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

Bridget Hatfield (Janssen), Iowa State University, EEOB 563

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 A recent study (Hatfield and Beattie, unpublished data) shows that the bacteriophytochrome in *Pseudomonas syringae* pv.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. 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). For example, a foliar pathogenic bacterium may use its circadian rhythm as a way to evade host defenses. 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). Stomata, openings in the bottom of the leaf which allows for airflow, are also a point of entry for foliar pathogens. The closing and opening of the stomata are controlled by clock genes that also control additional plant defenses to raise defenses around the stomates when they first open. Due to the higher level of plant defensive gene activation 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.

Due to virulence and colonization being controlled by a bacteriophytochrome which directly regulates the *kaiC* gene in *Pseudonomas syringae* pv. syringae B728a I was curious about the phylogeny of *kaiC* in various bacterial species that have similar and different lifestyles from B728a. This study describes a phylogenetic analysis of *kaiC* genes from bacterial species that are naturally found in the soil, foliage, and aquatic environments. This study also took into consideration the pathogenicity or beneficial nature of the organisms and whether they are photosynthetic or not.

Methods

Data Collection

The dataset was assembled using the NCBI protein database and searching for bacterial species that contained KaiC proteins. A broad range of species were selected to represent bacteria that are both pathogenic and non-pathogenic, located in the soil and foliar surfaces, as well as two outgroup species. Data was downloaded in the FASTA format and converted to PHYLIP format using the following 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 the KaiC gene 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 several models were applied and all models yielded the same resulting tree. The applied models included WAG, JTTF, auto, autoaic, and autobic with a random number seed (-p) of 34271. The resulting trees were compared for similarity and further analyzed with prior information on each species.

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 knowledge of natural location, pathogenicity, and other noticeable traits.

At first glance the most noticeable clade group is that of the *Pseudomonads* these closely related species group together unsurprisingly. Additionally, the foliar pathogen strains are more closely related than the *pseudomonad* that is not a pathogen and is located primarily in the soil.

There is a second group of foliar pathogens whose closest related ancestral *kaiC* gene is most closely related to the *kaiC* gene from the nitrogen fixing soil bacterium *Bradyrhizobium japonicum.*

Soil associated bacterium make up a small clade that consists of *Rhizobium leguminosarum* and *Rhizobium radiobacter*, but that is where the phenotypic relatedness ends for this clade as one of the organisms is a beneficial nitrogen fixing bacterium and the other is a pathogen.

