Phenotypic pan- and core-genome characterization of bioluminescent Vibrios

Aya Brown Kav
Microbial Diversity 2012

Abstract

Bacteria and Archaea can have a large variation in the genetic content, even among closely related strains. The genetic content which is not shared by all members of the taxonomical group is generally termed pan-genome. In this study, a phenotypic and metabolic approach was employed in order to explore the pan and core genome of bioluminescent members of the Vibrio genus. Overall, our findings suggest that some phenotypes such as colony morphology and over all carbon utilization profiles portray a similar differentiation as described by the molecular phylogeny. These can generally be considered as phenotypes which are coded by core-genome genes. Other phenotypes such as specific carbon utilization profiles (such as D-glucoronic acid) as well as bioluminescence timing and intensity cannot be explained by the molecular phylogeny and may be considered as phenotypes generated by the pan-genome. Finally, phenotypes which correspond to the molecular phylogeny are more congruent with the 16S tree than the hsp60 tree, suggesting that perhaps 16S rRNA makes a better molecular phylogenic marker than Hsp60 for this system.

Introduction

Phylogenetic classification and the ultimate structure of the tree of life has been a focus of research for decades. Microbes have posed challenges to their own classification due to two major reasons. First, they are very small and unlike multicellular organisms, tracking morphology and behavior can be tricky, therefore science has turned to molecular phylogeny as the main measure. Secondly, microbes tend to transfer genes horizontally which add complexity and entanglement into the process of molecular phylogeny (Doolittle, 1999). More recently, it became apparent the 'species concept', at large, may be inapplicable to taxonomy of microbes. Advances in the sequencing capacities of bacterial genomes have a thought us genetic content of even closely related bacterial genomes can vary to an extent which was not previously anticipated. These observations gave rise to the concept of pan-genome. The pan-genome includes the gene content which is not shared among most members of the taxa (Snipen *et al.*, 2009; Welch *et al.*, 2002).

Member of the genus Vibrio have been studied extensively in the context of pan genomics, having their pan-genome estimated at 26,500 genes (Thompson *et al.*, 2009; Vesth *et al.*, 2010). Most of our current knowledge on the Vibrio pan-genome comes from comparative genomics and sequence analysis (Chun *et al.*, 2009; Preheim *et al.*). Here, a phenotypic and metabolic approach was employed in order to explore the pan and core genome of bioluminescent members of the Vibrio genus. The premise for this study was that varying phenotypes which differentiate the bacteria according to their molecular clustering arrangement (either 16S gene or hsp60) can be considered part of the 'core-genome'. Functions which do not differentiate according to the molecular clustering are most likely not a part of the 'core-genome' and may be considered part of the 'pan-genome'.

Material and Methods

Strain collection and media

Sea water was collected from Eel pond, MBL beach and Garbage beach at 2 PM, constantly, and was plated on sea media plates, 100µl per plate. Plates were incubated overnight at room temperature. In the next day, plates were inspected in a dark room. Colonies which displayed bioluminescence were transferred to a fresh plate for further isolation. Isolates were transferred several times until colonies seemed isolated.

Sea water media (per 1 Liters):	
Yeast extract	1 g
Bacto-tryptone	5 g
Glycerol	3 ml
Artificial sea water	750 ml
DDW	250 ml
Adjust pH to 7 with 0.1M NaOH. For Agar plates, add 30 g of agar. Autoclave media for 40 minutes.	

16S rRNA gene and hsp60 PCR

Each of the isolates (60 total) was subjected to colony PCR using two primer pairs. Amplification of the 16S rRNA gene sequence was carried out using universal bacterial primers 8F (5'-AGAGTTTGATCCTTGGCTCAG-3') and 1492R (5'-GCYTACCTTGTTACGACTT-3'). Amplification of the hsp60 gene was carried out using primers H279 (5'-GAATTCGAIIIIGCIGGIGA(TC)GGIACIACIAC-3') and H280 (5'-CGCGGGATCC(TC)(TG)I(TC)(TG)ITCICC(AG)AAICCIGGIGC(TC)TT-3') as previously described (Goh *et al.*, 1996). PCR reactions were carried out using Promega master mix. Products were analyzed

on 1% agarose gel electrophoresis. Reactions which had a band corresponding to the expected amplicon length were chosen for further were purified using Exosap and subjected to Sanger sequencing. Generally, only isolates for which both genes were amplified were subjected to sequencing.
Sequences were trimmed using PHRED and each of the two sequence sets was aligned using MUSCLE alignment software, with the *E. coli* corresponding gene as an out group. Finally, a tree was built for each of the gene sets using FastTree software and the trees were inspected for isolate selection.

