Phylogenetic analysis of the circadian clock regulating gene *kaiC* from various bacterial species

Bridget Hatfield (Janssen), EEOB563, 4/10/2018, https://github.com/BridgetHatfield/project

Introduction

Light and dark exposure are a part of everyday life for almost all organisms on the planet. Due to day/night cycles, creatures have evolved to maintain a host of sensors to respond accordingly to changes in their light environment. Phytochromes are a light sensor found in many species that have a variety of response outcomes. In foliar bacteria, phytochromes can cause changes in mobility, ability to colonize, virulence, lesion size, and potentially other phenotypes depending on the specific wavelength of light present (McGrane and Beattie, 2017; Wu et al., 2013). A recent study (Hatfield, Haili, and Beattie, unpublished data) shows that the bacteriophytochrome in *Pseudomonas syringae* pv. syringae B728a regulates a gene named *kaiC*. The *kaiC* gene has homologs in many organisms and is known to function as a circadian clock regulating gene (Rosbash, 2009).

The circadian clock is a mechanism used by organisms to develop a natural rhythm that typically coincides with normal day/night or light/dark cycles. The circadian clock can be set by various factors, namely light and temperature. One hypothesis is that organisms use this internal clock to anticipate environmental changes and react accordingly, or in advance to increase their levels of fitness (Reece et al., 2017). Foliar pathogenic bacterium may use its circadian rhythm as a way to evade host defenses. Additionally, foliar associated bacteria would have different evolutionary pressures than soil or aquatic bacteria as the environmental signals would differ greatly based on location. For example, foliar bacteria would have a greater exposure to light and temperature flux than soil borne organisms. This study aims to determine if environmental niche is a driving force of *kaiC* evolution in bacteria that are isolated from various environments.

Methods

Data collection and preparation

The dataset was assembled using the NCBI protein database and searching for bacterial species that contained KaiC proteins. One sequence (*Pseudomonas syringae* pv. tomato DC3000) was obtained from http://pseudomonas.com. A broad range of species were selected to represent bacteria that are plant soil/root associated, foliar associated, or aquatic, as well as one outgroup species (*Escheria coli*) which resides commonly in animal gastrointestinal tracts. Data was downloaded in the FASTA format, aligned with Mafft –auto, and converted to PHYLIP format using the site: <http://phylogeny.lirmm.fr/phylo_cgi/data_converter.cgi>.

