2016

Genetic Diversity of *Colletotrichum Sublineola* on Sweet Sorghum and Wild Sorghum Relatives in Kentucky and the Southeastern U.S.

Katia Viana Xavier
University of Kentucky, katiaxavier@uky.edu
Digital Object Identifier: https://doi.org/10.13023/ETD.2016.440

Recommended Citation

Viana Xavier, Katia, "Genetic Diversity of *Colletotrichum Sublineola* on Sweet Sorghum and Wild Sorghum Relatives in Kentucky and the Southeastern U.S." (2016). *Theses and Dissertations--Plant Pathology*. 20.
https://uknowledge.uky.edu/plantpath_etds/20

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Katia Viana Xavier, Student

Dr. Lisa J. Vaillancourt, Major Professor

Dr. Lisa J. Vaillancourt, Director of Graduate Studies

GENETIC DIVERSITY OF *COLLETOTRICHUM SUBLINEOLA* ON SWEET SORGHUM AND WILD SORGHUM RELATIVES IN KENTUCKY AND THE SOUTHEASTERN U.S.

DISSERTATION

A dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in the College of Agriculture
at the University of Kentucky

By

Katia Viana Xavier

Lexington, Kentucky

Director: Dr. Lisa J. Vaillancourt, Professor of Plant Pathology

Lexington, Kentucky
2016

ABSTRACT OF DISSERTATION

GENETIC DIVERSITY OF *COLLETOTRICHUM SUBLINEOLA* ON SWEET SORGHUM AND WILD SORGHUM RELATIVES IN KENTUCKY AND THE SOUTHEASTERN U.S.

Anthracnose, caused by the fungus *Colletotrichum sublineola*, is one of the most important diseases on sorghum (*Sorghum bicolor*) in the United States (U.S.) and worldwide. The production of sweet sorghum for feedstock has been increasing in the Southeastern U.S. (SE), and anthracnose has emerged as a significant production constraint. Anthracnose is also common on the ubiquitous wild sorghum relative, johnsongrass (*Sorghum halepense*). The degree to which the pathogen population on johnsongrass contributes to disease epidemics on cultivated sorghum (*S. bicolor*) in the SE is unknown. Genetic and pathological diversity was characterized among a collection of *Colletotrichum* strains recovered from *S. bicolor* and *S. halepense* in Kentucky, Alabama, Georgia, and Florida. The pathogenicity of five *C. sublineola* isolates from sweet and grain sorghum and from johnsongrass was measured on the susceptible sweet sorghum inbred Sugar Drip in the field. Isolates from cultivated sorghum were more aggressive than isolates from *S. halepense,* which generally caused little or no disease. The disease levels observed in the field had no effect on the yields of sorghum biomass, grain, or juice, or on Brix levels. Removal of sorghum seed heads increased sugar levels in the plants, but this had no effect on susceptibility to anthracnose. Greenhouse and laboratory assays were developed that gave rankings of relative strain aggressiveness that were consistent with the field results. Marker analyses with repetitive fingerprinting probes were used to evaluate several hundred *Colletotrichum* strains isolated from *S. bicolor* and *S. halepense* across the SE. Results revealed that, with a few exceptions, isolates from cultivated sorghum were genetically distinct from isolates from *S. halepense*. A restriction fragment length polymorphism (RFLP) analysis based on probes against individual sequences presumed to encode effectors and secondary metabolism enzymes confirmed that most isolates from johnsongrass grouped separately from most isolates from cultivated sorghum. The RFLP analysis revealed the presence of three distinct groups within the population that were distinguished by fixed allelic variations, or by presence-absence polymorphisms, of some of these putative pathogenicity genes. Phylogenetic trees were inferred based on a sampling of isolates from both host species and representing each of the three groups by using the internal transcribed spacer (ITS) sequence of the ribosomal DNA; portions of the DNA lyase gene (*Apn2*) and the manganese superoxide dismutase gene (*Sod2*); and a region between the *Mat1* and the *Apn2* genes (*Mat1/Apn2*). The trees were found to be congruent, and to identify three distinct species, including *C. sublineola* and two previously undescribed species. One of these novel species was found associated only with *S. halepense* throughout the SE. This new species was named *Colletotrichum halepense* (Xavier & Vaillancourt). There was evidence for cross-infection of *S. bicolor* and *S. halepense* by *C. sublineola* and the second new species, which was named *C. caselae* (Xavier & Vaillancourt). These findings have significant implications for the development and deployment of resistant sweet sorghum varieties in areas where johnsongrass is common. This information will help to evaluate the potential for *Colletotrichum* spp. to cause epidemics in sweet sorghum if acreages continue to expand in the SE in the future.

KEYWORDS: Anthracnose, johnsongrass, sweet sorghum, *Colletotrichum caselae*, *Colletotrichum halepense*.

Katia Viana Xavier
December 1, 2016

GENETIC DIVERSITY OF *COLLETOTRICHUM SUBLINEOLA* ON SWEET SORGHUM AND WILD SORGHUM RELATIVES IN KENTUCKY AND THE SOUTHEASTERN U.S.

By
Katia Viana Xavier

Lisa J. Vaillancourt

Director of Dissertation

Lisa J. Vaillancourt

Director of Graduate Studies

December 1, 2016

Date

To my mother, Maria Helena Viana da Silva, my father Jeronimo Xavier da Silva, my sister Karla Viana Xavier, my brothers Bruno Ferreira Xavier and Francisco F. S. Neto, and to her who has been always with me Juliana M. Soares.

Acknowledgments

I am very thankful to all who have been part of my life during my five years at University of Kentucky. In particular, I owe my gratitude to the best advisor in the world, Dr. Lisa Vaillancourt, who taught me how to be a better scientist and who cares about me as my own mother. Thanks to my committee members Dr. Mark Farman, Dr. Hongyan Zhu, Dr. Kenneth Seebold and Dr. Nicole Ward Gauthier for their insightful suggestions and ideas for development of this project.

Additionally, I would like to thank the lab and greenhouse technicians and undergraduate students Doug Brown, Etta Nuckles, Sarah Holton, Andrew Gennett and Justin Judy for their support.

I am very grateful to Ester Buiate, Maria Torres, Misbakhul Munir, for their help, support and friendship. I admire all of them. I am also thankful to Sladana Bec and Douglas Parreira. I extend my gratitude to all staff of the Plant Pathology department for taking care of all the administrative details, allowing me to focus one hundred per cent on my research.

Finally, thank you to all the friends I made in Lexington during these years, and especially to my best friend Juliana, who has been by my side for more than 13 years.

Table of contents

List of tables

List of figures

Chapter 1 - Literature review

Importance and history of sweet sorghum

Sorghum [*Sorghum bicolor* (L.) Moench] ranks fifth in importance among cereal crops worldwide (56). Sorghum is an annual C4 grass with chromosome number 2n=20 (56). It is characterized by its high photosynthetic efficiency, and also by its relative resistance to environmental stresses, including drought, salinity and high temperature (148). Sorghum can produce high yields in marginal soils (81).

The center of diversity for sorghum is in Africa. Early domestication was based on selection and hybridization to increase the number of branches within the inflorescence; decrease the length of the internode rachis; and increase the seed size (86). Sorghum varieties have been further improved by plant breeders by selection of traits such as photoperiod insensitivity, reduced height in order to reduce lodging, drought tolerance, and pest and disease resistance (140). Some of the newer sorghum varieties produce very high levels of green leaf area at flowering and high yields, and also resist insect pests (15).

Cultivated *S. bicolor* is divided into five major races: bicolor; guinea; caudatum; kafir; and durra. A comprehensive analysis of the diversity of sorghum collections worldwide, based mainly on molecular markers, demonstrated that the population is highly polymorphic (1, 4, 75). Different types of sorghum have been selected and optimized for production of grain, fiber, forage, or syrup (209). Sweet sorghum varieties grown for syrup production accumulate high levels of sugar in the stalk, and so they can also be used for the production of ethanol biofuel from cellulosic biomass or directly from sorghum sugar (12, 133, 158, 197, 208). The sweet sorghum plant can serve multiple purposes: the grain can be used for human and animal food, and the bagasse that remains after pressing the stalks, including the stripped leaves, makes a high-quality animal feed (6, 191). Thus, sweet sorghum is of significant and increasing interest as a potential multiuse biofuel crop.

Sweet sorghum is genetically very diverse (1, 142). It was taken from Africa to China (perhaps more than once) and cultivated there for more than 1000 years before being introduced to France in 1851. It was first brought to the United States (U.S.) in 1853 (160). In 1857, Leonard Wray brought 16 additional varieties of sweet sorghum to the U.S. from Africa, and since then the U.S.D.A has introduced more varieties from Australia, China, and Africa. The varieties that Wray introduced include Orange, Sumac, Planter, Gooseneck, Honey, Sapling, African Millet, and Sourless (166). The grain sorghum varieties Brown Durra and White Durra were introduced to the U.S. about 20 years after sweet sorghum. A forage type of sorghum called sudangrass [*S. sudanense* (Piper) Stapf)] was introduced in 1909 (166).

It is believed that almost all the varieties of sweet sorghum grown currently in the U.S. were developed from Wray's introductions, which were improved through selection and hybridization (Figure1.1). The selection of sweet sorghum during the 1900s was done based mainly on agronomic traits, including sugar content (Brix), plant maturity, plant height, biomass yields, and environmental adaptation (166). Today there are approximately 4,000 varieties of sweet sorghum grown throughout the world (77). There are 2,163 accessions in the U.S. sweet sorghum collection (133). Sweet sorghum is grown in the U.S. primarily for the production of table syrup. Kentucky is one of the biggest producers of sweet sorghum, and is among eight states responsible for 90% of the total U.S. output of sweet sorghum syrup (12, 188).

The potential of sweet sorghum as a biofuel

Currently, sugarcane (*Saccharum officinarum* L.) and maize (*Zea mays* L.) are the primary crops used for the production of ethanol biofuel (32, 49, 155). Sugars in the extracted sugarcane juice can be fermented directly, while the starch in maize grains must first be enzymatically converted to sugar before fermentation (192, 197). The sugar content in juice extracted from the stalks of sweet sorghum is up to 20% Brix (compared with 20-24% in sugarcane) (139). Thus, the potential value of sweet sorghum for the direct production of ethanol is similar to that of sugarcane. In addition to the juice, sorghum grains could be used for production of biofuels after enzymatic conversion of starch to sugar, and sorghum feedstock (pressed stalks and stripped leaves) could be used for cellulosic biofuel production. Sweet sorghum is remarkably efficient in production of biomass per acre compared to other potential cellulosic biofuels crops such as switchgrass (*Panicum virgatum* L.) and miscanthus (*Miscanthus* sp.) (37, 110). A study was performed to evaluate the potential of sweet sorghum versus switchgrass for the production of biofuel. Over four years, sweet sorghum had an average production of 15-20t of dry matter ha^{-1}. This was superior to switchgrass, the highest-yielding perennial crop, which produced only 8-15t of dry matter ha^{-1} (81).

Sweet sorghum uses resources more efficiently than either maize or sugarcane. Thus, maize needs 20% more water and 35% more nitrogen than sweet sorghum to produce the same amount of sugar (9, 76, 123, 142, 204). The growing season for sorghum is shorter than that for sugarcane, permitting two crops per year in warmer climates, while sugarcane can produce just one (9, 47).

Sweet sorghum is a seed-propagated annual plant and so, unlike perennial biofuel crops like switchgrass, it can easily be grown in rotations with food crops. This gives the farmer more flexibility and the ability to react to market forces. It also facilitates use of the same crop management practices and farm machinery as for other annual grain crops (134). There have been several studies of sweet sorghum in crop rotations and some advantages have been reported, including improvements in soil physical properties and fertility, soil moisture, increased biological activity, and increased content of organic matter (134). Significant increases in sorghum yield when it was planted in rotation with soybean or with maize, versus continuous cultivation, have been reported (34, 206, 85).

There are many potential advantages of sweet sorghum as a biofuel crop, but significant barriers remain to its widespread use in the U.S. One is a lack of appropriate equipment and infrastructure to harvest and move stalks or juice quickly from the farm to a processing plant. Furthermore, there are relatively few adapted cultivars for the U.S., and we have little information about yield parameters and agronomic qualities. Sweet sorghum is prone to lodging which interferes with machine harvest. We also lack data about disease and pest susceptibility. Rotation of sorghum with maize could increase the risk of losses from insects or disease, since many of the same pests and pathogens affect both crops (145).

Wild sorghum (johnsongrass and shattercane) in the United States

Sorghum bicolor has a large number of wild relatives, and two of the most widespread and important of these are the weeds shattercane (*Sorghum bicolor*) and johnsongrass (*Sorghum halepense* L. Pers.).

The first appearance of shattercane has not been documented, but it is thought to have resulted from hybridization among different varieties of cultivated grain or forage sorghum and wild sorghum relatives (56). Shattercane is so similar to cultivated sorghum that is difficult to selectively control it when it is growing in sorghum fields (23).