Growth Curves and Bioluminescence measurements

For each of the selected clones, growth curves were generated. Prior to the actual growth curve experiments, the isolates were grown overnight in a liquid culture. Overnight cultures were diluted by a thousand fold into fresh sea water media, and 100µl were aliquoted, in triplicates, in standardized, sterile, clear 96-well plates. Cultures were grown in 25°C over a time period of 24 hours shaken continuously and optical density at 630nm was measured at 15 minute increments using Synergy Mx plate reader. Measuring growth rates at different media pH was carried out using the same protocol. The sea water media was supplemented with 10mM MOPS buffer and the pH was adjusted to 6.6, 7, 7.4 and 7.8 using 0.1mM NaOH. Measuring the bioluminescence was carried out similarly while using black 96-well plates measuring luminescence at 15 minute increments.

Carbon Utilization Profiles

In order to measure the carbon utilization profile of the isolates, the Biolog (NG II) plates were used, testing the utilization of 95 different carbon sources. Each individual isolate was picked from a plate into minimal media (content bellow) until O.D 600 nm was at 0.1 (values between 0.08-0.12. 150µl of the diluted culture was added to every well. For each individual isolate, three triplicate plates was set up and incubated at 25°C. Color precipitate was already visible for some carbon sources, plates were read at 590 nm after 40 hours of incubations.

Biolog minimal media:

NaCl	25 g
MgCl2	8 g
KCl	0.5 g
MOPS	10mM

Add 950 ml of DDW, adjust pH to 7.5 and complete with DDW to a final volume of 1 L.

The results, optical density at 590 nm, were than analyzed by Qiime pipeline. The average OD for each of the carbon sources, for each of the isolates was transformed into an "OTU Table" as in input for Qiime. The rest of the analysis was carried out using Qiime scripts. A distance matrix, using the Bray Curtis similarity index, was created by the 'beta_diversity.py' script. Following the similarity matrix generation – an UPGMA tree was constructed in order to illustrate the distance between the different carbon utilization profiles of the isolates (using 'upgma_cluster.py').

Results and discussion

Selection of Isolates

The sequences obtain from the two sequences genes, hsp60 and the 16 rRNA gene, were aligned and a tree was constructed (figure 1, A and B). As hsp60 has been described as a better molecular clock for Vibrio bacteria, the hsp60 tree (figure 1A) was used to select the isolates for analysis. As can be seen in the hsp60 tree, most of the isolates form a "hair brush" like clade in the tree indicating that they are very closely related. Three other clusters were visible in the hsp60 tree. In order to test the hypothesis, four isolates from the same clade were chosen (GB47, GB52, GB19 and GB59 – highlighted in blue) as the similar isolates, while one isolate was chosen from each of the three additional clades (GB12, GB48 and GB18 – highlighted in pink. When inspecting the 16S rRNA based tree is was clear that most of the isolates take up roughly the same organization on the tree, with one exception, GB74, which does not appear in the closely related clade and is closer to the other isolates.