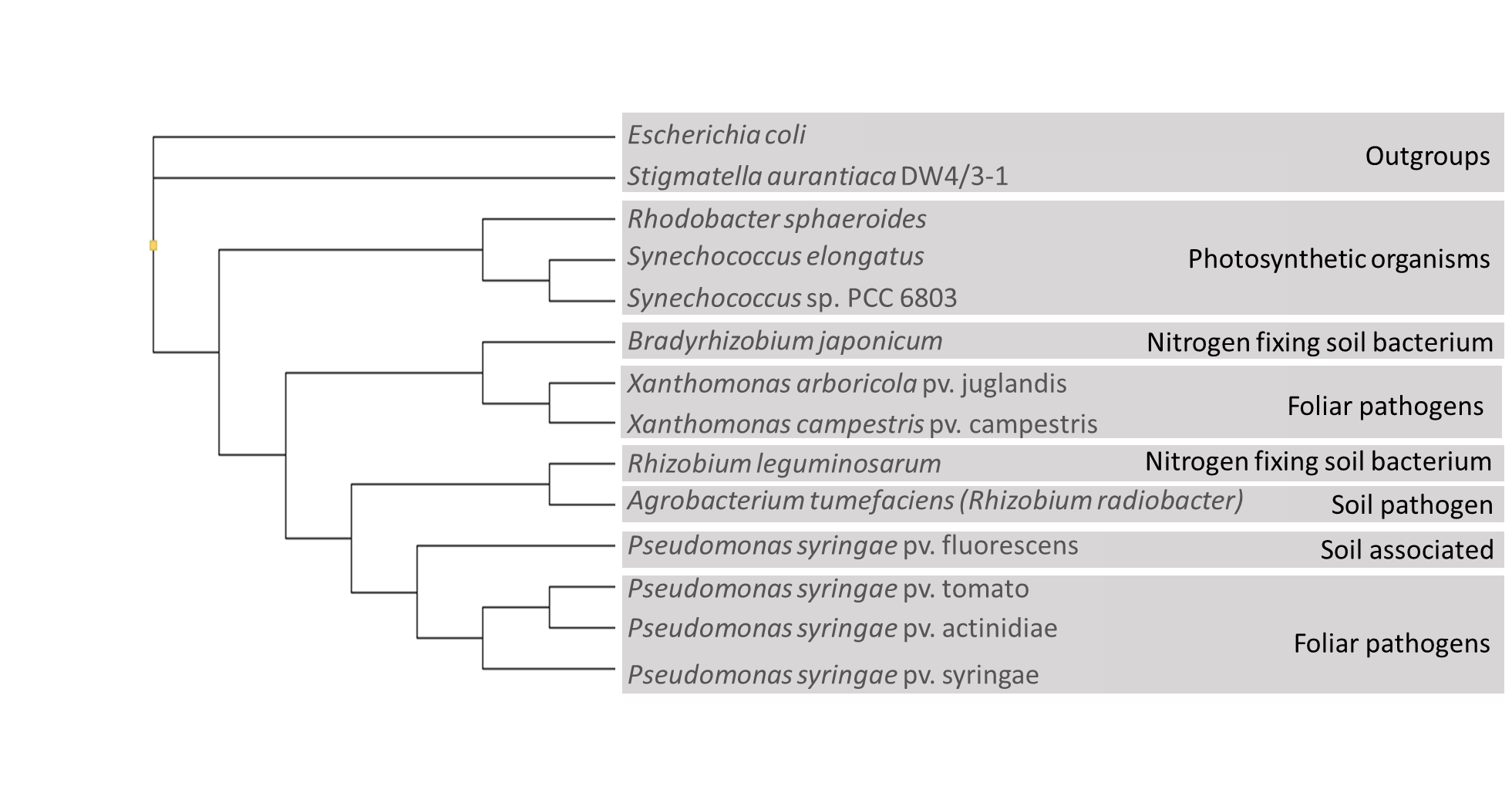


Fig. 1- Phylogenetic analysis of bacterial species shows grouping patterns based on relatedness of circadian clock regulating *kaiC* gene. Bold print at right of tree denotes typical location, pathogenicity, and photosynthetic ability.

Fig. 1

The final large clade is made up of organisms that are photosynthetic. Both *Synechococcus* strains are cyanobacteria typically found in aquatic environments and their *kaiC* genes are most closely related. The *Rhodobacter* strain is photosynthetic when necessary, can fix nitrogen, and is also found in aquatic environments.

Two strains came out as outgroups, *Escherichia coli* commonly found in diverse environments and gastro-intestinal tracts, and *Stigmatella aurantiaca* which is a myxobacteria primarily found in the soil with a very unique lifestyle unlike that of most other bacteria.

Discussion

Circadian rhythm is an important tool for both plants and bacteria when it comes to defenses and evasion respectively. Due to this and other phenotypes that we know are controlled by light regulation I was interested in looking at how a group of diverse bacterial strains, all with a *kaiC* gene, would group phylogenetically based on the alignment of this gene.

When looking at Fig. 1, the foliar pathogens group together, but in two distinct clades. This may be due to the similarity of the diseases that are caused in each group, or due to the fact that one group is made up of *Pseudomonads* and the other made up of *Xanthomonads.* What is interesting is that while the foliar pathogens group together, the soil borne bacteria have a much looser grouping. This could be interpreted to mean that the soil bacteria have a lesser exposer to light than the foliar bacteria and therefore the *kaiC* gene is not as conserved among the soil bacteria.

Another clade to make note of is the clade that groups photosynthetic organisms together. Not only are these organisms photosynthetic, but they are also all found in aquatic environments. Either the environmental location of these bacterium, 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.

In order to gain a deeper understanding of the functional categories of these phylogenetic groups future studies could take additional genes controlled by KaiC and bacteriophytochromes and investigate if the clade organizational structure changes. It would be useful to look at genes that are specific to virulence and host defense evasion as well as genes that are known to be related to beneficial interactions with hosts. Whole genomes could also be analyzed to look at total relatedness of these species.

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**.