The sorghum relative *Sorghum halepense*, thought to be native to the Mediterranean, was introduced into the U.S. before 1830 (5). It was planted and widely promoted as a forage by Colonel William Johnson near Selma, Alabama, when it came to be known as johnsongrass (5, 116). By 1900, johnsongrass had spread throughout the southern and eastern U.S. as a noxious weed. It has hybridized with various cultivated sorghum varieties, and the current population is comprised of numerous ecotypes containing various quantities of the *S. bicolor* genome (132). Johnsongrass is a perennial that reproduces by seeds and by the production of rhizomes, and it is considered to be one of the most destructive weeds in the U.S. and worldwide (84, 116). Johnsongrass could become an especially severe problem in a maize-sorghum-wheat rotation.

Anthracnose on sorghum

The fungus *Colletotrichum* is the casual agent of anthracnose leaf blight and stalk rot of grasses (3, 76, 92, 130). Anthracnose is one of the most important diseases on both maize and sorghum in the U.S. and worldwide (193). Anthracnose on maize is caused by *Colletotrichum graminicola* (Ces.) Wils., while the closely related species *Colletotrichum sublineola* (Henn.) ex Sacc. & Trotter produces similar symptoms on sorghum. *Colletotrichum sublineola* is also reported to be a pathogen of the common related weeds johnsongrass and shattercane (118, 194). Morphological, molecular phylogenetic, and biochemical analyses have indicated that *C. graminicola* and *C. sublineola* diverged from a common ancestor relatively recently (39, 40, 61, 157, 165, 167, 181). There are no reports of *C. graminicola* infecting sorghum, or of *C. sublineola* infecting maize, in the field, and inoculation of most healthy tissues of maize by *C. sublineola*, or of sorghum by *C. graminicola*, does not result in infection. Thus, these two species appear to be host-specific (45, 92, 105, 202). *Colletotrichum sublineola* can infect stalk epidermal cells of healthy maize plants, but the infection does not progress to the stalk interior (187). *Colletotrichum sublineola* can more extensively colonize the leaves and stalks of maize plants that are senescing, drought stressed, or light stressed (Vaillancourt unpublished data) and it can colonize maize leaf sheaths that have been killed by a localized application of liquid nitrogen (175). The factors that are responsible for host specificity of *C. graminicola* and *C. sublineola* are unknown. However, host specificity in plant-pathogen interactions seems in many cases to be related to the presence or absence of pathogen virulence factors, including specialized secondary metabolites (SSMs), and small secreted proteins (SSPs) (7, 36, 48, 59, 90, 103, 180).

Wherever grain sorghum is widely grown, anthracnose is the most important disease problem. It can reduce crop yields by 50% or more under warm and humid conditions in a susceptible variety (3, 83). *Colletotrichum sublineola* can infect all parts of the plant including the stem, peduncle, panicle, grain, and leaf blades and midrib (68). When it occurs in the stem, it produces a disease known as red stalk rot. The disease cycle of *C. sublineola* is shown in Figure 1.2. The pathogen overwinters on crop debris on the soil surface, as conidia and acervuli (193) (Figure 1.2 A). In the spring, when the weather conditions are favorable (i.e. high humidity and warm temperatures) the conidia are disseminated by water splash (Figure 1.2 B). Dried conidial masses can also be spread by the wind to new plants. After a conidium lands on a healthy plant surface (Figure 1.2 C), it geminates to produce a germ tube, which eventually forms an appressorium, (195) (Figure 1.2 D). A penetration peg emerges from the appressorium and forces its way through the plant cell wall. Once inside the plant cell, the pathogen produces thick primary hypha. Primary hyphae are surrounded by a membrane, and the plant cell stays alive during this initial biotrophic phase of development (176, 195) (Figure 1.2 E). Later, the colonized plant cells die and the pathogen switches to necrotrophic growth. During this phase, thin secondary hyphae are produced from the primary hyphae, and the plant cells are destroyed by the activity of lytic enzymes that are produced by these secondary hyphae (175, 176) (Figure 1.2 F). It is at this stage that visible lesions develop. Finally, the pathogen completes its life cycle by producing acervuli containing conidia and setae (195) (Figure 1.2 G). These spores can be disseminated to other plants or tillers during the same season as secondary inoculum, and at the end of the season they serve as overwintering structures (193) (Figure 1.2 H).

Symptoms of anthracnose on the leaf blades and midribs are similar on johnsongrass and on sweet sorghum (Figure 1.3). Symptoms begin as small, round to elliptical spots that vary in color from yellow to red or gray. The spots increase in number and size, eventually coalescing to cover large portions of the leaves. Signs (acervuli containing conidia and setae) are produced in the necrotic centers of the lesions and are visible to the naked eye as small black dots.

The occurrence of anthracnose in both maize and sorghum is associated with stresses during grain filling, including foliar diseases, drought, soil saturation, cold weather, and nutritional deficiencies (55). Reducing the stresses that predispose the plants, and using resistant or tolerant varieties, helps to minimize the impact of the disease. However, no variety of maize or sorghum is immune to anthracnose. Leaf anthracnose has been reported to occur in sweet sorghum and to cause reductions in biomass yield in Kentucky (12). Red stalk rot is common in Kentucky sweet sorghum, as is stalk lodging. The relationship of these syndromes to *Colletotrichum* infection has not been established, although it is often assumed that they are connected.

Anthracnose has been studied most in grain sorghum. The primary control for anthracnose in grain sorghum is the use of resistant cultivars (28). However resistance frequently fails, which is thought to be related to the high degree of genetic variability in the population of *C. sublineola* affecting grain sorghum (25, 27).

Genetic diversity within the *C. sublineola* population on grain sorghum has been characterized by the use of molecular markers, including random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLP) and amplified fragment length polymorphisms (AFLP) (30, 184). The ability of *C. sublineola* to infect a series of different cultivars of sorghum has been used to identify races in this species (3, 14, 30, 143, 184). One study suggested that the population of isolates infecting johnsongrass did not contribute to the population infecting grain sorghum varieties grown in nearby fields in Texas (143). However, there have been no additional studies of this topic in the U.S., and there is no information in the literature about population or race structure of *C. sublineola* on sweet sorghum or weedy sorghum relatives in the southeastern U.S.

The interaction between sorghum and *C. sublineola* was suggested to follow the gene-for-gene pattern (26). Resistance of grain sorghum to *C. sublineola* is controlled mostly by dominant single genes (14, 35, 104). In 1993, Tenkouano reported that one source of resistance in sorghum was controlled by a single genetic locus with multiple allelic forms (170). The value of these major gene sources of genetic resistance has been limited by the ability of the pathogen population to adapt in different locations, and also by variation in the effectiveness of resistance alleles in different genotypes (30, 135). Very little is currently known about the presence or basis for resistance to anthracnose in sweet sorghum.

Although sweet sorghum has been grown in the U.S. for more than 150 years, it has been on a limited scale, primarily for forage and for the production of syrup. If sweet sorghum is cultivated on a commercial scale in the U.S. in the future for biofuel, diseases, including anthracnose, may become a significant limiting factor for production.

Dissertation Goals and Predictions

The goals of my dissertation research were: (1) to study the impact of anthracnose disease on sweet sorghum in the field in Kentucky; (2) to characterize the degree of genetic and pathological diversity among the population of *Colletotrichum* on cultivated and wild sorghum in Kentucky and the southeastern U.S.; and (3) to develop and test several greenhouse and laboratory assays for sorghum anthracnose disease.

In chapter 2 of this dissertation, I compared the pathogenicity of *C. sublineola* isolates from sweet and grain sorghum with that of isolates from the wild sorghum relative johnsongrass on Sugar Drip sweet sorghum in the field. The impact of the anthracnose disease on yields of biomass, seeds, juice, and Brix was determined. The same set of isolates was also tested in greenhouse and laboratory assays, to determine whether these protocols would be useful for screening sorghum germplasm for new sources of resistance or to characterize new isolates. The predictions for chapter 2 were: (1) *C. sublineola* from sweet sorghum will be more pathogenic and/or aggressive on sweet sorghum than isolates from johnsongrass or grain sorghum; (2), anthracnose disease will have a negative impact on yield and Brix; (3) deheading sorghum will result in increased Brix and decreased levels of disease; and (4) the results of greenhouse and leaf sheath assays will be correlated with field inoculations.

In chapter 3, I investigated the genotypic and phenotypic diversity of a large population of *Colletotrichum* strains recovered from cultivated sorghum and johnsongrass, in Kentucky and the Southeastern U.S. The predictions for chapter 3 were: (1) isolates from sweet sorghum will resemble those on grain and forage sorghum more than they will resemble isolates from johnsongrass; (2) the population of strains recovered from johnsongrass will be more diverse genetically than the population recovered from cultivated sorghum; and (3) isolates from different hosts will vary in their virulence and aggressiveness to sweet sorghum varieties.

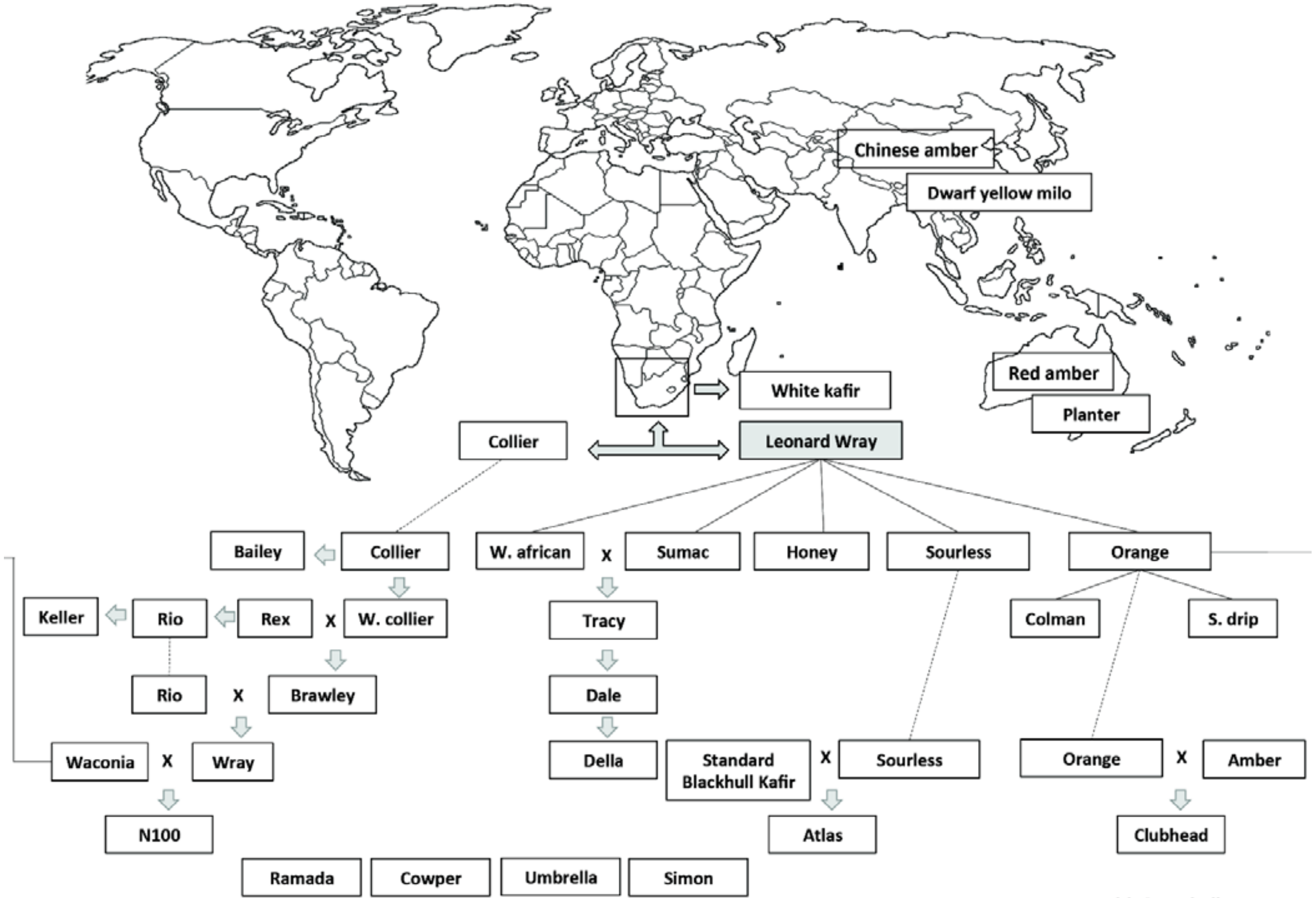
In chapter 4, I characterized and compared *C. sublineola* versus *C. graminicola* genes that were predicted to encode SSM proteins as part of a larger comparative genomic analysis of these two species. I utilized some of these genes, and also some that were predicted to encode SSPs, as probes for an RFLP analysis of a representative sample of isolates from the *Colletotrichum* population affecting cultivated sorghum and johnsongrass in the southeastern U.S. The predictions were: (1) that the sibling species *C. sublineola* and *C. graminicola* would differ in the presence of one or more genes predicted to encode SSPs and SSM-associated proteins (aka pathogenicity genes); (2) that pathogenicity genes found in *C. sublineola* but not in *C. graminicola* would be conserved among all members of the population infecting *Sorghum bicolor* in Kentucky and the southeastern U.S.; and (3) that populations affecting *S. bicolor* would be distinguishable from populations affecting *S. halepense* by allelic differences in individual pathogenicity genes.