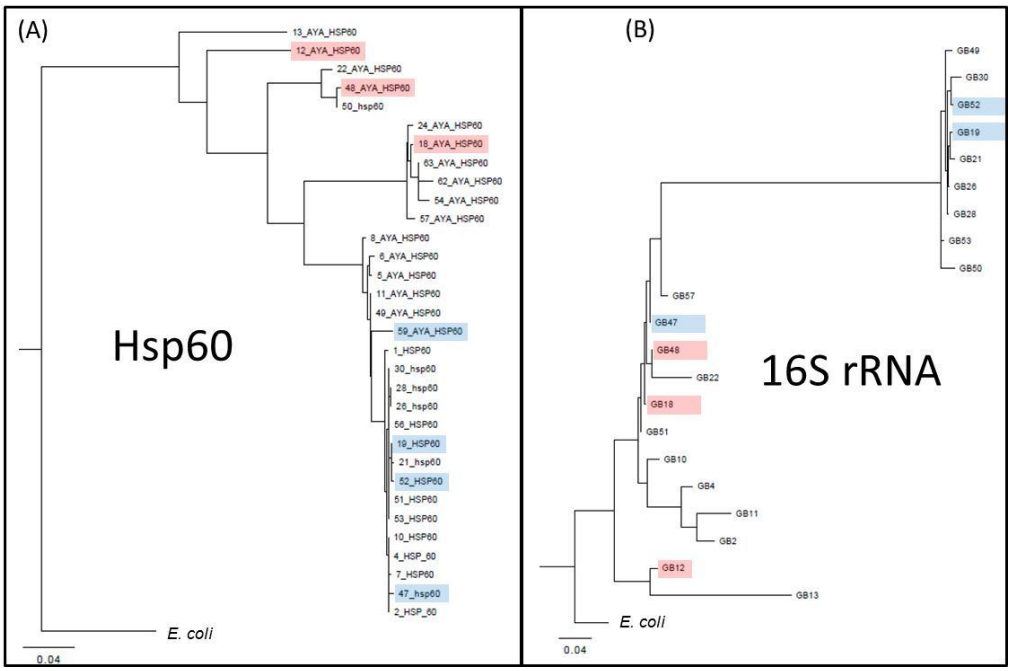


Figure 1. Phylogenetic trees created for the aligning Sanger sequences of both genes using MUSCLE. Tree construction was done by FastTree. (A) The tree generated using the alignment of the hsp60 gene. (B) The tree generated using the alignment of the 16S rRNA gene.

Growth curves

Growth curves experiments were performed in triplicates for each of the seven isolates under several conditions. Figure 2 illustrates the different growth curves generated by each of the isolates. As can be seen, the growth curves do not portray a phenotypic difference between the different isolates, and therefore, cannot be used for to distinguish between isolates. It can be assumed that genes which control growth rate under these tested conditions are a part of the core genome as this phenotype is share among all isolates. It should be noted that measuring the growth curves under different pHs (6.6, 7, 7.4 and 7.8) did not yield different growth curves as well. Hence, it can also be assumed that genes related to pH adaptation are also part of the core-genome.

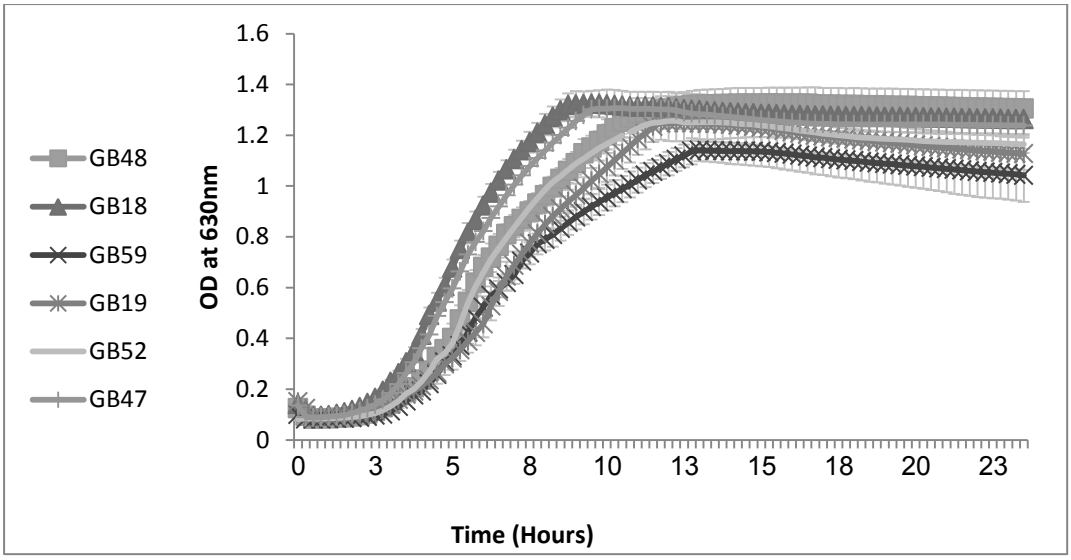


Figure 2. Growth curves generated for the isolates over a time course of 24 hours in 25°C.