Data

>Pseudomonas syringae pv. syringae B728a

MNTSKAATGIEGLDDILAGGLSRSHVFLLEGEPGTGKTTVALQFLQAGAAAGEISLYITLSETESELRSG

AASHGWELDEHINIFELTPPESLLDADHHQSLLYSSDLELGEATRQIFEVVERVKPTRVVIDSLSEIRLL

AQSSLRYRRQILAIKHYFTRYNATVLLLDDLTTEALDKTVHSVAHGVIRLEALTPTYGAERRRLKIVKYR

GQKYRGGFHDFTIAENGIHVFPRLVAAEHRSNYSRSQLSSGIPELDNLLGGGIEGGSSTLILGPAGTGKS

LISLVFAVQAVARGERVGLFIFDEEMGLLFERMLKLGIDLRALQETGNLVIEQIDAAELSPGEFAHRVRR

AVDKKQIKTVVIDSINGYQAAMPEESALVLHMHELLLYLNRQGASTFMTVAQHGLVGDMRSPVDITYLAD

SVILLRYFEALGQVRRAISIIKKRTGTHESTIREYRISSNGLKIGEPLQAFQGVLRGVPSYMGDKKPLLE

DDDL

>Pseudomonas syringae pv. actinidiae

MNTSKAETGIEGLDDVLSGGLSRSHVFLLEGEPGTGKTTVALHFLQAGAAAGEISLYITLSETERELRAG

AASHGWELDPQIHIYELTPPESLLDAEHHQSLLYSSDLELGEATRQIFEVVERIKPSRVVIDSLSEIRLL

AQSSLRYRRQILAIKHYFTRYNATVLLLDDLTTEALDKTVHSVAHGVIRLEALTPTYGAERRRLKIVKYR

GQKYRGGFHDFTIAEDGIHVFPRLVAAEHRSNYSRIQLSSGIKELDNLLGGGIESGSSTLILGPAGTGKS

LISLVFAVQAVARGERVGLFIFDEEMGLLFERMLKLGIDLRALQETGNLVVEQIDAAELSPGEFAHRVRS

AVDRKQIKTVIIDSINGYQAAMPEESALVLHMHELLLYLNRQGAATFMTVAQHGLVGDMRSPVDITYLAD

SVILLRYFEALGQVRRAISIIKKRTGTHESTIREYRINSSGLTIGEPLEAFQGVLRGVPSYFGDKKPLLA

DDDQ

>Pseudomonas fluorescens

MSTSKQLLSEKAATGVEGLDDILAGGLSRSHVFLLEGEPGTGKTTVALHFLRAGAQNGERCLYITLSETE

HELRQGAKSHGWDLDEHIHIFELTPPESLLNADHQQSLLYSSDLELGEATRQIFEVVERVKPTRVVVDSL

SEIRLLAQSSLRYRRQILAIKHYFVRYDATVLLLDDLTTESLDKTVHSVAHGVIRLEELTPTYGAERRRI

RVVKYRGQKYRGGFHDFTIMGDGIHVFPRLVAAEHRGGYNRQTLSSGIEELDSLLGGGIETGSSSLILGP

AGTGKSLISMIFAAAAVARGEKAALFIFDEELGLLFERMKNMGIDLAALRDTGNLLIEQVDAAELSPGEF

SHRVRRCVDERGIKTVVIDSINGYQAAMPEENALILHMHELLLYLNRRGAATFMTVAQHGLVGDMQTPVD

ITYLADTVILLRYFEALGKVRRAISIIKKRTGSHESTIREYRIGSRGMTVGVPLDNFQGVLRGIPTYMGA

GSPLLKDEG  
>Pseudomonas syringae pv. tomato  
MNTSKAETGIEGLDDILSGGLSRSHVFLLEGEPGTGKTTVALQFLQAGAAAGEISLYITLSETERELRAG