Phylogenetic analysis using RAxML

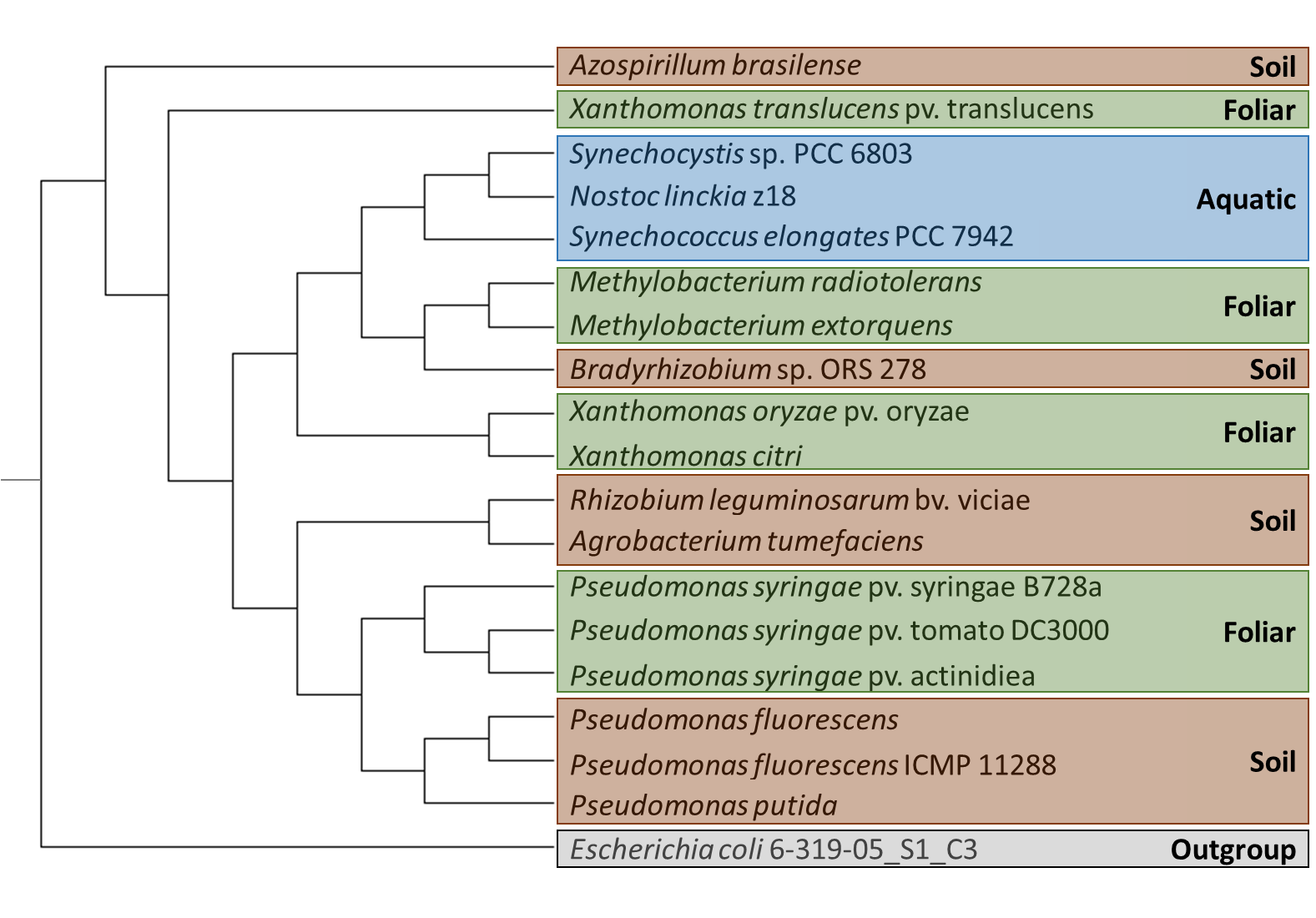
Maximum likelihood analysis was conducted using RAxML to determine the phylogenetic structure of *kaiC* of the selected species. This program was chosen based on its ability to appropriately determine phylogenetic structure of any size dataset. RAxML is a quick and efficient program that is easily executed by the user and has parameters that are simply changed for each option (Stamatakis, 2014). For this analysis WAG and AUTO models were applied with a random number seed (-p) of 864 for the WAG model. The outgroup (-o) function was also used to set the outgroup to *E. coli*. The resulting trees were compared for similarity, found to be identical and further analyzed with prior information on each species. The program calls used were as follows:

raxmlHPC-PTHREADS-SSE3 -T2 -m PROTGAMMAWAG -p 567 -s KaiCecoli.phy -#40 -n ecoli 1

raxmlHPC-PTHREADS-SSE3 -T2 -p 864 -m PROTGAMMAAUTO -o E.coli -s KaiCecoli.phy -n AUTO3