Bioluminescence

The bioluminescence activity during growth of each of the isolates was measure during a 24 hours period of time, using a luminescence plate reader (Synergy Mx). As can be seen in figure 3, the bioluminescence activity of each of the isolates varied significantly, both in time and in intensity. In order to better clarify the relationship between the growth curve stage and the bioluminescence activity, the measurements were also plotted for each of the strains individually (Figure 4). As can be seen, two of the isolates, GB18 and GB47, although selected for bioluminescence did not show bioluminescence activity under the tested conditions. Three other phenotypes were visible: (1) those whose bioluminescence activity peaks early in the exponential grown phase – GB59, GB52 and GB48). (2) Those whose bioluminescence activity peaks later in the exponential grown phase – GB19. (3) Those whose bioluminescence intensity is roughly 2 fold higher than the rest of the isolates.

A correlation between the bioluminescence phenotype and the position of the isolates portraying this phenotype in either of the trees (16S rRNA and hsp60) could not be found. The different phenotypes are displayed randomly by different isolates in different clades of the trees. This phenotypic observation, thus cannot be related to the core genome, and may be considered part of the pan-genome.

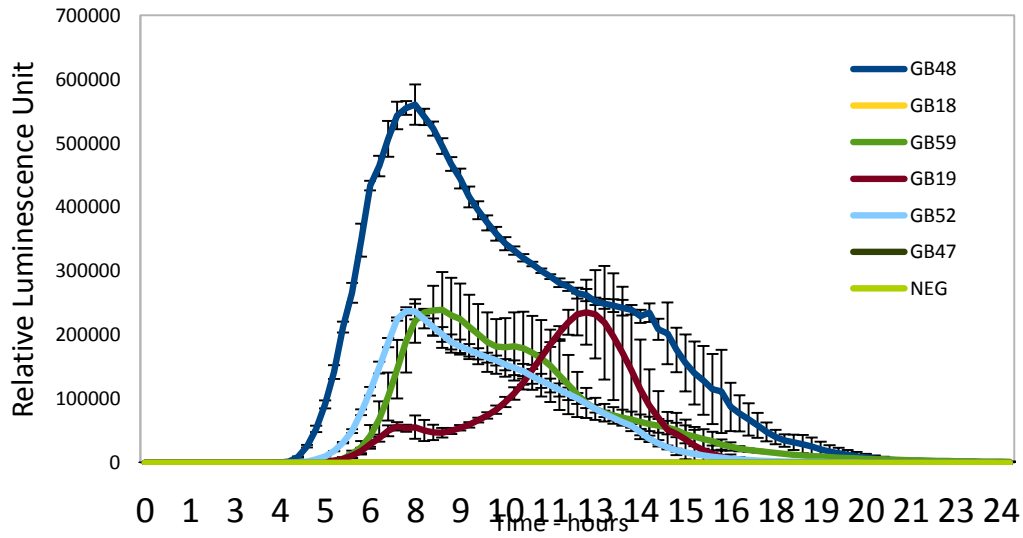


Figure 3. Relative Luminescence Units (RLU) as were measuring during a growth curve of each of the isolates.

