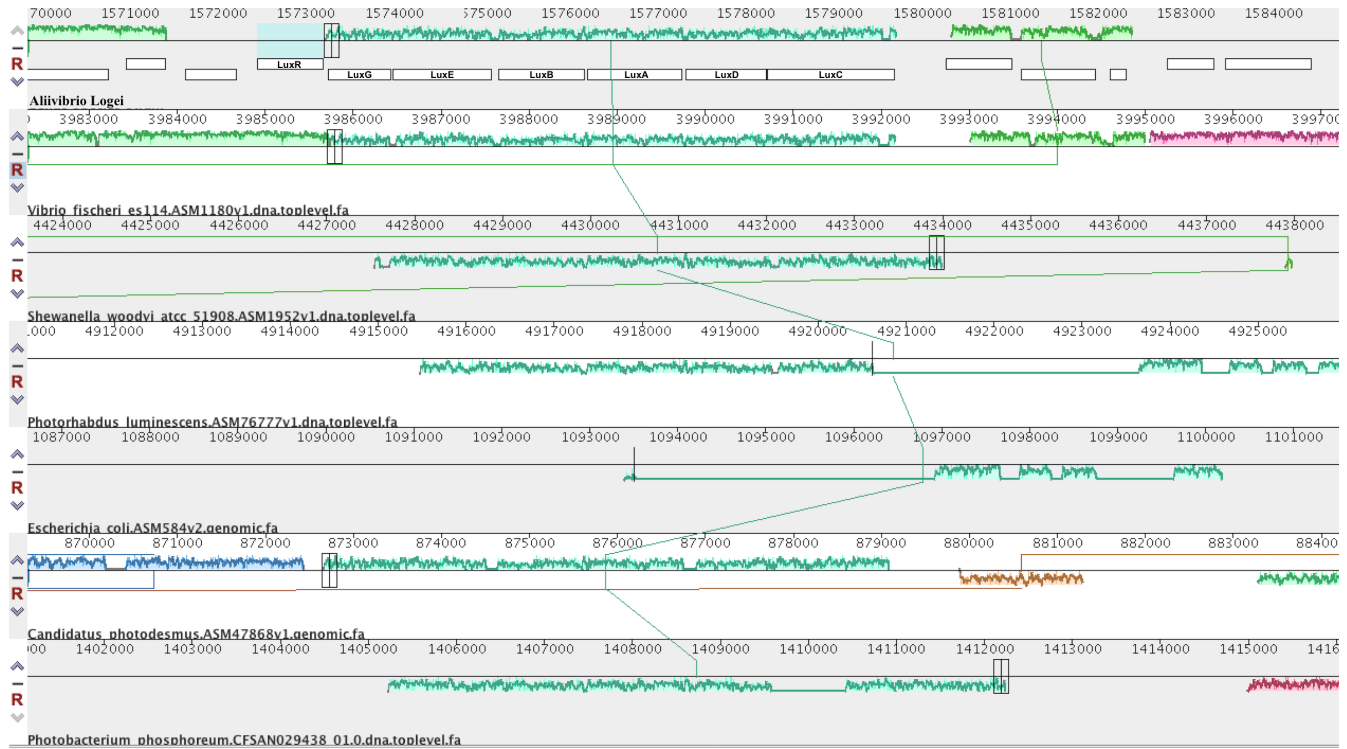
**Figure 5.** Heat map of GO Terms from eggNOG-mapper for the 22 bacterial genomes. The X-axis lists over 6000 GO Terms, and the Y axis lists the species analyzed in this report. Red indicates GO terms that were absent, whereas green represents over 100 genes pertaining to a GO term.

this specific interaction may differ greatly from the other symbiotic bioluminescent prokaryotes, which could result in the difference in functional distribution, and therefore GO term sets, between *Candidatus photodesmus katoptron* and the other analyzed species. *Candidatus photodesmus katoptron* has been found to be difficult to culture in a laboratory setting, due to it being an obligate symbiont and requiring the stringent conditions of its host. These factors make the nature of this symbiotic relationship difficult to characterize and compare.<sup>33</sup> This result differs from the expected. It was initially expected that *E. coli* would be the outgroup, as it does not exhibit bioluminescence or the respective *lux* operon. However some strains of *E. coli* are also involved in a mutually beneficial symbiotic relationship, as many of the other selected species are.<sup>34</sup>