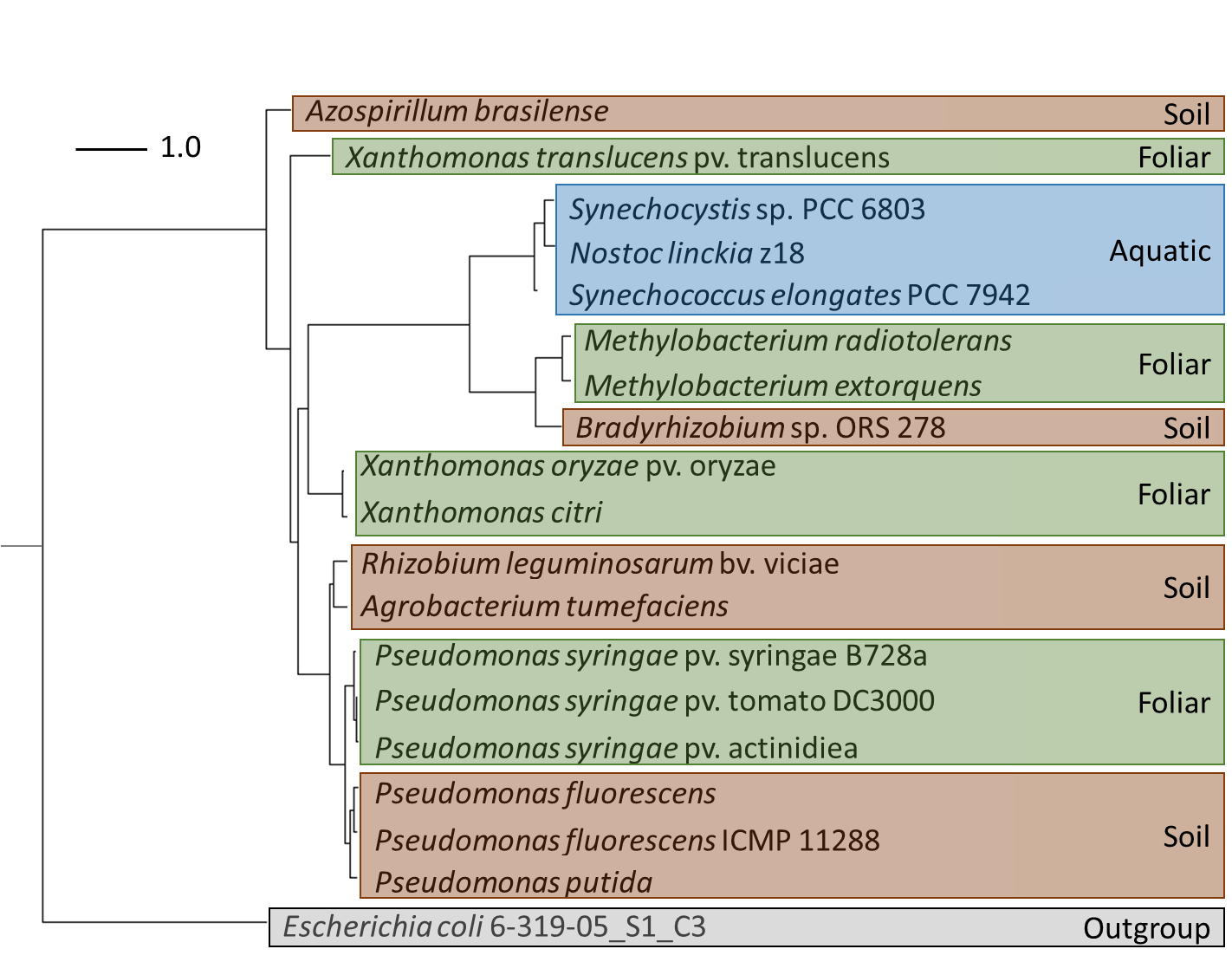
Trees were visualized using Dendroscope.

Results

Phylogenetic analysis using RAxML produced a tree that showed grouping of bacterial species based off of relatedness of the *kaiC* gene. Once the analysis was ran clade groups were evaluated based on prior knowledge of each specie’s ecological niche.

**Figure 1.** **Cladogram of bacterial species** shows grouping patterns based on relatedness of circadian clock regulating *kaiC* gene. Bold print at the right of boxes denotes typical ecological niche of the taxa.

In Fig. 1 the *Pseudomonads* group closely together, however, there is a separation between the plant associated bacteria that are found in foliar environments from those that are located in the soil. Tight grouping also occurs among the cyanobacterial species. Both *Synechococcus* strains are cyanobacteria typically found in aquatic environments and their *kaiC* genes are most closely related. The *Nostoc* is also a cyanobacteria found in aquatic environments. Among the other plant associated strains there does not appear to be any pattern in *kaiC* related clustering based on ecological niche, leading to the conclusion that *kaiC* evolution is likely not dependent on environment.

Figure 2 shows a phylogram that takes distance into account when graphing the relatedness of *kaiC* visually. The phylogenetic tree shows that the *kaiC* gene from cyanobacteria, M*ethylocacteria*, and *Briadyrhizobium* sp. ORS 278 has experienced more genetic changes than the *kaiC* gene from the other species evaluated.

**Figure 2.** **Phylogenetic tree** shows phylogenetic relatedness of circadian clock regulating *kaiC* gene in various bacterial species. Bold print at the right of boxes denotes typical ecological niche of the taxa.