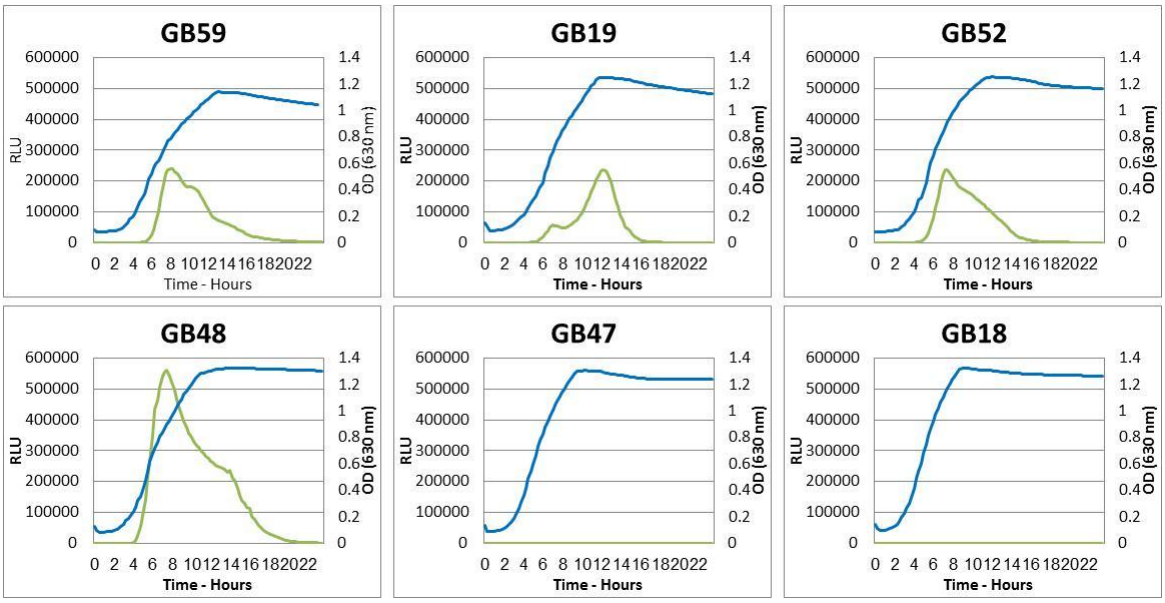


Figure 4. The bioluminescence (left Y axis) and the OD (right Y axis) is plotted for each of the isolates over a time period of 24 hours.

Carbon Utilization profiles

The carbon utilization profile of each of the isolates was established using the Biolog GN II plates. Figure 5 illustrates the different carbon utilization profiles of each of the isolates. Two separate groups can be seen (also visible in figure 6). Generally, the three left isolates show a higher degree of similarity to each other than the three right ones, who also show higher similarity among them. However, utilization of some carbon sources are quite different and cannot be explained but either tree (16s rRNA or hsp60). Such carbon sources include β-Hydroxy Butyric Acid, Mannose, Tween 40 and Arabitol. These carbon sources and more can be used to investigate the pan-genome of these Vibrio isolates. another interesting observation is that D-glucoronic acid is correlated with the bioluminescence activity as high bioluminescence is visible in the same isolates that show high utilization rate of D-glucoronic acid. Further investigation is required in order to elucidate this correlation. Over all, the general carbon utilization profile is able to distinguish between the two groups.