AASHGWELDPQIHIYELTPPESLLDAEHHQSLLYSSDLELGEATRQIFEVVERIKPSRVVIDSLSEIRLL

AQSSLRYRRQILAIKHYFTRYNATVLLLDDLTTEALDKTVHSVAHGVIRLEALTPTYGAERRRLKIVKYR

GQKYRGGFHDFTIAEDGIHVFPRLVAAEHRSNYSRAQLSSGIRELDNLLGGGIESGSSTLILGPAGTGKSL

ISLVFAVQAVARGERVGLFIFDEEMGLLFERMLKLGIDLRALQETGNLVVEQIDAAELSPGEFAHRVRSAV

DRKQIKTVIIDSINGYQAAMPEESALVLHMHELLLYLNRQGAATFMTVAQHGLVGDMRSPVDITYLADSVI

LLRYFEALGQVRRAISIIKKRTGTHESTIREYRINSSGLKIGEPLEAFQGVLRGVPSYLGDMKPLLADDDQ

>Synechococcus elongatus

MTNLPEHQSSPTEQSSAEVKKIPTMIEGFDDISHGGLPQGRTTLVSGTSGTGKTLFAVQFLYNGITIFNE

PGIFVTFEESPQDIIKNALSFGWNLQSLIDQGKLFILDASPDPDGQEVAGDFDLSALIERIQYAIRKYKA

TRVSIDSVTAVFQQYDAASVVRREIFRLAFRLKQLGVTTIMTTERVDEYGPVARFGVEEFVSDNVVILRN

VLEGERRRRTVEILKLRGTTHMKGEYPFTINNGINIFPLGAMRLTQRSSNVRVSSGVKTLDEMCGGGFFK

DSIILATGATGTGKTLLVSKFLETGCQQGERALLFAYEESRAQLSRNASSWGIDFEELERRGLLRIICAY

PESAGLEDHLQIIKSEIADFKPSRVAIDSLSALARGVSNNAFRQFVIGVTGFAKQEEITGFFTNTTDQFM

GSNSITESHISTITDTILLLQYVEIRGEMSRAINVFKMRGSWHDKGIREYVITEKGAEIRDSFRNFEGII

SGTPTRISVDEKTELARIAKGMQDLESE

>Xanthomonas arboricola pv. juglandis

MGNLKPMIANRITTGTTGLDTILRGGLPPNRLYLLEGQPGSGKTTASLQFLLDGAAKGESCLYVTLSETI

DELNEVAASHDWSLEALHIFELASAEAFLGDGRQQSILHPWEMELDGTIKLIQSEVDRVKPTRVVFDSLS

ELRLLAQDSLRYRRQVLALKQFFAPRNITVFLVDDLTGENGERDAHLHSLCHGVISLERMTLDFGAARRR

MQVQKLRGVDFIAGYHDLTIATGGMRIFPRLIACDHHVPFIGDAVSSGVARIDALLHGGPLRGTSTLLTG

PAGSGKTNIALQYVTAACERGEHCCILEFDERTGTLLTRAESLGMDLRKYLDAGLLELHQMDPAELTPGE

FAWAVRASVEQRNCRVLVIDSLNGYLTSMPQEKQLMLQMHELLSYLNQSGVTTFLVNPQHGLVGTMSTGN

LNISYMADTVILFRFFEAQGRIRKAVSVIKNRSGAHEDSIRELRIGTGGIHLSEPLEKFHGVLTGTPHFV

GDDTPLLDRPLDYL

> Escherichia coli

MTDNFYAPPHSIEAEQAVIGGLLLDDDSSERVQKVLAMLKPDSFYSRPHKILFEEITRMHREQKPVDGLT

LFDELERKSLTASVGGFAYIAEIAKNTPSAANIVAYAMQVRETAMERYAINRMTEATELLYSRNGMTATQ

KYEAIQAIFTQLTDHAKTGSRRGLRSFGEVMEDWVSDLEKRFDPSGEQRGMSTGIPSLDRMLSPKGLVKG

SLFVIGARPKMGKTTLYSQMAINCAVHEKKPALMFSLEMPGDQILEKLVGQKSGVNPNIFYLPATNDADD

GYQGDYDGDFNRAIETANRLSEIDLLYIDDTPGLSLAQIVSESRRIKREKGCVGMILVDYLTLMTAEKAD

RNDLAYGMITKGLKNLAKELDCVVVLLTQLNRALESRTNKRPLPSDSRDTGQIEQDCDYWVGIHREGAFD