*Escherichia coli* was chosen as an outgroup for this analysis because *E. coli* is animal associated rather than plant associated. *E. Coli* is commonly found in diverse environments and gastrointestinal tracts in humans and other animals and has no strong plant association making it an ideal outgroup for this analysis.

Discussion

Circadian rhythm is an important tool for both plants and bacteria when it comes to defenses and evasion respectively. This study was driven by the need to understand how environmental niche may be related to circadian clock regulation via the *kaiC* gene. Plants have many functions that are controlled by their circadian rhythms. One very important function that is controlled by the clock in plants is the defense system (Lu et al., 2017). Stomates, openings in the bottom of the leaf which allow for airflow, are also a point of entry for foliar pathogens. The closing and opening of stomates is controlled by plant clock genes that also raise defenses around the stomates when they first open. Due to the higher level of plant defense at certain times of day, bacteria could avoid this attack by using their own clock controlled genes to wait until a more favorable time to invade the host. Foliar bacteria are exposed to high levels of light so their circadian evasion of defenses may be different than that of root associated bacteria.

When looking at Fig. 1, the *kaiC* genes from the *Pseudomonads* group together in two distinct clades, one of foliar associated bacteria and one of soil associated bacteria. This may be due to differences in their habitats. However, the conclusion that the *kaiC* gene has evolved due to soil vs. foliar habitat is not well supported because the other soil and foliar designated taxa do not group together. In fact, the grouping appears to be based more on species than environment.

Another clade to make note of is the clade that groups cyanobacteria together. Not only are these organisms all cyanobacteria, they are also photosynthetic, and all found in aquatic environments. Either the fact that these organisms are all cyanobacteria, the environmental location, or the fact that they are photosynthetic and light is important to their survival have led to *kaiC* being highly conserved in this clade compared to the bacteria that are naturally found in the soil or foliar environments.

Fig. 2 looks more specifically at changes in the *kaiC* gene over time. In this figure it is clear that *kaiC* from cyanobacteria, *Methylobacteria*, and *Bradyrhyzobium* have undergone a greater number of genetic changes than the other species in the analysis. This could be an additional reason that these taxa group apart from the other taxa with similar environmental classification.

Going forward a repeat of this analysis should take several aspects into consideration to build a more robust conclusion based on phylogeny. Gene synteny could be taken into account as some organisms contain an operon of *kaiABC* and other organisms only contain *kaiC*. It would be interesting to see if *kaiC* is more highly conserved in the operon or in its orphan gene form. An additional necessity would be to use a broader range of species that are not so closely related to one another. For example, many of the *Pseudomonad* species evaluated are known to be very close relatives. If this group of organisms could be increased in number of taxa evaluated the results may become more clear. Whole genomes could also be used to look at the relatedness of species as a rather than a single gene. Likewise, specific sequences could be added to the dataset that are related to the circadian clock to determine if environment is the primary cause of adaptation to the clock or if other factors are the driving force. The analysis itself could be strengthened by doing a bootstrap analysis to provide greater support values to the results. All in all, at this point it would be difficult to definitely state that ecological niche is an important causal agent to *kaiC* evolution.

References

Lu, H., McClung, C. R., and Zhang, C. (2017). Tick Tock: Circadian Regulation of Plant Innate Immunity. *In* "Annual Review of Phytopathology, Vol 55" (J. E. Leach and S. E. Lindow, eds.), Vol. 55, pp. 287-311. Annual Reviews, Palo Alto.

McGrane, R., and Beattie, G. A. (2017). Pseudomonas syringae pv. syringae B728a Regulates Multiple Stages of Plant Colonization via the Bacteriophytochrome BphP1. *Mbio* **8**.

Reece, S. E., Prior, K. F., and Mideo, N. (2017). The Life and Times of Parasites: Rhythms in Strategies for Within-host Survival and Between-host Transmission. *Journal of Biological Rhythms* **32**, 516-533.

Rosbash, M. (2009). The Implications of Multiple Circadian Clock Origins. *Plos Biology* **7**, 421-425.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313.

Wu, L., McGrane, R. S., and Beattie, G. A. (2013). Light Regulation of Swarming Motility in Pseudomonas syringae Integrates Signaling Pathways Mediated by a Bacteriophytochrome and a LOV Protein. *Mbio* **4**.