In chapter 5, I tested the prediction that genetically distinct groups that I had observed within the populations of *Colletotrichum* affecting cultivated sorghum and johnsongrass actually represent different species. Sequences from the *Sod2* gene, the *Apn2* gene, and from the intergenic region between *Apn2* and *Mat1*[1], have been used previously to identify new species within the *C. graminicola* species complex that affects graminaceous hosts (40). To test my prediction, I analyzed these sequences together with the internal transcribed spacer (ITS) of the ribosomal DNA, as well as measuring various morphological traits in culture, for several representative *Colletotrichum* isolates.

In chapter 6, I present a summary of my dissertation findings and a discussion of their importance and relevance to sorghum cultivation in the region.

Colletotrichum species contain only the HMG-domain MAT gene, which is by convention known as MAT2 (22). Thus, the gene referred to as MAT1 in (24) should really be called MAT2. Nonetheless, in this dissertation I will use the terminology of (24) to refer to the intergenic region adjacent to this gene (*Mat1/Apn2*).

Figure 1. 1. Pedigree of sweet sorghum varieties (Smith and Frederiksen, 2000).



Smith & Frederiksen, 2000

Figure 1. 2. Disease cycle of *Colletotrichum sublineola*.

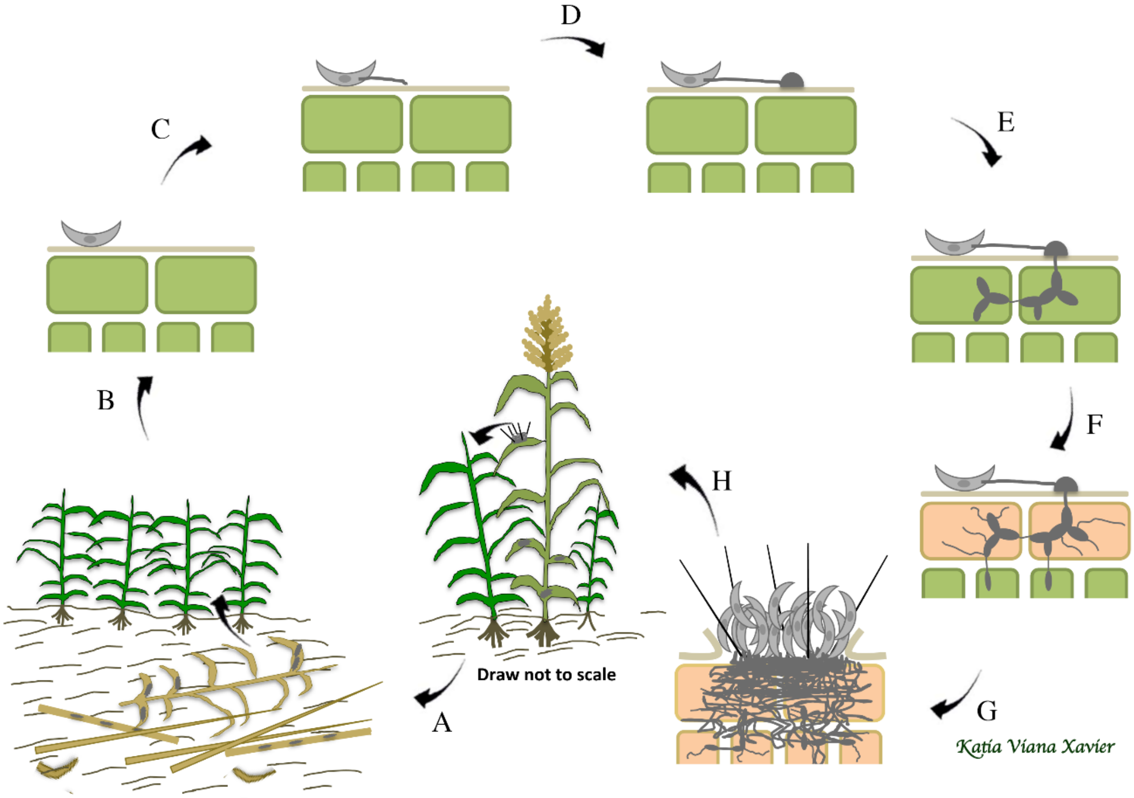
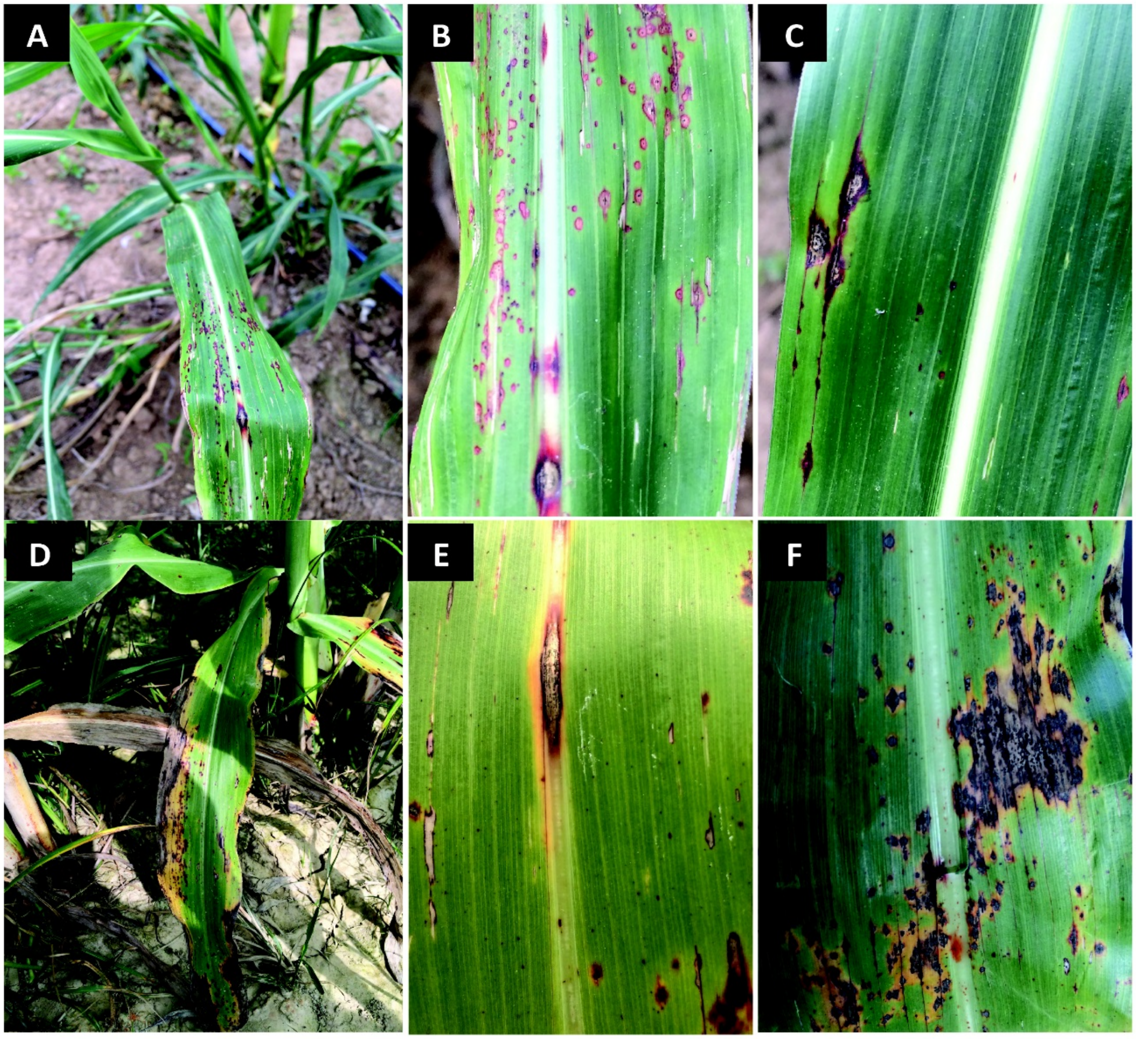


Figure 1. 3. Typical anthracnose symptoms on johnsongrass (A, B, C) and on sweet sorghum (D, E, F).



Chapter 2 - Impact of *Colletotrichum* strains from cultivated and wild Sorghum spp. on the sweet sorghum variety Sugar Drip in field, greenhouse, and laboratory assays

Abstract

Sweet sorghum (*Sorghum bicolor*) has been grown in the southeastern United States for more than 150 years on a relatively limited scale, primarily for forage and for the production of table syrup. However, interest in the crop has increased recently due to its potential as a feedstock for biofuels. *Colletotrichum sublineola* is reported to be the causal agent of anthracnose on cultivated sorghum and on the wild sorghum relative johnsongrass (*S. halepense*). Anthracnose is the most important disease of grain sorghum worldwide, but comparatively little is known about its impact on sweet sorghum in the United States. The pathogenicity of several *C. sublineola* isolates from sweet and grain sorghum and from johnsongrass was measured as disease incidence and severity on the susceptible heirloom sweet sorghum inbred Sugar Drip in the field. The isolate from sweet sorghum was consistently the most aggressive, while the johnsongrass isolates caused only minimal disease symptoms. Disease incidences of up to 95%, and severities of up to 20% of leaf area affected, had no effect on the yield of biomass, grain, juice, or Brix levels in the affected tillers. Removal of sorghum seed heads increased sugar levels in the stalks and leaves, but this had no effect on susceptibility to anthracnose. The same group of fungal isolates was evaluated for pathogenicity in greenhouse assays on juvenile plants, and in the laboratory on seedlings and detached leaf sheaths. Results of these assays were generally consistent with the field results; thus these protocols might be useful for pre-screening

sorghum germplasm for new sources of resistance or for characterizing the pathogenicity of new fungal isolates.

Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is an annual C4 grass characterized by its high photosynthetic efficiency and its relative resistance to environmental stresses, including drought, salinity, marginal soils, and high temperatures (81, 148). Sorghum is grown primarily for grain, which is used mainly for animal feed and ethanol. Sorghum ranks fifth in importance among cereal crops worldwide (57).

In addition to grain sorghum varieties, which have been bred to produce a large, starchy seed head on a short stalk, there are also sweet sorghum types. These produce more biomass than the grain varieties, with many reaching heights of three meters or more, and they accumulate large amounts of soluble sugar in the stalks. Sweet sorghums are not as commonly grown as the grain varieties, and are used traditionally for forage and for the production of sweeteners (142). Recently however, interest in sweet sorghum has increased because of its potential for the production of ethanol biofuel from cellulosic biomass or directly from sorghum sugar (12, 133, 158, 197, 208). Sweet sorghum is well adapted for cultivation in the southeastern (SE) United States (U.S.), where the crop has been grown on a small scale for more than 150 years for the production of table syrup. Kentucky (KY) and its neighboring SE states are responsible for 90% of the total U.S. output of sweet sorghum syrup (12, 188).

Anthracnose is the most important disease of grain sorghum worldwide, and has been reported to cause reductions of as much as 50% in grain yield (83, 173). *Colletotrichum sublineola* (Henn.) causes anthracnose on all varieties of *S. bicolor*, and can infect all parts of the plant including the stem, peduncle, panicle, grain, and leaf blades and midrib (67). Foliar anthracnose is the most common phase of the disease in grain sorghum. Foliar anthracnose occurs everywhere in the world where grain sorghum is produced, and is especially severe under conditions of high temperature and humidity, as are typical of the SE U.S. (3). Infection of the stem produces a disease called red stalk rot which, when severe, can result in lodging. The occurrence of anthracnose on grain sorghum has been associated with stresses during grain filling including drought, soil saturation, cold weather, and nutritional deficiencies (54).

During World War II, epidemics of foliar anthracnose severely limited the productivity of sweet sorghum that was widely grown in the Gulf region to provide a substitute for cane sugar (106). Anthracnose is still common on sweet sorghum in the SE U.S. (12, 79), but relatively little is known about the impact of this disease currently, or its potential for damage if sweet sorghum is grown more widely for biofuel in the future.

The main strategy for the management of foliar anthracnose in grain sorghum is the use of resistant cultivars (28, 124). Resistance fails frequently, and this is thought to be related to the high degree of genetic variability in the population of *C. sublineola* (25, 27). New sources of resistance must constantly be developed and deployed in order to manage the pathogen effectively. Another strategy recommended for the management of anthracnose on sorghum is to use crop rotation with other plant genera. However, *C. sublineola* also reportedly causes foliar lesions on the ubiquitous weed johnsongrass (194), and johnsongrass could serve as a refuge for *C. sublineola* during crop rotation schemes. Furthermore, johnsongrass could be a source of novel pathogenic strains, since this weed is genetically highly diverse (98). A prior study suggested that johnsongrass isolates did not significantly contribute to the populations of *C. sublineola* on nearby grain sorghum in Texas (73, 144). However, the pathogenicity of johnsongrass strains to sweet sorghum has not been directly tested.

The primary goal of the work described here was to compare the pathogenicity of KY *C. sublineola* isolates from different hosts, including johnsongrass, on the susceptible sweet sorghum variety Sugar Drip. A second goal was to evaluate the impact of anthracnose on yields of biomass, juice, and Brix levels. Prior studies have suggested that removing the seed head from sorghum plants increases the amount of sugar in the stalks (63), and other studies have reported a correlation between increased levels of stalk sugar and decreased susceptibility of maize to anthracnose stalk rot (121). Thus a third goal of this study was to evaluate the effect of seed head removal on levels of sugar and disease in Sugar Drip.