Further comparison between the trees in figure 3 and figure 5 reveals slightly different branching schemes, which indicates that the bacteria's evolutionary relationships does not exactly correlate with functional, metabolic similarities. The two *Photorhabdus* species were seen to be more closely related to



the *E. coli* outgroup on the NCBI taxonomy tree, but on the GO term-derived tree *Photorhabdus* was more closely grouped to other bioluminescent bacteria. In the phylogenetic tree derived from NCBI taxonomy, the genera *Vibrio* and *Aliivibrio* are found in very distinct groups, while in the tree derived from GO term similarity, species of these genera are intermixed, indicating overall functional similarities.



**Figure 6.** Visualization of *lux* operon synteny in 7 representative species. The *luxCDABE* operon of *A. logei* was mapped to *A. fischeri*, *S. woodyi*, *P. luminescens*, *E. coli*, *Candidatus Photodesmus katoptron*, and *P. phosphoreum*, and shows a high degree of synteny in this operon in most of the organisms.

The *lux C*, *D*, *A*, *B*, and *E* genes are known to make up the *lux* operon, and were found to be present in the majority of analyzed species upon gene set comparison.<sup>10</sup> A representative species from each genus analyzed in this project was selected for whole-genome synteny comparison, which revealed that the order of these *lux* genes, and therefore the organization of the *lux* operon, was conserved among the six bioluminescent genera. The position of the operon in each genome differs, but the high level of conservation of the operon's structure implies functional importance of the *lux* operon, which is shown in figure 6. The only species shown to have no sign of the operon was *E. coli*, which was expected.

## Conclusions

To summarize, prokaryotic bioluminescence pathways are not well understood. The *lux* operon plays a crucial role in the pathway, with *luxA* and *luxB* encoding the heterodimeric luciferase protein. It is possible that eukaryotic luciferase proteins arose through convergent evolution, although further analysis is required to support this. The *luxCDABE* operon displays syntenic gene order indicating functional importance of each of the components in the bioluminescence pathway. It appears that the *Vibrio* genus employs a unique two component pathway involving *Lux N, O, Q, R, U* genes to regulate bioluminescence. These genes do not appear to be highly represented in the other genera we analyzed. When comparing the branching schemes of the phylogenetic tree and and GO term tree, it was determined that functional relationships between species do not parallel the phylogenetic relationships. Branching of functional relationships is likely due to other factors such as specific symbiotic relationships. It is also thought that the bioluminescent capabilities of some genera, such as *Shewanella* and *Photorhabdus*, are most likely due to horizontal gene transfer from other bioluminescent species. Overall, bioluminescence has large functional importance in many species, multiple applications in molecular biology research, and possibly many other uses yet to be discovered.

## Challenges

The analyses performed for this project came with many challenges and complications. One limitation was the computational power of the devices used to carry out analyses, especially the genome annotations. Functional annotation of 22 prokaryotic genomes using pFastGO required over 40 hours of continuous analysis, which was made more manageable by dividing the analysis among four computers. The time-consuming nature of these genomic analyses limited the number of genomes that could be investigated.

The specific computational methods chosen for this project can make the task of interpreting derived results difficult. The use of different pipelines for genome annotation may yield different results, and some pipelines may be more effective at identifying and annotating every gene in the genome.<sup>35</sup> As a result of this limitation, there may be *lux* genes or *lux*-like genes present in the analyzed genomes which were not annotated by pfastGO. This lowers the confidence of the obtained results, as well as making it difficult to

compare these results to those in the literature that were derived using alternate pipelines.

As bioluminescence is not well understood, all genes critical to this process may not have been identified and characterized. Bioluminescence has also been studied extensively in some species (ie. *Aliivibrio fischeri*), but very little in others, which makes comparing bioluminescent pathways between species difficult, due to this sequencing bias.

## Future Directions

To further investigate bioluminescence pathways and their functional significance, more detailed analyses should be performed. The percentage of genes in each bacterial genome dedicated to bioluminescence and its related pathways could be investigated through construction of functional gene profiles. This would provide a better indication of how functionally important bioluminescence is to members of different genera. This is of specific interest in deep sea marine species where light acts as the major form a communication.<sup>36</sup>

There appears to be potential involvement of *lux*-like genes in the bioluminescence pathway. These genes should be investigated to determine how their sequences and functionality are similar or different from *lux* genes. Specifically, species which have been observed to perform bioluminesce but do not appear to contain luciferase genes, such as *luxA* or *luxB* should be examined.