DSVPPGETELILRLNRHGNTGTVYCIQANGAIYDTDQQSAEMRRREREEPQSKKKGGF

> Xanthomonas campestris pv. campestris

MTSTRIATGTSGLDTILRGGLPSNRLYLLEGQPGSGKTTLALQFLLEGASKGESCLYITLSETLDELQEV

AQSHGWSLDALHIFELASAEAFLGDGRQQSILHPWELELDGTVKLIQAEVDRVKPQRVVFDSLSELRLLA

QDSLRYRRQVLALKQFFAPRNITVLLVDDLTGTSEGRDSHLHSLCHGVVSLERLTLDFGAARRRMQVQKL

RGVDFIAGYHDLSILRGGLRIYPRLIASDHHVTFIGEAVSSGVQEVDDLLHGGPLRGTSTLLTGPAGSGK

TNVALQYVAAACARGEHCCIFEFDERTGTLLARAESLGLDLRRHLESGMLELLQMDPAELSPGEFAWQVR

ASVEQRNCRLLVIDSLNGYLSSMPQEKQLMLQMHELLSYLNQSGVTTFLINPQHGLVGTMSTGNLNISYM

ADTVILFRFFEAQGRIRKALSVIKNRSGAHEDAIRELRIGNSGIRLSAPLNDFHGVLTGTPHFIGDEAPL

LGNDLARR

>Rhizobium leguminosarum

MNTAAPATKARTGVSGLDTVLAGGLSPGHVFLLEGNPGAGKTTIALQFLIEGARLGEQGLYITLSETESE

LRAGAASHGMVIDGNIEIFEVVPPESLLDADQQQSLLYSSDLELGETTKEIFAAFERVKPRRVVLDSLSE

IRLLAQSSLRYRRQILALKHYFARQGATVLLLDDLTSDVLDKTVHSVVHGVIHLEEMAPSYGSERRRLRV

IKYRGQAFRGGYHDFIIQTGGVVVFPRLVAAEHRSSYARDQISCNIAELDLLLGGGLERGSSTLILGPAG

TGKSTFSFQFLVAAVARGEKVAAFIFDEELGLLFTRLKALGIDLEAMRDAGHIHIEQLDAAELSPGEFAH

RVRNCVDKSDAKTVIIDSINGYQASMPDENSLILHMHELLQYLNRQGANTFLTVAQHGLVGDMKAPVDVT

YLADTVILLRYFEAAGKVRRAVSVIKKRTGFHEDTIREYRIDSSGLRFGDPLVGFQGVLRGVPEFIATST

PLLKTDGGDSGNS

>Agrobacterium tumefaciens

MMREIDSKATTGVEGLDDILSGGLTRRHVFLLEGSPGTGKTTISLQFLIEGAVRGERCLYISLSESEEEL

RDSAASHNMVLHENIEIFELVPPESLLDADQQQSLLYSSDLELGETTKLIFDAFERIKPDRVVIDSLSEI

RLLAQSSLRYRRQILALKHFFSKSGATVLLLDDLTADTLDKTVHSVVHGVIHLEELSPNYGSERRRVRVL

KYRGQSYRGGYHDFVIRTGGVAVFPRLRAVEHKTTFVRDTLSSGIDSLDGLLGGGIDRGSSTLLIGPAGT

GKSIFGIQFVAAALARGEKAAMFIFDEELGLLFNRMRHIGIDLEAMRDEGSLYIEQLDAAELSPGEFAQR

VREHVAKYDAKTVLIDSINGYQASMPEENALILHMHELLQFLNRQGANTFLTVAQHGLVGDMKSPVDVTY

LADTVILLRYFEAVGRVRRAVSVIKKRTGMHEDTIREYKINESGLTLGEPISSFQGVLRGVPFLMPEKSI

DSPD

>Synechocystis sp. PCC 6803

MNLPIVNERNRPDVPRKGVQKIRTVIEGFDEITHGGLPIGRTTLVSGTSGTGKTLLAVQFLYQGIHHFDY

PGLFITFEESPSDIIENAYSFGWDLQQLIDDGKLFILDASPDPEGQEVVGTFDLSALIERIQYAVRKYKA

KLVSIDSVTAVFQQYDAASVVRREIFRLVARLKQLQVTSIMTTERVEEYGPIARFGVEEFVSDNVVVLRN