Sorghum breeders have developed improved anthracnose-resistant varieties of sweet sorghum for the SE U.S., and this work has increased in importance as the potential of sorghum for biofuels production has become more widely recognized (18, 43, 46, 188). Breeders typically test the performance of new varieties in field trials (64, 156). However, field screening is resource-intensive, and results are complicated by environmental variables including temperature, humidity, and the presence of other abiotic and biotic stresses. A rapid, high throughput resistance assay that could be performed in the greenhouse or laboratory would allow pre-screening of a larger number of germplasms against a wider range of pathogen populations under controlled conditions, thus speeding the process of development of resistant sweet sorghum germplasm. To contribute to this goal, I also measured the pathogenicity of the *Colletotrichum* isolates to Sugar Drip in several greenhouse and laboratory assays in order to test the ability of these assays to predict results in the field.

Materials and Methods

Fungal strains: Six *Colletotrichum* isolates were used in this study (Table 2.1). The isolates were recovered from small pieces of plant tissue cut from the edges of foliar lesions, soaked in 70% ethanol for 30 seconds, transferred to 10% sodium hypochlorite (Clorox®) for 2 minutes, and then washed 3 times with sterile water. Disinfested pieces were blotted on sterile paper towels and then placed on potato dextrose agar (PDA, Difco Laboratories, Detroit, MI, USA) containing ampicillin (0.1 mg/ml). All the isolates were genetically purified by single-sporing and preserved on silica gel at -80°C (179). Strains were identified based on their morphology, including the characteristic falcate spore shape and presence of acervuli with setae, and on blastn matches of the ribosomal DNA internal transcribed spacer (ITS) sequences (>99% identity) to the *C. sublineola* species type (accession NR_111191) or the *C. graminicola* species type (accession NR_111190) (Figure 2.1). All fungal strains were routinely cultured on PDA at 23°C under continuous fluorescent light.

Plants: The heirloom sweet sorghum inbred Sugar Drip was used for all experiments. Seeds were obtained from Dr. Todd Pfeiffer at the University of Kentucky (UK), and increased in 2012 at the UK Agronomy Farm in Lexington, KY. Twelve additional genetically diverse sorghum varieties (Honey, Simon, Chinese Amber, Dwarf Yellow Milo, Cowper, Orange, Planter, Bailey, Dale, Della, N100, and Keller) were used for some experiments (Table 2.2). Seeds of these varieties were obtained from Dr. Gary Pederson at the Plant Genetic Resources Conservation Unit in Griffin GA, and increased in 2012 at the UK Agronomy Farm.

Preparation of fungal inoculum for field and greenhouse experiments: Fungal inoculum was produced by using a modified protocol of Prom and colleagues (136). Forty grams of autoclaved Sugar Drip sorghum grains were placed aseptically in a Petri dish. Falcate spores were harvested from two-week-old fungal PDA cultures by adding 10 mL of sterile water to each plate and scraping the surface gently with a mini-pestle. The spores were collected and washed three times with sterile water to remove the mucilage. A volume of spore suspension was added to achieve a final concentration of $5x10^6$ spores per gram of sorghum grains. The plates were incubated for 14 days at 23°C in the dark to allow the fungus to colonize the grains.

Field experiment: The field experiment was repeated in two sequential years (2014 and 2015) at the UK South Farm in Lexington, KY. Sorghum seeds were sown on 4 May 2014, and 12 May 2015, in 72 cell-flats (each cell 1/2" L x 1 1/2" W and 2 1/4" deep) containing Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada) in the greenhouse. After 4-6 weeks, the seedlings were transplanted into the field with a between-row spacing of 1.4m, and a within-row spacing of 0.7m. A randomized complete block design with four blocks was used. Each experimental unit consisted of two rows of five plants each. Each experimental unit was separated from the others by a row of maize to prevent the dissemination of inoculum between treatments.

Two different experimental plots, both with Maury silt loam soils, were used in the sequential years. Fertilizer (19-19-9 NPK) was applied at a rate of 56 kg/ha in 2014, and 89 kg/ha in 2015, prior to transplanting. The rate in 2015 was higher because the plot used that year had previously been in pasture. Weeds were controlled mechanically: no herbicides were used. Drip irrigation was applied as necessary.

Approximately five weeks after transplanting, the main stem of each plant was cut back to the crown to promote the development of multiple tillers per plant. Tillers were inoculated by using a modified protocol of (136). Tillers were inoculated at the V5 or V6 growth stage by placing 20 colonized sorghum grains (see above) into the whorl. Non-colonized grains, which otherwise had been treated identically to the colonized grains, were used as a water control treatment. The optimal growth stage for inoculation was determined in preliminary experiments demonstrating that maximal levels of disease resulted when plants were inoculated at this stage (See Appendix 1 of this dissertation). Two randomly selected tillers per plant were inoculated, for a total of 20 inoculated tillers per experimental unit.

The distribution and amount of rainfall differed in 2014 and 2015. The amount of rainfall between transplanting and inoculation was much less in 2014 (56.6 mm) than in 2015 (233.2 mm). However, after inoculation, rainfall was much higher in 2014 (218.7 mm) versus 2015 (81.8 mm) (http://www.kymesonet.org/historical_data.php).

Seed head removal (de-heading) was accomplished by removing all the developing panicles continuously from all of the plant tillers as they appeared, from boot stage until harvest. Control treatments (heads intact) retained the seed heads until harvest. The treatments were as follows: (1) SS1.001, heads intact; (2) GS.1.001, heads intact; (3) JG9.001, heads intact; (4) JG18.001, heads intact; (5) water control, heads intact; (6) M30.001, negative control, heads intact; (7) SS1.001, de-headed; (8) GS1.001, de-headed; (9) JG9.001, de-headed; (10) JG18.001, de-headed; (11) water control, de-headed; (12) M30.001, negative control, de-headed.

Inoculated tillers were harvested by cutting them at the crown at physiological maturity, 45 days after flowering. Five inoculated tillers were randomly selected from each treatment for yield measurements. In 2014, seed heads were removed from the tillers, but leaves were retained, while in 2015, both leaves and seed heads were removed prior to analysis. The tillers were weighed, and then pressed at 12.4 MPa of pressure in a grooved, two-roller mill powered by a 1.5-HP electric motor. The juice was collected, weighed, and the Brix was measured immediately with a hand-held refractometer (ATAGO, Tokyo, Japan). One degree Brix corresponds to one kilo of fructose per 100 kilos of solution. Mature seed heads were collected from the stems and pooled for each replication of the head-intact treatments, and the seeds were threshed and weighed.

Greenhouse experiment: The ability of the KY *Colletotrichum* strains to cause disease symptoms on the sweet sorghum variety Sugar Drip was evaluated in the UK Department of Plant Pathology greenhouse in Lexington, KY. Two replicate experiments were conducted. Four sorghum seeds were sown per 8-inch pot, in a mixture of three parts Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada) and two parts sterile topsoil. The pots were maintained on the bench with 14 hours of light, watered every other day to saturation by using an automated overhead watering system, and fertilized with a solution of 150 ppm of Peters 20-10-20 (Scotts-Sierra Horticultural Product Co., Marysville, OH) once a month, beginning one week after seedling emergence. The pots, each containing four plants, were arranged in a randomized complete block design with four blocks. Each experimental unit was comprised of three pots containing 12 plants. Plants were grown to the V5 or V6 stage, and then inoculated by placing 15 colonized sorghum grains (see above) into the whorl of each plant. Non-colonized grains, which otherwise had been treated identically to the colonized grains, were used as a negative control treatment. Treatments were as follows: (1) SS1.001; (2) GS.1.001; (3) JG9.001; (4) JG18.001; (5) water control; (6) M30.001, negative control.

Disease evaluation for field and greenhouse experiments: Anthracnose disease symptoms were evaluated on a weekly basis for the field and greenhouse experiments beginning at 14 days after inoculation, for a period of four weeks. A standardized rating scale was used, with values ranging from 0-5, based on the one described and used in (22, 25, 184). On this scale, a “0” indicates no symptoms or very minor red flecks; a “1”is defined as small red dots on the leaves; a “2” indicates small-to-large red spots on the leaves; a “3” is lesions with necrotic centers but without acervuli; a “4” specifies necrotic lesions with acervuli in the centers of the lesions; and a “5” indicates the presence of coalescent necrotic lesions and abundant production of acervuli (Figure 2.2). This scale divides samples into two classes: resistant, including values of 0-3; and susceptible, including values of 4 and 5. Disease incidence was calculated as the number of infected (susceptible, ratings of 4 or 5) plants at each time point for the field and greenhouse experiments.

Disease severity was also evaluated in the field experiments at a single time point, six weeks after inoculation. Leaves were collected from the 6th and 7th internodes from five randomly chosen inoculated tillers per treatment, brought back to the laboratory, and scanned (by using an Epson GT-1500 scanner). The percentage of unhealthy (necrotic or chlorotic) tissue was evaluated for each leaf by using the software Quant (182).

Laboratory Assays:

Seedlings: Sugar Drip sorghum seeds were sown in 72 cell-flats (each cell 1/2" L x 1 1/2" W and 2 1/4" deep) containing Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada) in the greenhouse. Seedlings were maintained in the greenhouse with 14 hours of light, and watered every other day to saturation using an automated overhead irrigation system. After seven days, seedlings were brought to the laboratory and inoculated by pipetting 10 μl of a spore suspension (10^7 spores per mL) into the whorl. Spore suspensions were prepared from 2-week-old PDA cultures and used immediately for inoculations. An experimental unit was a group of 11 seedlings. The experimental units were arranged in a completely randomized design. The inoculated seedlings were maintained in a humidity chamber at 23 °C under continuous light for 10 days. Disease evaluation was performed by using a dissecting stereoscope. Disease incidence was recorded as the number of seedlings with visible acervuli and pathogen sporulation.

Detached leaf sheath: Seeds of Sugar Drip sweet sorghum were sown in 3.8 x 21 cm containers (Super SC-10 UV stabilized, Stuewe & Sons, Inc. Oregon, USA) containing a mixture of two parts sterile topsoil and three parts Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada). Seedlings were maintained in the greenhouse with 14 hours of light, watered every other day to saturation using an automated overhead irrigation system, and fertilized beginning one week after emergence two or three times per month as needed with a solution of 150 ppm of Peters 20-10-20 (Scotts-Sierra Horticultural Product Co., Marysville, OH).

The detached sheath assay was based on the protocol described by Kankanala et al. (97), as modified by Torres et al. (97, 175). Briefly, sorghum seedlings at the V5 stage were cut at the soil line and brought back to the laboratory (see Appendix 2 of this dissertation for details regarding the optimization of this assay). Leaf sheaths of the youngest fully expanded leaves were removed and cut into three-inch segments. A pipette was used to fill the rolled segments with 150 µL of a fungal spore suspension ($5X10^5$ spores per mL). Water was used as a negative control. The filled sheaths were supported, with the midrib at the bottom, inside a Petri dish containing moistened Whatman paper (Whatman No. 1, Whatman, Hillsboro, OR, USA) to maintain high humidity, and incubated at 23 °C under continuous light for periods of up to 96 hours. Treatments were as follows: (1) SS1.001; (2) GS.1.001; (3) JG9.001; (4) JG18.001; (5) CgSl1; and (6) M30.001, negative control. The treated sheaths were arranged in a completely randomized design.

After incubation the leaf blades were removed from each sheath segment, leaving only the midrib section. This was inverted on a wet microscope slide, and a single-edged razor blade was used to remove excess tissue from the abaxial surface until only a thin layer of mostly epidermal cells remained. For evaluation of disease incidence, a total of 50 fungal infection sites (consisting of appressoria and associated structures, if any) on each of 20 leaf sheaths were observed with the light microscope for each treatment at 60 hpi, and the number that had invaded more than one host cell was recorded.

Statistical analysis:

<u>Field and Greenhouse Experiments:</u> Analysis was performed with SAS PROG GLM (version 9.3, SAS Institute, Inc., Cary, NC) to test the effect of removing the seed head of the plants, and to evaluate the relationship between plant yield and BRIX and the aggressiveness of the fungal strains. The incidence and severity data from 2014 and 2015 experiments were analyzed using Least Squares Means PROC GLM, and multiple mean comparisons among the different treatments were performed with t-tests. General linear

model was used with post-hoc comparisons. P-values less than 0.01 were considered significant. Incidence values were used to calculate the area under the disease progress curve (AUDPC) by the trapezoidal integration method (20). Analyses among the treatments were evaluated by using the general linear models (GLM) procedure of SAS (version 9.4) (SAS Users Guide, SAS Institute, Cary, NC). Means were separated with least squares estimates of marginal means (LSMEANS statement). P-values less than 0.05 were considered significant.

The analysis for the grain yield was performed using SAS version 9.3. Treatment means were compared using the Tukey's multiple comparison test procedure of the SAS software package (SAS Users Guide, SAS Institute, Cary, NC). P-values of less than 0.05 were considered significant.