The genes and bioluminescence pathways are different in prokaryotes and eukaryotes. This project focuses on prokaryotic bioluminescence, but leaves the eukaryotic mechanisms to be investigated. Firefly luciferase is the most well-known eukaryotic luciferase, but the whole genome of the firefly has yet to be fully sequenced and assembled due to the abundance of highly variable repetitive elements.<sup>37</sup> Once the firefly genome information, as well as genome information of other bioluminescent eukaryotes is published, a comparative genomics analysis of these species would be informative.

## References

1. Simone M Marques and Joaquim CG Esteves da Silva. Firefly bioluminescence: a mechanistic approach of luciferase catalyzed reactions. *IUBMB life*, 61(1):6–17, 2009.
2. J Woodland Hastings. Chemistries and colors of bioluminescent reactions: a review. *Gene*, 173(1):5–11, 1996.
3. J Marshall, J Kent, and T Cronin. Visual adaptations in crustaceans: spectral sensitivity in diverse habitats. In *Adaptive mechanisms in the ecology of vision*, pages 285–327. Springer, 1999.
4. T Schmidt and M Schaechter. *Topics in ecological and environmental microbiology*. Elsevier, USA, 1 edition, 2012.
5. F Stephenson. *Calculations for molecular biology and biotechnology: A guide to mathematics in the laboratory*. Elsevier, USA, 1 edition, 2003.
6. NCA De Ruijter, J Verhees, W Van Leeuwen, and AR Van der Krol. Evaluation and comparison of the gus, luc and gfp reporter system for gene expression studies in plants. *Plant Biology*, 5(02):103–115, 2003.
7. R Dubey. *Advanced biotechnology*. S. Chand & Company Pvt. Ltd., New Delhi, 1 edition, 2014.
8. Edith A Widder. Bioluminescence in the ocean: origins of biological, chemical, and ecological diversity. *Science*, 328(5979):704–708, 2010.
9. Edward A Meighen. Molecular biology of bacterial bioluminescence. *Microbiological reviews*, 55(1):123–142, 1991.
10. Edward A Meighen. Bacterial bioluminescence: organization, regulation, and application of the lux genes. *The FASEB journal*, 7(11):1016–1022, 1993.
11. KEGG Pathway. Quorum sensing: map02024. [http://www.genome.jp/dbget-bin/www\\_bget?pathway:map02024](http://www.genome.jp/dbget-bin/www_bget?pathway:map02024).
12. UniprotKB. LUXS\_ECOLI. <http://www.uniprot.org/uniprot/P45578>.

13. Stephan Schauder, Kevan Shokat, Michael G Surette, and Bonnie L Bassler. The luxs family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. *Molecular microbiology*, 41(2):463–476, 2001.
14. Anthony H. Rose and John Frome Wilkinson. *Advances in microbial physiology*, volume 34. Academic Press, Great Britain, 1993.
15. Erik Hjerde, Marit Sjo Lorentzen, Matthew TG Holden, Kathy Seeger, Steinar Paulsen, Nathalie Bason, Carol Churcher, David Harris, Halina Norbertczak, Michael A Quail, et al. The genome sequence of the fish pathogen *aliivibrio salmonicida* strain Ifi1238 shows extensive evidence of gene decay. *BMC genomics*, 9(1):616, 2008.
16. Olivier Tenaillon, David Skurnik, Bertrand Picard, and Erick Denamur. The population genetics of commensal *escherichia coli*. *Nature Reviews Microbiology*, 8(3):207–217, 2010.
17. Tory A Hendry, Jeffrey R de Wet, Katherine E Dougan, and Paul V Dunlap. Genome evolution in the obligate but environmentally active luminous symbionts of flashlight fish. *Genome biology and evolution*, 8(7):2203–2213, 2016.
18. Susumu Yoshizawa, Minoru Wada, Kumiko Kita-Tsukamoto, Eiko Ikemoto, Akira Yokota, and Kazuhiro Kogure. *Vibrio azureus* sp. nov., a luminous marine bacterium isolated from seawater. *International journal of systematic and evolutionary microbiology*, 59(7):1645–1649, 2009.
19. Graciela M Dias, Cristiane C Thompson, Brian Fishman, Hiroaki Naka, Margo G Haygood, Jorge H Crosa, and Fabiano L Thompson. Genome sequence of the marine bacterium *vibrio campbellii* ds40m4, isolated from open ocean water. *Journal of bacteriology*, 194(4):904–904, 2012.
20. Eva Tarazona, Teresa Lucena, David R Arahal, M Carmen Macián, María A Ruvira, and María J Pujalte. Multilocus sequence analysis of putative *vibrio mediterranei* strains and description of *vibrio thalassae* sp. nov. *Systematic and applied microbiology*, 37(5):320–328, 2014.
21. Jesús L Romalde, Ana L Diéguez, Aide Lasa, and Sabela Balboa. New *vibrio* species associated to molluscan microbiota: a review. *Frontiers in microbiology*, 4:413, 2014.
22. I Letunic. phyloT: phylogenetic tree generator. <http://phylot.biobyte.de/>, 2015.