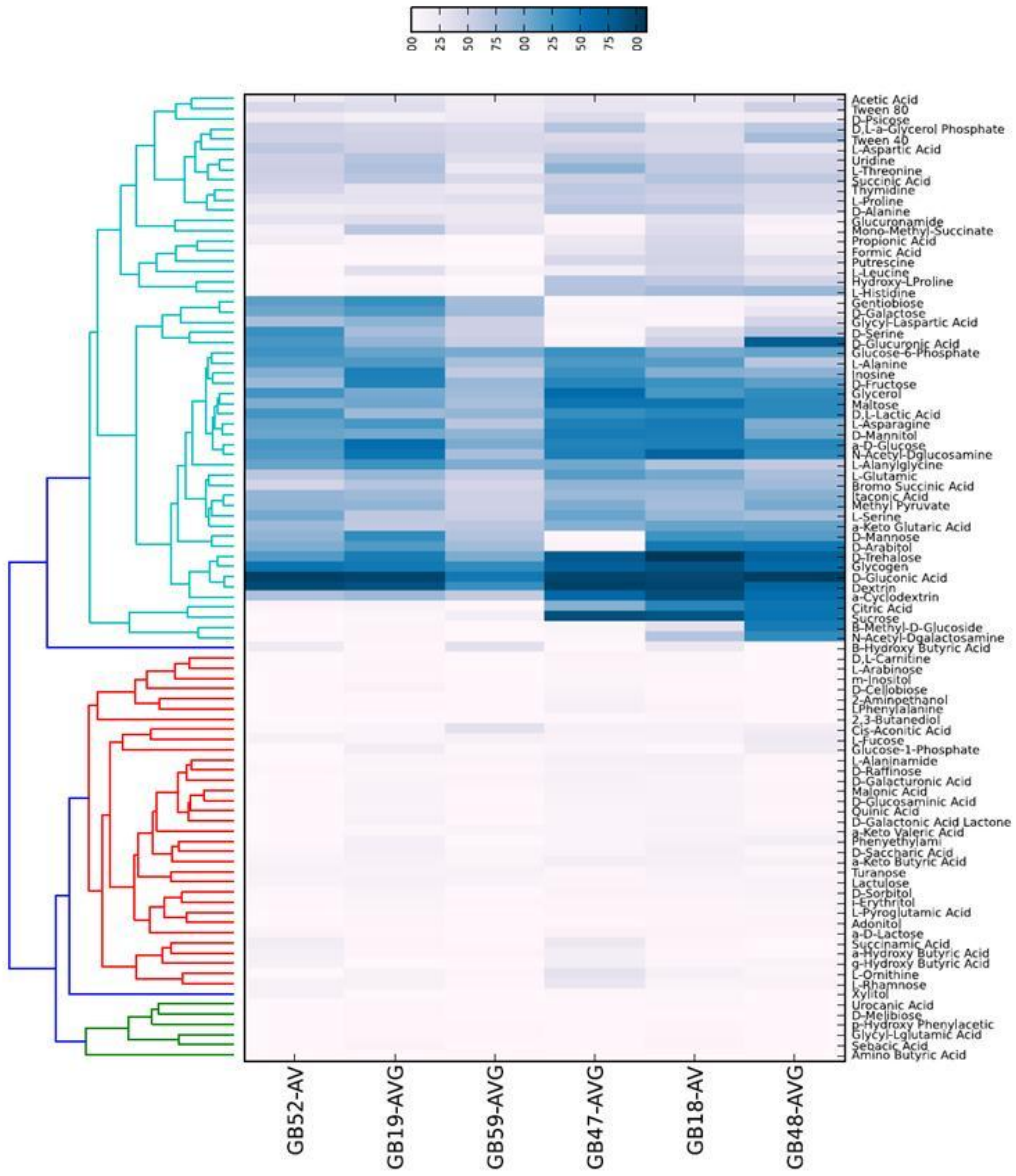
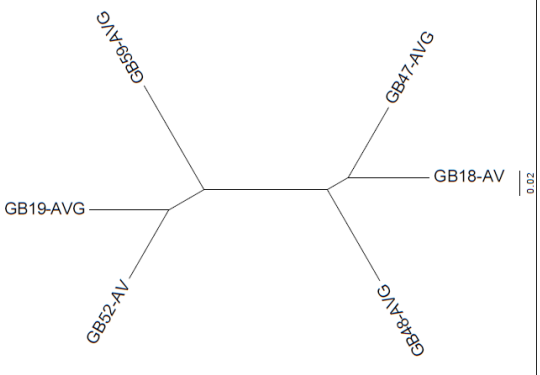


Figure 5. A heatmap was generated to illustrate the different carbon utilization of the isolates. The color intensity is correlated to isolates' ability to utilize each carbon source. The dendograms on the left were constructed using Bray Curtis similarity index, and illustrate the relationship between the different carbon sources and their utilization by the different isolates.

Figure 6. UPGMA tree describing the relationship between the isolates according to their ability to utilize 95 different carbon sources using the Bray Curtis similarity index.



As can be seen in figure 6, two separate clades generated by the carbon utilization analysis. One clade includes isolates GB19, GB52 and GB59. All three isolates were picked from the closely related clade in the hsp60 tree. The other clade in figure 6 includes GB47 which was also visible in the in same clade as GB19, GB52 and GB59 in the hsp60 tree. This data, together with colony morphology data collected, which illustrating a similar colony morphology among GB19, GB52 and GB59 and different from GB47, GB18 and GB48 suggest that phenotypes that do differentiate the isolates according to their molecular clustering arrangement, and can be considered part of the 'core-genome', are much better explained by the 16S tree than the hsp60 tree. One possible explanation for this phenomenon is that the hsp60 gene of isolate GB47 was recently horizontally transferred from one of the closely related clade members and explains its differential location in the two trees. Another possibility is that for this system, 16S phylogeny is a better molecular clock than hsp60, which can provide a better explanation to the phylogenetic relationship among these isolates.