Whole seedling and detached sheath assays: The analyses for the detached leaf sheath and whole seedlings assays were performed using SAS version 9.3. The data were subjected to analysis of variance (ANOVA). Treatment means were compared using the Tukey's multiple comparison test procedure of the SAS software package (SAS Users Guide, SAS Institute, Cary, NC). P-values of less than 0.05 were considered significant.

Results

Field experiments

Effects of De-heading: Brix values for the intact versus de-headed treatments were significantly different ($P<0.01$) in both years of the study (Table 2.3). The year-by-treatment interaction was not significant. There was no statistical effect of the removal of the seed heads on the yields of plant biomass or juice in either year (Table 2.3). Yields of both biomass and juice were statistically reduced in 2015 in comparison with the year before.

Development of Disease in the Field: In both years, typical anthracnose disease symptoms developed, within two days after inoculation (dpi), on the midrib and on the leaf blades of the inoculated plants. Symptoms were first visible as small, elliptical reddish spots that increased in number and size, eventually coalescing to cover large portions of the leaves (Figure 2.3 A). Signs (acervuli containing conidia and setae) were produced in the necrotic centers of the lesions within 5 dpi. The pathogen spread from the inoculated tillers, resulting in the development of symptoms and signs on the leaves of neighboring, uninoculated tillers by the end of the experiment in the most heavily affected treatments. However, typical disease symptoms and signs were not observed in the water or non-pathogen control treatments, indicating that there was no natural infection, and no inter-plot pathogen movement.

De-heading the plants had no effect on the disease incidence, or on disease severity with the exception of SS1.001 in 2014, which produced more severe symptoms on the de-headed treatment (Table 2.4). Since de-heading had no effect on disease incidence or yield, the control and de-headed treatments were pooled for subsequent analyses of these parameters.

Disease Incidence and Severity: Generally, disease incidences were higher in the field in 2014 versus 2015 (Table 2.5, Figure 2.4). All of the strains from *Sorghum* spp. resulted in incidences of disease that were statistically greater than the negative controls (water and the strain from maize) in 2014. However in 2015, only the isolates from *S. bicolor* and johnsongrass isolate JG18.001 were different from the negative controls (Table 2.5).

The amount of leaf damage was also higher in 2014 than in 2015 (Table 2.5, Figure 2.5). Disease severity did not exceed 20% of the leaf area damaged on the inoculated tillers for any treatment in either year of the study. The sweet sorghum and grain sorghum isolates were the most damaging, while the johnsongrass isolates were not statistically different from the negative controls in either year (p value > 0.01) (Table 2.5).

Although the overall amount of disease was different between the two years, the relative rankings of isolate aggressiveness were consistent. In both years of the study SS1.001, which was isolated from sweet sorghum, was the most aggressive, while the two strains isolated from johnsongrass (JG18.001 and JG9.001) produced much lower levels of disease.

<u>Effect of anthracnose on plant yield and Brix:</u> Pooled data from intact and de-headed sub-treatments by year revealed that yields of plant biomass and juice and Brix levels in all but one case, did not differ from the two controls for any level of disease that was achieved during the two years of this study (Table 2.6). Brix for the water control in 2014 was lower than for the treated samples (Table 2.6). Grain yields were also unaffected by any of the disease treatments (Table 2.6).

Greenhouse experiments

Two repetitions of the greenhouse experiment were performed. Symptoms started to appear within 2 dpi as red spots on the leaf blade and midrib. Within 5 dpi, necrotic lesions containing *Colletotrichum* fruiting structures were visible. Symptoms produced in the greenhouse were similar to those in the field, although they were generally much milder (Figure 2.3B). Incidences of disease resulting from inoculation with the strains isolated from cultivated sorghum (SS1.001 and GS1.001) were statistically higher than the control treatments (p value > 0.01) in both replications of the experiment. In the first repetition, SS1.001 was the most aggressive while in the second replication, SS1.001 and GS1.001 were equally aggressive (p value = 0.001). The strains from johnsongrass did not differ statistically from the water or non-pathogen controls in either replication (p value > 0.01) (Table 2.5). Both of the johnsongrass isolates caused typical foliar anthracnose lesions, including acervuli with setae, when the spores were applied to leaves of johnsongrass plants grown from seed collected in Fayette County KY, confirming that they were pathogenic to *S. halepense* (data not shown).

Seedling assays

Due to the necessity for containment, the Indiana grain sorghum isolate CgSl1 was not used for field assays. All three isolates from cultivated sorghum (CgSl1, SS1.001, and GS1.001) produced typical dark, sporulating anthracnose lesions on the seedlings within 10 dpi (Figure 2.3C). In the first replication of the study, these three isolates could be statistically separated based on incidence, with CgSl1 being the most aggressive of the three, and GS1.001 the least (p value < 0.05). In the second replication, the isolates were not statistically different from each other (p value < 0.05) (Figure 2.3C, Table 2.5). The isolates from johnsongrass caused only relatively mild symptoms, and did not differ in incidence from the negative controls in either replication (Figure 2.3C, Table 2.5).

Detached leaf sheath assays

The process of infection of sweet sorghum sheaths by *C. sublineola* was evaluated over a time course from 0 to 108 hours after inoculation (hpi) (Figure 2.5). At 0 hpi, only ungerminated spores were visible on the surface (Figure 2.5 A). The pathogen had formed appressoria by 12 hpi (Figure 2.5 B), and small primary hyphae could be seen within the host epidermal cells by 24 to 48 dpi (Figure 2.5 C, D). By 60 hpi, the hyphae in a compatible interaction had progressed to colonize adjacent cells (Figure 2.5 E).

Colonization continued to expand into additional cells by 72 hpi (Figure 2.5 F), and narrower secondary hyphae had begun to develop as branches from the hyphae behind the colonization front (not shown). By 108 hpi, tissue collapse was evident, and sporulation had occurred (Figure 2.5 G).

Based on observations of hundreds of individual inoculation sites, I found that whenever the pathogen was able to colonize more than one plant cell within 60 hpi, it almost always progressed to sporulation. On the other hand, if the pathogen had not colonized beyond one cell in that time, its growth was arrested and it almost always failed to complete its life cycle even up to 6 days after inoculation. Thus, I considered that a successful infection had occurred if the pathogen had progressed beyond one plant cell within 60 hpi. Isolate CgSl1, from grain sorghum in Indiana, was the most aggressive on Sugar Drip in this detached leaf sheath assay (Figure 2.3D, Figure 2.6). SS1.001 and GS1.001 did not differ from one another, but both were significantly more aggressive than the johnsongrass isolates, which did not differ from the negative control (the maize isolate M30.001) (Figure 2.3D, Figure 2.6).

Sheaths of 12 additional sorghum lines were inoculated with the six fungal strains, and the johnsongrass isolates did not differ from the negative control on any of these lines (Figure 2.7). The ability of the three isolates from *S. bicolor* to colonize each of the lines varied. The grain sorghum isolates CgSl1 and GS1.001 were each able to colonize five of these lines, whereas SS1.001 could only colonize one (Chinese Amber). None of the isolates successfully infected Cowper or Keller.

Discussion

Anthracnose is listed as one of the three most important diseases of sweet sorghum in KY (12). I made multiple visits to KY sweet sorghum growers in Clay, Russell, and Morgan Counties during the summers of 2013, 2014, and 2015. Foliar anthracnose lesions could be found on sweet sorghum (varieties Dale and M81E) and johnsongrass on all of these farms each year, but overall disease levels were very low. Red rot and lodging in sweet sorghum are often attributed to *C. sublineola*, however red rot was not observed during the years of this study. Lodged plants were relatively common, but I did not recover *Colletotrichum* from any of the lodged plants.

The results of the field study demonstrated that *C. sublineola* has significant potential to cause anthracnose disease on a susceptible variety of sweet sorghum in KY. Sugar Drip is an heirloom variety of unknown origin that was very popular in KY up through the 1980s. It is fast-growing, thus well adapted to the state and suited for late planting, and it produces very high quality syrup (12). However, it is highly susceptible to many sorghum diseases, including anthracnose, and for this reason it has been largely displaced by more resistant varieties.

Incidences of foliar anthracnose on inoculated tillers of up to 95%, and severities of close to 20% of the leaf area affected, were achieved in the field experiment, and I also observed frequent secondary spread of the disease to neighboring uninfected tillers in the most heavily affected plots. No red rot or lodging was observed, suggesting that the pathogen did not significantly colonize the stalk from the foliage when it was inoculated at this stage of development. Disease levels achieved in my experiments had no negative impact on the yield of biomass, grain, juice, or Brix. Given that the levels of disease in my experimental fields were much higher then I observed at the collection sites, I conclude that the economic impact of this disease in KY on yield at present is probably minimal. It is important to point out that I inoculated two tillers per plant. Most of the plants eventually produced up to 10 additional tillers, and these may have compensated for the tissue loss by remobilizing photosynthate to the diseased tillers (159).

Removal of seed heads has been recommended as a way to increase yields of sugar for syrup (13). Maize anthracnose is known to be negatively affected by increases in stalk sugars (121), and so I wanted to see if removal of seed heads would also increase resistance of sweet sorghum to anthracnose. In the current study, removal of seed heads did increase the Brix, but this had no significant effect on the levels of foliar anthracnose.

Although I followed the same experimental protocol in both years of the study, disease levels varied significantly between years. This was probably due to environmental factors, especially the amount of rainfall. In 2014, rainfall levels were unusually high throughout the season. It rained on the day of inoculation, and it continued to rain frequently up through harvest. In 2015, there was torrential rainfall and flooding at the beginning of the season, but it stopped before the inoculation, and after that the weather became hot and dry. The higher levels of leaf moisture during and subsequent to inoculation probably favored the development of the pathogen in 2014, leading to more disease. One of the johnsongrass isolates did not differ from controls in 2015, but did in 2014. This suggests that high levels of leaf wetness, and low light intensities resulting from overcast conditions, may be important determinants of infectivity for less aggressive isolates. Similar results have been reported for anthracnose on grain sorghum (135).

Despite higher levels of disease, yields of biomass, grain, and juice were significantly higher in 2014 than in 2015. The difference in biomass may be due in part to the removal of leaves from the stalk samples in 2015 prior to pressing, whereas they were included in 2014. Additionally, the plants were noticeably damaged and stunted early in the season in 2015, probably as a result of the unusually high levels of rainfall and flooding that occurred just after transplanting (Figure 2.8).

Control plots (mock-inoculated or inoculated with the maize *Colletotrichum* isolate) were disease-free, indicating that there was little or no natural inoculum in the fields. This is in spite of the fact that johnsongrass was common in the grassy areas surrounding the fields, and that *Colletotrichum* could be readily recovered from these plants. My observations are consistent with a prior study that reported that johnsongrass isolates did not contribute to the population of *C. sublineola* infecting nearby grain sorghum in Texas (73, 144). Even though they were significantly less aggressive, the KY johnsongrass isolates in the current study were nonetheless both able to infect inoculated Sugar Drip plants in the field and complete their life cycle. This suggests that, if conditions for infection are ideal, johnsongrass could serve as a reservoir for the pathogen during crop rotation schemes, and as a source for new races.

The sweet sorghum strain SS1.001 used in my study was isolated from a foliar lesion on the sorghum variety Sourless on the UK Agronomy Farm in Fayette County. This isolate was the most aggressive to Sugar Drip in both years in the field study, but with so few isolates included in the experiment it would be premature to conclude that sweet sorghum isolates are generally more aggressive than isolates from grain sorghum to sweet sorghum. It is necessary to test more isolates in order to address this question. Field experiments are limited by time and resources, and it is possible to investigate only relatively small numbers of local isolates. I developed and tested three alternative greenhouse and laboratory (controlled environment) assays for more rapid evaluation of larger numbers of strains, and exotic strains that require containment.

All three of the controlled-environment assays gave results that were consistent with the field results, in that they confirmed that the johnsongrass isolates were significantly less aggressive to Sugar Drip than the isolates from cultivated sorghum. Tests on 12 additional lines using the sheath assay revealed that the johnsongrass isolates were unable to infect any of them. In the greenhouse and laboratory assays, the two johnsongrass isolates were completely nonpathogenic, but the field assays demonstrated that these isolates did have some potential to complete their life cycles on Sugar Drip sweet sorghum, especially when environmental conditions were highly favorable to the development of disease, as occurred in 2014. Thus, a lack of pathogenicity in the controlled environment assays will not necessarily mean that the interaction will be incompatible in the field, but it does suggest that the disease severity will be relatively low. The ability of the three isolates from *S. bicolor* to colonize the sheaths varied depending on the sorghum variety, and this could suggest the presence of race interactions. *Colletotrichum sublineola* is known to exist as multiple races (25, 27). Two of the sorghum lines, Cowper and Keller, were resistant to all three of the *S. bicolor* isolates in the sheath assay. Tests involving whole plants in the greenhouse and in the field will be necessary to confirm these results.