23. Mira V Han and Christian M Zmasek. phyloXML: XML for evolutionary biology and comparative genomics. *BMC bioinformatics*, 10(1):356, 2009.
24. Torsten Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068–2069, 2014.
25. Doug Hyatt, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11(1):119, 2010.
26. D Wüthrich. pfastGO. <https://github.com/danielwuethrich87/pfastGO>, 2016.
27. Jaime Huerta-Cepas, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering, and Peer Bork. Fast genome-wide functional annotation through orthology assignment by eggno-mapper. *Molecular biology and evolution*, page msx148, 2017.
28. Aaron E Darling, Bob Mau, and Nicole T Perna. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one*, 5(6):e11147, 2010.
29. Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, and Minoru Kanehisa. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 27(1):29–34, 1999.
30. Jeremy A Freeman and Bonnie L Bassler. Sequence and function of LuxU: a two-component phosphorelay protein that regulates quorum sensing in *vibrio harveyi*. *Journal of bacteriology*, 181(3):899–906, 1999.
31. Henryk Urbanczyk, Jennifer C Ast, Melissa J Higgins, Jeremy Carson, and Paul V Dunlap. Reclassification of *vibrio fischeri*, *vibrio logei*, *vibrio salmonicida* and *vibrio wodanis* as *aliivibrio fischeri* gen. nov., comb. nov., *aliivibrio logei* comb. nov., *aliivibrio salmonicida* comb. nov. and *aliivibrio wodanis* comb. nov. *International journal of systematic and evolutionary microbiology*, 57(12):2823–2829, 2007.

32. Henryk Urbanczyk, Jennifer C Ast, Allison J Kaeding, James D Oliver, and Paul V Dunlap. Phylogenetic analysis of the incidence of lux gene horizontal transfer in vibrionaceae. *Journal of bacteriology*, 190(10):3494–3504, 2008.
33. Tory A Hendry and Paul V Dunlap. The uncultured luminous symbiont of anomalops katoptron (beryciformes: Anomalopidae) represents a new bacterial genus. *Molecular phylogenetics and evolution*, 61(3):834–843, 2011.
34. VA Gritsenko and OV Bukharin. The ecological and medical aspects of the symbiosis between escherichia coli and man. *Zhurnal mikrobiologii, epidemiologii, i immunobiologii*, (3):92–99, 2000.
35. Roland J Siezen and Sacha AFT Van Hijum. Genome (re-) annotation and open-source annotation pipelines. *Microbial biotechnology*, 3(4):362–369, 2010.
36. Jean-François Rees, Bertrand De Wergifosse, Olivier Noiset, Marlene Dubuisson, Bernadette Janssens, and Eric M Thompson. The origins of marine bioluminescence: turning oxygen defence mechanisms into deep-sea communication tools. *Journal of Experimental Biology*, 201(8):1211–1221, 1998.
37. Sarah Sander Lower, J Spencer Johnston, Kathrin Stanger-Hall, Carl E Hjelman, Shawn J Hanrahan, Katharine Korunes, and David Hall. Genome size in north american fireflies: Substantial variation likely driven by neutral processes. *Genome biology and evolution*, page evx097, 2017.