Isolate GB12

This isolate, was selected for it high intensity of bioluminescence. The 16s rRNA sequence suggests that this is a member of the Photobacterium species but the hsp60 sequences is annotated as Vibrio (it is possible that this is a mistake – for further investigation this isolate must be sequenced again). Generally, it had a different phenotypic profile than all other isolates and therefore was excluded from the further analysis. However, a few interesting observations were made. It has a much slower growth rate, reaching stationary at a much lower density (Figure 7a). Bioluminescence activity was at least 10 fold more intense than any of the other isolates (Figure 7a) and pH sensitivity was also observed (figure 7b). Different colony morphology was also visible.

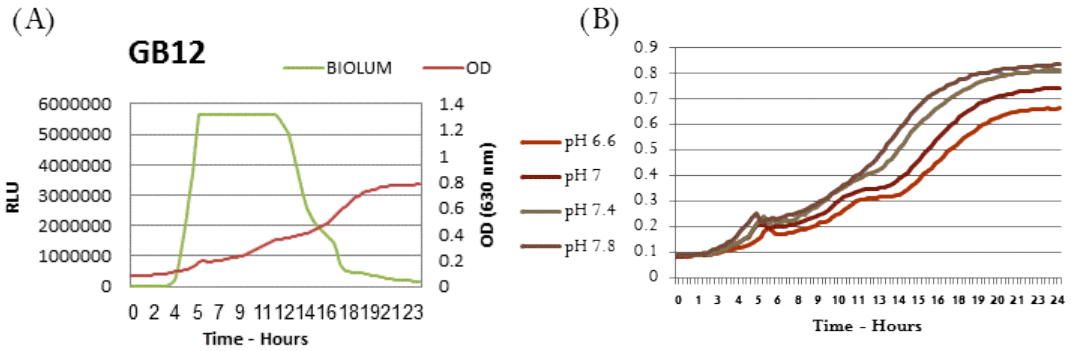


Figure 7. (A) – Growth curve (red) and bioluminescence curve (Green). For the bioluminescence measurement, the peak is not portrayed as it exceeded the detected capacity of the pate reader; a flat line is illustrated instead. (B) The growth curve under four different pHs.

Acknowledgments

I would like to thank Prof. Dan Buckley and Prof. Steve Zinder for directing a great course. I would like to also thank all of the TAs for helping me throughout my summer endeavors and my classmates for being there, always. I couldn't have done it without any of you! This summer has been one of the most fantastic, rewarding and magical experiences of my life, and I am positive that all the people mentioned above contributed to that. Having been surrounded, all summer, by people who love science, especially microbiology, as much as I do, was a dream come true.

I finally would like to sincerely thank Pfizer Inc. and the Department of Energy who made my participation in Microbial Diversity, class of 2012, possible.

References

Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ *et al* (2009). Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic Vibrio cholerae. *Proc Natl Acad Sci U S A* **106:** 15442-7.

Doolittle WF (1999). Phylogenetic classification and the universal tree. *Science* **284:** 2124-9.

Goh SH, Potter S, Wood JO, Hemmingsen SM, Reynolds RP, Chow AW (1996). HSP60 gene sequences as universal targets for microbial species identification: studies with coagulase-negative staphylococci. *J Clin Microbiol* **34:** 818-23.

Preheim SP, Timberlake S, Polz MF Merging taxonomy with ecological population prediction in a case study of Vibrionaceae. *Appl Environ Microbiol* **77:** 7195-206.

Snipen L, Almoy T, Ussery DW (2009). Microbial comparative pan-genomics using binomial mixture models. *BMC Genomics* **10:** 385.

Thompson CC, Vicente AC, Souza RC, Vasconcelos AT, Vesth T, Alves N, Jr. *et al* (2009). Genomic taxonomy of Vibrios. *BMC Evol Biol* **9:** 258.

Vesth T, Wassenaar TM, Hallin PF, Snipen L, Lagesen K, Ussery DW (2010). On the origins of a Vibrio species. *Microb Ecol* **59:** 1-13.

Welch RA, Burland V, Plunkett G, 3rd, Redford P, Roesch P, Rasko D *et al* (2002). Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli. *Proc Natl Acad Sci U S A* **99:** 17020-4.