There were higher levels of disease in the field in both years than in the greenhouse. Similar results were observed by Ferreira and colleagues when evaluating 23 grain sorghum cultivars for resistance to *C. sublineola* in the greenhouse and in the field (67). Plants in the field are exposed to multiple stresses including insects; wounds that facilitate pathogen entry; abiotic stressors (wind, high temperature, drought or flooding, hail, etc.); and biotic disease (bacteria and other fungi). In contrast, the plants in the greenhouse are maintained at optimal conditions of temperature and moisture and protected from most other biotic and abiotic stressors. Stress is well-known to predispose grain sorghum to anthracnose (54). Disease incidence was higher in the second biological repetition of the greenhouse experiment than in the first. The second replication took place in the mid-winter months, whereas the first was done in the late summer, and thus light intensity may have played a role. It has been reported that higher light intensities result in lower anthracnose disease severity in maize, due to an increase in the ability of the plant to express resistance (82, 126, 150). Although supplemental lighting is provided in the greenhouse, it does not match natural summer light intensities. The sweet sorghum isolate (SS1.001) was significantly more aggressive than the grain sorghum isolate (GS1.001) in the first replication, similar to the field observations. However, it did not differ from the grain sorghum isolate in the second replication, which may indicate that a reduction in resistance caused by lower light intensities favors less aggressive pathogens more than highly aggressive ones.

The sheath assay provides a uniquely detailed view of the host-pathogen interaction in living tissues. Synder at al. (1991) described the accumulation of red vesicles at the infection sites of the non-pathogen *C. graminicola* (161). Wharton and Julian (1996) compared the colonization of susceptible versus resistant grain sorghum leaf sheaths by *C. sublineola*. They described how the resistant plants responded by accumulating red vesicles at the sites of infection, which eventually burst and released their contents, filling the colonized cells and also often coating the fungal hyphae as well (195). In contrast, in susceptible interactions the accumulation of pigmented vesicles was relatively slow, and the pigmentation was more orange than red (195). It has been shown that the dark red vesicles formed in incompatible interactions with *C. graminicola* contain anthocyanidin phytoalexins, which have also been strongly implicated in the resistance of sorghum to *C. sublineola* (89, 112, 129). My observations in sweet sorghum leaf sheaths (Figure 2.3D) were generally consistent with the published observations of grain sorghum. Thus, I observed a rapid and intense production of vesicles containing a bright red pigment in Sugar Drip responding to the non-pathogen *C. graminicola*. None of the strains from sorghum produced as intense a reaction in Sugar Drip sheaths. Strains SS1.001 and GS1.001 (both virulent) induced the production of a diffuse reddish pigment throughout the colonized cells, whereas CgSl1 (virulent) and JG9.001 (avirulent) didn't cause any change in color. Strain JG18.001 (avirulent) caused only a slight accumulation of red pigments surrounding the infection hyphae. Production of these more diffuse reddish compounds was not correlated with resistance, and it is possible that they are not related to the anthocyanidin phytoalexins that have been previously described. Biochemical assays will need to be performed to address the question of their identity and significance. The sheath assays are well-suited for this kind of detailed analysis of the host-pathogen interaction.

Summary

My predictions for chapter 2 were: (1) *C. sublineola* from sweet sorghum will be more pathogenic and/or aggressive on sweet sorghum than isolates from johnsongrass or grain sorghum; (2), anthracnose disease will have a negative impact on yield and Brix; (3) de-heading sorghum will result in increased Brix and decreased levels of disease; and (4) the results of greenhouse and leaf sheath assays will be correlated with field inoculations. Prediction 1 was supported, in that SS1.001 was the most aggressive isolate in the field study. However, it was not always the most aggressive in the other assays, and since I could include only one isolate from sweet sorghum, it seems premature to conclude that sweet sorghum isolates will always be more aggressive to sweet sorghum. Prediction 2 was not supported. The levels of disease achieved in my field study had no negative effect on yield parameters or BRIX. However, since I only inoculated only two tillers, and the plants eventually developed additional tillers, these uninoculated tillers may have been able to compensate for the diseased tissues. Prediction 3 was partially supported. De-heading the sorghum plants did increase BRIX. However, this increase had no effect on levels of anthracnose disease. Prediction 4 was supported. Greenhouse and laboratory assays generally agreed with the results of the field assays, in that they confirmed that the two johnsongrass isolates were much less aggressive than the isolates from cultivated sorghum.

Table 2.1. *Colletotrichum* isolates used in this study.

Strain	Host	Species	State	County	Year Collected	ITS sequence Accession	Reference
CgSl1	Grain sorghum	*C. sublineola*	IN	Tippecanoe	1982	KY006070	Jamil & Nicholson, 1987
SS1.001	Sweet sorghum	*C. sublineola*	KY	Fayette	2012	KY006074	This study
GS1.001	Grain sorghum	*C. sublineola*	KY	Hopkins	2013	KY006071	This study
JG9.001	Johnsongrass	*C. sublineola*	KY	Russell	2012	KY006073	This study
JG18.001	Johnsongrass	*C. sublineola*	KY	Fayette	2012	KY006072	This study
M30.001	Maize	*C. graminicola*	KY	Fayette	2012	KY006075	This study

This chapter has been submitted for publication to *Plant Disease* as: Xavier, K.V., Pfeiffer, T, Parreira, D.F., Chopra, S., and Vaillancourt, L., Impact of *Colletotrichum* strains from cultivated and wild sorghum spp. on the sweet sorghum variety Sugar Drip in field, greenhouse, and laboratory assays.

Table 2.2. Pedigree and origin of sweet sorghum varieties used in the leaf sheath assay.

Variety	Origin	Year Imported/Released	Pedigree (If Known)	References
Chinese Amber	China	1853	Unknown	Smith and Frederiksen 2000
Honey	South Africa	1857	Unknown	Swanson & Laude, 1934
Orange	South Africa	1857	Unknown	Swanson & Laude, 1934
Planter	South Africa	1857	Unknown	Swanson & Laude, 1934
Dwarf Yellow Milo	Texas and Oklahoma	1900	From Standard Yellow Milo (Texas – 1890)	Swanson & Laude, 1934
Dale	Mississippi	1973	Tracy/MN 960 (PI 152857)	Ali, Rajewski et al., 2008
Keller	Mississippi	1982	Mer. 50-1/Rio	Ali, Rajewski et al., 2008
Bailey	Georgia	1984	Wiley (Collier/MN822/MN2046)/Tracy	Ali, Rajewski et al., 2008
N100	Nebraska	1990	Waconia/Wray	Ali, Rajewski et al., 2008
Della	Virginia	1993	Dale*2/ATx622	Ali, Rajewski et al., 2008
Cowper	Unknown	Before 1951	Unknown	LeBeau, Stokes et al., 1951
Simon	Unknown	Unknown	Unknown	Ali, Rajewski et al., 2008

Table 2.3. Effect of removing the seed head of the plants on the plant yield (fresh biomass and juice content) and BRIX.

Year	Plant top	Brix ¥∂	Fresh biomass∫	Juice content∫
2014	Intact Head	12 a ß	4.74 a	2.4 a
	De-head	14.1 b	5.0 a	2.5 a
2015	Intact Head	15.2 c	1.17 b	0.66 b
	De-head	16.4 d	1.05 b	0.62 b

¥ One (1) degree Brix corresponds to 1 Kg of fructose in 100 Kg of solution
∫ Data average weight in Kg of five inoculated tillers
∂ Data average based on extraction of five inoculated tillers
ß Means followed by the same letter, for each column, are not significantly different according to Least Square Means (t-test significant at P = 0.001)

Table 2.4. Effect of heading and de-heading on incidence (transformed in area under the disease progress curve) and severity (percentage damaged area) in the field.

Treatments	2014				2015			
	Incidence[¥π]		Severity		Incidence[¥π]		Severity	
	Head	De-head	Head	De-head	Head	De-head	Head	De-head
GS1001	48.5 b	51.8 b	8.2 c	9.5 bc	25 b	20 b	7.9 bcd	11.4 abc
SS1001	66.8 a	72.8 a	12.4 b	16.7 a	59.3 a	59.5 a	16.8 a	12.8 ab
JG18001	33 c	27.8 cd	2.8 ed	3.9 d	5.8 cd	7.4 c	3.1 de	2.6 de
JG9001	19 d	20.5 d	3.2 ed	2.4 ed	1.6 cd	5.7 cd	4.7 de	5.8 cde
M30001	0 e	0 e	0.8 e	2.2 ed	0 d	0 d	2.0 e	2.0 e
Water	0 e	0 e	1.1 ed	1.6 ed	0 d	0 d	2.9 de	1.2 e

¥ Incidence, number of infected tillers out of 20 tillers in the field.

π Incidence data was used to calculate the least-square means for area under the disease progress curve (AUDPC).

Means followed by the same letter, for each column, are not significantly different according to Least Square Means (t-test significant at P = 0.001).

Table 2.5. Field and greenhouse pathogenicity assay of five strains of *Colletotrichum* on sweet sorghum var. Sugar Drip.

Treatments	Field		Greenhouse		Lab/seedlings	
	AUDPC $^{\yen\pi}$		AUDPC $^{\yen\pi}$		Incidence$^{\yen}$	
	2014	2015	1	2	1	2
CgSl1	NA	NA	NA	NA	2.7 a	6.5 a
GS1.001	50.1 b	22.5 b	1.3 b	23.3 a	1 c	6 a
SS1.001	69.8 a	59.4 a	3.7 a	22 a	1.7 b	6.5 a
JG18.001	30.4 c	6.6 c	0 c	0.3 b	0 d	0 b
JG9.001	19.8 d	3.6 cd	0 c	0 b	0 d	0 b
M30.001	0 e	0 d	0 c	0 b	0 d	0 b
Water	0 e	0 d	0 c	0 b	0 d	0 b

Incidence, number of infected plants out of 11 seedlings in the lab or 12 plants in the greenhouse, or number of infected tillers out of 20 tillers in the field (pooled for intact and de-head treatments).

π Least-square means for area under the disease progress curve (AUDPC) for the greenhouse and field pathogenicity assays of five strains of *Colletotrichum* on Sugar Drip. For the AUDPC means followed by the same letter, for each column, are not significantly different according to Least Square Means (t-test significant at P = 0.001)

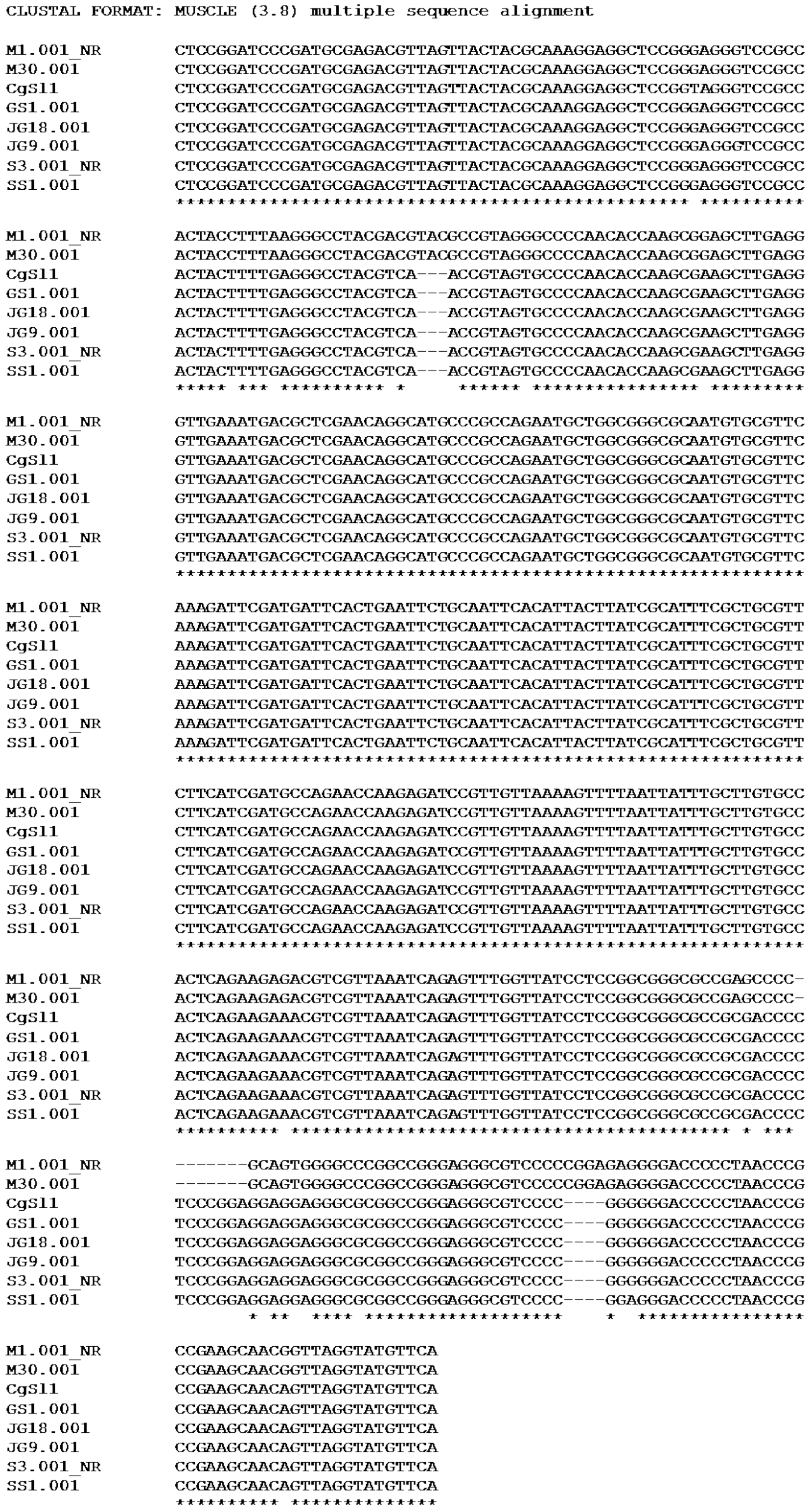
For incidence means followed by the same letter, for each column, are not significantly different according Tukey’s multiple comparison test procedure

Table 2.6. Effect of disease treatments on plant yield (fresh biomass, grain, and juice content) and BRIX.