VLEGERRRRTVEILKLRGTTHMKGEYPFTITHDGINIFPLGAMRLTQRSSNARISSGVQTLDEMCGGGFF

KDSIILATGATGTGKTLLVSKFLQEGCRQRERAILFAYEESRAQLSRNASSWGIDFEEMEHKGLLKLLCT

YPESAGLEDHLQMIKSEISEFKPSRIAIDSLSALARGVTNNAFRQFVIGVTGYAKQEEITGFFTNTTDQF

MGAHSITESHISTITDTILMLQYVEIRGEMSRALNVFKMRGSWHDKGIREYSISHDGPDIRDSFRNYERI

ISGSPTRISVDEKSELSRIVRGVKDKTAE

>Bradyrhizobium japonicum

MTSLDDPTRIPTGVSGLDDILGGGLTSHRVYLVEGTPGTGKTTLSLQFLLEGARRGEAGLYITLSETAEE

LSAVAATHGWSLDPLSLFELVNQGELDPEREQTILHPSEVELGETAREVMKQIEELRPSRVVFDSLSEMR

LLAQNPLRYRRQILALKQFFASRDCTVLMLDDRTSDPNDLQLHSIAHGVISLEQAARDFGAERRRLRVVK

MRGIKFRGGYHDFILDRGGLTVFPRLVAAEHGAEFAPKVASTGTPELDELLGGGLMPGTNTLLSGPSGVG

KTTTAMFCIRAALDRGERATYFLFDEGIKTLLVRCKALGIDLASYIERGVLDIQQIDPAELSPGEFTSRA

RGAVEKSGSTFICIDSLNAYMQAMPGEQFLLLQMHELLNYLNQSGITTMLILGQHGLVGDMRSDIDLSYL

SDTILLFRFFEAKAEVRTAISVVKSRTSSHERSIREFKLGAGGLEVGEALRDFEGVLSGLPSYKGDTALL

QSSTN

>Stigmatella aurantiaca DW4/3-1

MTDTPSSQQIFLSGIPSFDALLGGGIPRRQSLIITGDPGCGKTILCGQVAFRAAARDVPVVLATVTSEPH

DKLVQALSGFSFFDQALLGEKLFLVSAYASLKKGAKEARDILLHTVRERGAKLLFIDGLRSIRDLWQDEA

KLREFLYELGIGLAAMDCIGLFTTEYPLEKLRALPEATTVDGIVALDVLRRGGRRLRRMEVVKLRGRNHL

TGEHFMRLGNEGVEVIPRLETQPPHDIDDLPAFKRSSFGLTEFDALMEGGLPSLSTTLLAGSMGIGKTLL

ATHFAAQGAREGEQTLFVSFFEPPGMLIARARRIGLDVAEHLARGTLKMLHLPPSEQEADAIVHHILEEV

ARLGVRRLVVDGLTELERSIVDPERRHLFLATLGAHLRHLGITSLFTKEVSKVAGTELDFNDTPIAMLGE

NLVLLRYVELRGRIHRVLSILKMRDSRYDGNLREFEIKDEGIRVLAPMRSAEGLLTGQARALGTAEGGES

A

>Rhodobacter sphaeroides

MTSQLGIGKSPTGIQGFDELTLGGLPTGRPSLVCGSAGCGKTLFASTFLINGVRDHGEPGVFVTFEERPE

DIVNNVASLGFELDKLIEEEKIAIEHIAVDPSEVAEIGDYDLEGLFLRLELAIDTVGAKRVVLDTIESLF

SAFSNPAILRAEIRRLFDWLKERGLTTVITAERGDGALTRQGLEEYVSDCVILLDHRVENQISTRRLRIV

KYRGTAHGTNEYPFLIDTDGFSVLPVSALGLLHQVHEERIGSGVPDLDAMMTGGGFFRGSSILVSGVAGA

GKSSLAAHFAAAACARGERAMYFSFEEAADQAVRNMRSLGLDLGRWRDAGLLRFMATRPTFYSLEMHLAV

ILREVMRFEPSVVVLDPISAFTESGDRLEVQSMLLRIVDFLKNRGITGIFTHLAHSQNEATTDAGLSSLM

DGWVLMLNREVNGEFNRELYLLKARGMAHSNQVREFLMSDRGISLLPPHLGEGGALTGTARRAEEARLRR

AEVERQTELGRLQQQIEQRRRRARAQIEALEAELQAEEIALKALVESESAHERQRLADADTLARSRGNER

FADLLMNKGE