Treatments	2014				2015			
	Weight[∫]	Juice[∫]	Brix [¥∂]	Grain*[√]	Weight[∫]	Juice[∫]	Brix [¥∂]	Grain*[√]
GS1001	4.9 ba [ß]	2.3 a	13.3 a	146.4 ba	1.1 ba	0.6 a	15.7 a	58.1 a
JG18001	4.8 ba	2.5 a	13.0 a	113.3 ba	1.2 a	0.7 a	15.8 a	55.9 a
JG9001	4.6 b	2.5 a	13.1 a	174.1 a	1.1 ba	0.6 a	16.2 a	83.5 a
M30001	4.7 b	2.6 a	13.5 a	148.1 ba	0.99 b	0.6 a	15.6 a	109.1 a
SS1001	4.9 ba	2.5 a	13.4 a	79.2 b	1.1 ba	0.6 a	16.1 a	52.7 a
Water	5.4 a	2.4 a	12.1 b	134.8 ba	1.2 a	0.7 a	15.6 a	62.0 a

¥ One (1) degree Brix corresponds to 1 Kg of fructose in 100 Kg of solution
∫ Data average weight in Kg of biomass or juice from five inoculated tillers
∂ Data average from five inoculated tillers
* Average grams of seeds from five seed heads from harvested inoculated tillers
ß Means followed by the same letter, for each column, are not significantly different according to Least Square Means (t-test significant at P = 0.001). The exception is grains, in which means followed by the same letter, for each column, are not significantly different according Tukey's multiple comparison test procedure

Figure 2. 1.Alignments of ITS sequences of CgSl1, SS1.001, GS1.001, JG9.001, JG18001 and M30.002 with *Colletotrichum sublineola*, species type S3.001_NR, and *Colletotrichum graminicola*, species type M1.001_NR. Alignments done with MUSCLE version 3.8.



```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment

M1.001_NR       CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
M30.001         CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
CgSl1           CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGTAGGGTCCGCC
GS1.001         CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG18.001        CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG9.001         CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
S3.001_NR       CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS1.001         CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
                ************************************************* **********

M1.001_NR       ACTACCTTTAAGGGCCTACGACGTACGCCGTAGGGCCCCAACACCAAGCGGAGCTTGAGG
M30.001         ACTACCTTTAAGGGCCTACGACGTACGCCGTAGGGCCCCAACACCAAGCGGAGCTTGAGG
CgSl1           ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
GS1.001         ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
JG18.001        ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
JG9.001         ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
S3.001_NR       ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
SS1.001         ACTACTTTTGAGGGCCTACGTCA---ACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGG
                ***** *** ********** *     ****** **************** *********

M1.001_NR       GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
M30.001         GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
CgSl1           GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
GS1.001         GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
JG18.001        GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
JG9.001         GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
S3.001_NR       GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
SS1.001         GTTGAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTC
                ************************************************************

M1.001_NR       AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
M30.001         AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
CgSl1           AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
GS1.001         AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
JG18.001        AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
JG9.001         AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
S3.001_NR       AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
SS1.001         AAAGATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTT
                ************************************************************

M1.001_NR       CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
M30.001         CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
CgSl1           CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
GS1.001         CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
JG18.001        CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
JG9.001         CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
S3.001_NR       CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
SS1.001         CTTCATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCC
                ************************************************************

M1.001_NR       ACTCAGAAGAGACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGAGCCCC-
M30.001         ACTCAGAAGAGACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGAGCCCC-
CgSl1           ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
GS1.001         ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
JG18.001        ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
JG9.001         ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
S3.001_NR       ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
SS1.001         ACTCAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCC
                ********** ****************************************** * ***

M1.001_NR       -------GCAGTGGGGCCCGGCCGGGAGGGCGTCCCCCGGAGAGGGGACCCCCTAACCCG
M30.001         -------GCAGTGGGGCCCGGCCGGGAGGGCGTCCCCCGGAGAGGGGACCCCCTAACCCG
CgSl1           TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGGGGGACCCCCTAACCCG
GS1.001         TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGGGGGACCCCCTAACCCG
JG18.001        TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGGGGGACCCCCTAACCCG
JG9.001         TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGGGGGACCCCCTAACCCG
S3.001_NR       TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGGGGGACCCCCTAACCCG
SS1.001         TCCCGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCC----GGAGGGACCCCCTAACCCG
                       * **  **** *******************    *  ****************

M1.001_NR       CCGAAGCAACGGTTAGGTATGTTCA
M30.001         CCGAAGCAACGGTTAGGTATGTTCA
CgSl1           CCGAAGCAACAGTTAGGTATGTTCA
GS1.001         CCGAAGCAACAGTTAGGTATGTTCA
JG18.001        CCGAAGCAACAGTTAGGTATGTTCA
JG9.001         CCGAAGCAACAGTTAGGTATGTTCA
S3.001_NR       CCGAAGCAACAGTTAGGTATGTTCA
SS1.001         CCGAAGCAACAGTTAGGTATGTTCA
                ********** **************
```

Figure 2. 2. Representative disease symptoms, illustrating the rating scale that was used for this work. (A) no symptoms=0; (B) small necrotic dots on the leaves=1; (C) small red spots on the leaves=2; (D) necrotic lesions without acervuli=3; (E) necrotic lesions with acervuli in the center of the lesions=4, and (F) coalescent necrotic lesions and abundant production of acervuli=5.

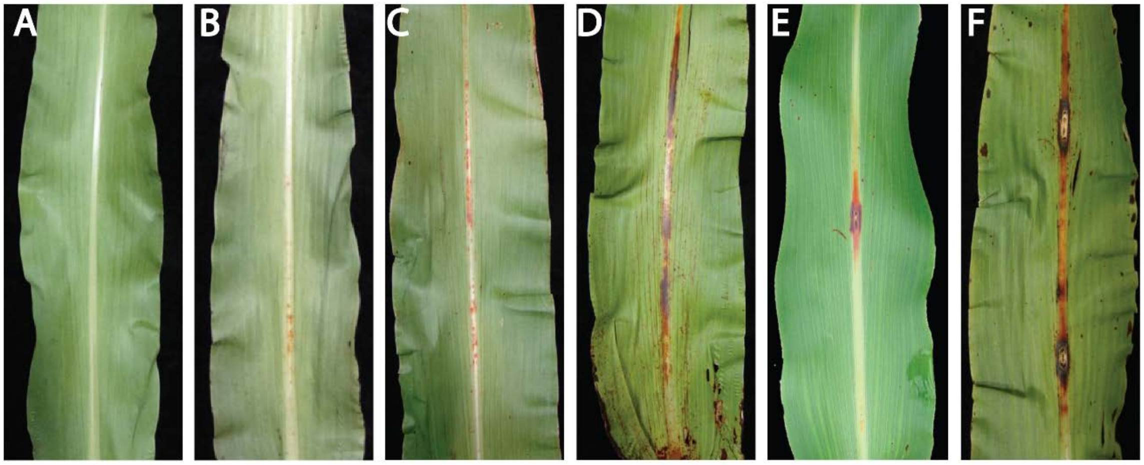


Figure 2. 3. Disease symptoms. (A) field experiment; (B) greenhouse experiment; (C) whole seedling inoculation in the lab; (D) detached leaf sheath assay, inoculation in the lab. Scale bars (C) 0.6 cm and (D) 50 µm.

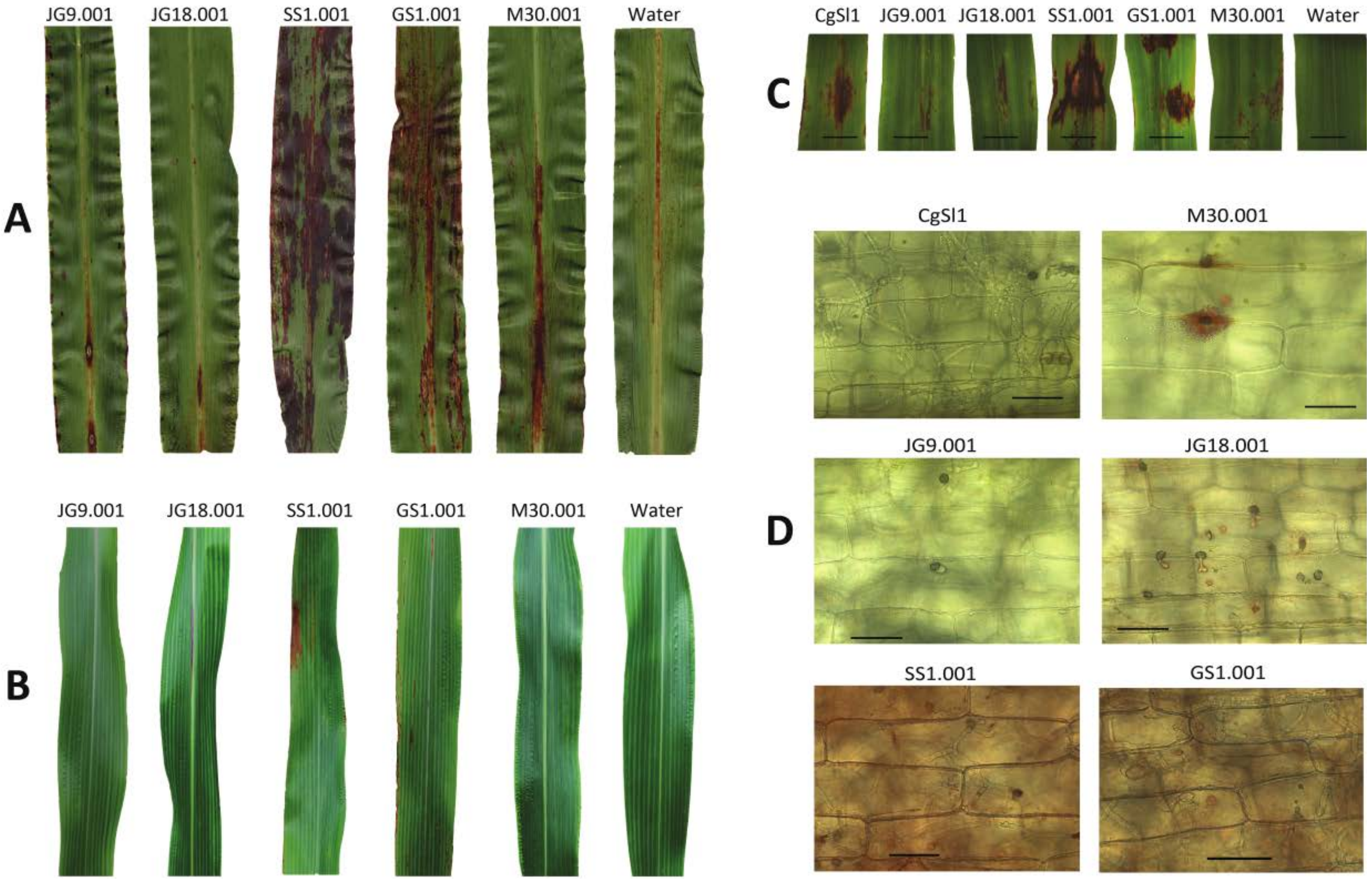


Figure 2.4. Average disease ratings (= the average number of plants rated 0-5) at 6 weeks post-inoculation in 2014 and in 2015.

Figure 2.5. Representative micrographs showing the infection process during the interaction between *C. sublineola* and sweet sorghum. A. Leaf sheath inoculation; (0) ungerminated spore, (1) Appressoria has been formed, (2) Hyphae smaller than appressoria, (3) Hyphae bigger than appressorium but restricted in one cell, (4) Hyphae growing in more than one cell, (5) Hyphae growing in more than three cells, and (6) Acervulus has been formed.

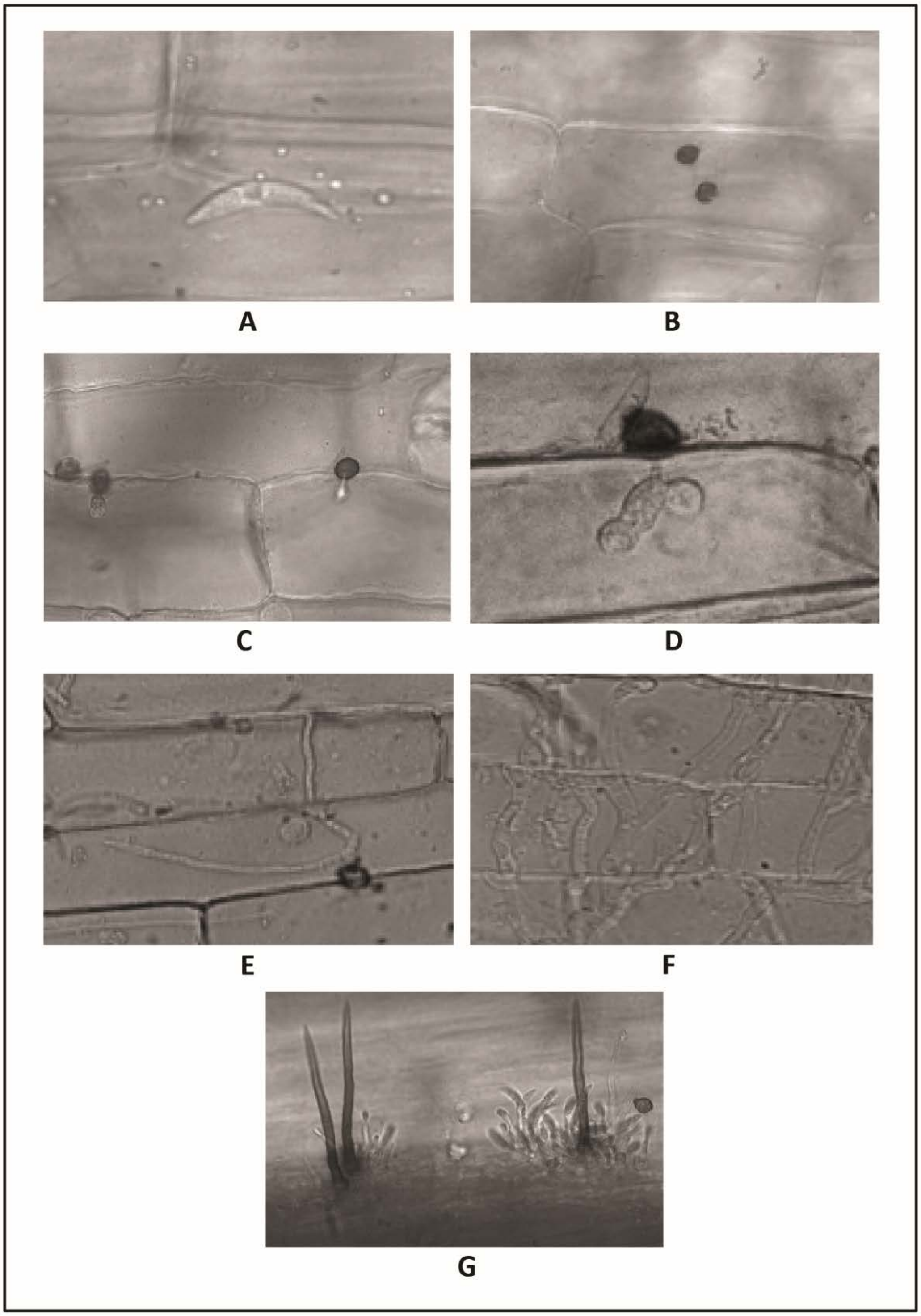


Figure 2.6. Leaf sheath assay of six strains of *Colletotrichum* on Sugar Drip. Disease evaluation was performed based on the number of infection sites in which the pathogen was able to grow beyond one plant cell at 60 hours after inoculation. Combined data from the evaluation of 50 infection sites on 20 leaf sheaths per treatment. The evaluation was performed under light microscope.

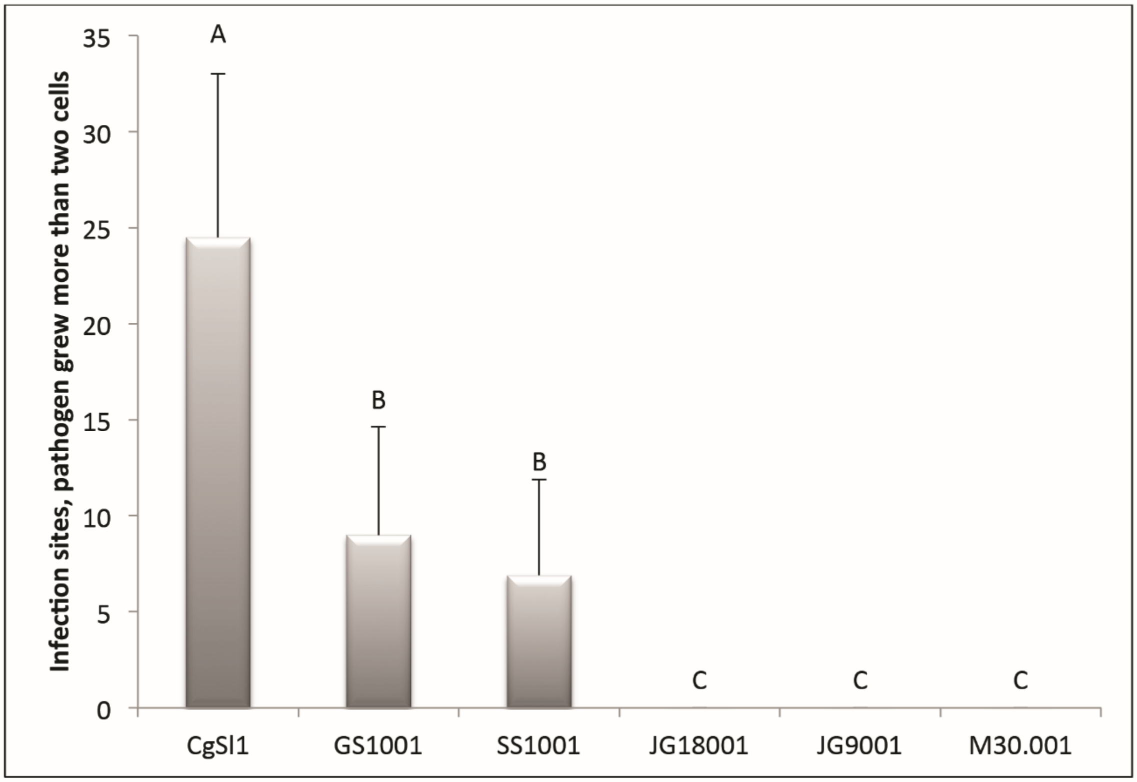
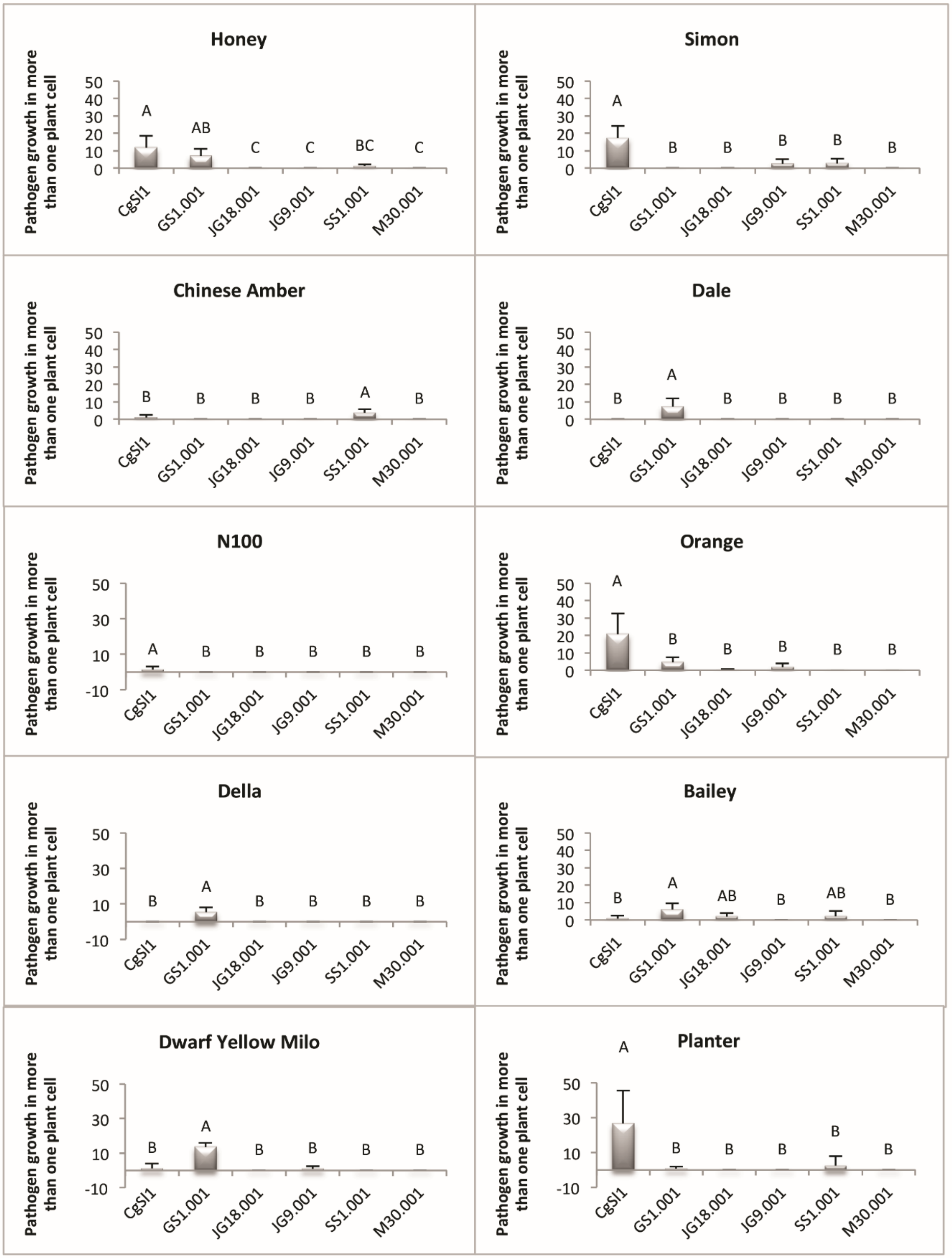


Figure 2.7. Leaf sheath assay comparing 6 fungal strains in each variety of Sorghum. Treatment means were compared among each other using Tukey's test.



Honey

Simon

Chinese Amber

Dale

N100

Orange

Della

Bailey

Dwarf Yellow Milo

Planter

Figure 2.8. Pictures illustrating the plant damage caused by the weather (high temperature and excess of rain) at the beginning of the season in 2015.

Chapter 3 - Genetic and pathogenic diversity of *Colletotrichum* isolates from *Sorghum bicolor* and *S. halepense* in the Southeastern United States

Abstract

Anthracnose caused by *Colletotrichum sublineola* is the most important disease of grain sorghum worldwide, and it is emerging as a significant production constraint for sweet sorghum grown in the southeastern United States. Anthracnose is also common on the wild sorghum relative, johnsongrass (*S. halepense*). Marker analysis using repetitive fingerprinting probes revealed that isolates from across the region were genetically highly diverse, with relatively few haplotypes that could be found in more than one location. With a few exceptions, the populations of strains recovered from cultivated sorghum were genetically distinct from the populations recovered from *S. halepense*. Nearly all of the isolates from cultivated sorghum were pathogenic to one or more of 13 genetically diverse lines of sorghum that were tested. A few of the isolates from *S. halepense* were also pathogenic to some of the lines of *S. bicolor*, while others caused no disease on any of the lines. The evidence suggested that cross-infection of sweet sorghum and johnsongrass by the same strains does occur, although relatively rarely, and this could be expected to complicate efforts to develop and deploy resistant sweet sorghum varieties in areas where johnsongrass is common.

Introduction

Anthracnose, caused by the fungus *Colletotrichum sublineola* Henn. ex Sacc. & Trotter (146), is the most important disease of sorghum (*Sorghum bicolor* (L.) Moench) worldwide (3, 29, 71, 92, 193). Anthracnose can reduce yields of grain sorghum by 50% or more under warm and humid conditions in a susceptible variety (3, 83, 171). *C. sublineola* can infect all parts of the plant including the stem, peduncle, panicle, grain, and leaf blades and midrib (68, 171). The primary control for anthracnose in grain sorghum is the use of resistant cultivars, although no cultivar is immune to the disease (28). Resistance frequently fails, and genotypes that provide high levels of resistance in one location often succumb to the disease in another. This is thought to be related to the high degree of genetic variability in the population of *C. sublineola* affecting grain sorghum (25, 27). Genetic diversity within the *C. sublineola* population on grain sorghum has been characterized by the use of various types of molecular fingerprinting markers, including random amplified polymorphic DNA (RAPD), restriction fragment length polymorphisms (RFLP), and amplified fragment length polymorphisms (AFLP) (30, 135, 184).

Resistance to *C. sublineola* in sorghum is controlled mostly by single dominant genes, some with multiple allelic forms (14, 35, 104, 170). The ability of *C. sublineola* to infect a series of different cultivars of sorghum has been used to identify a large number of races in this species (3, 14, 30, 143, 184). The interaction between sorghum and *C. sublineola* was suggested to follow the gene-for-gene pattern (26). The durability of these major-gene sources of genetic resistance has been limited by the genetic and pathogenic diversity of the fungal population (30, 135).

Sweet sorghum has been grown in the southeastern U.S. for more than 150 years, but it has been on a limited scale, primarily for forage and for the production of syrup. During the Second World War, production was increased in the region to provide a substitute for imported cane sugar (87). Anthracnose disease became a serious problem in those plantations, which led to the development of a breeding program to produce resistant varieties (106). Although production of sweet sorghum in the southeast decreased dramatically after the war, it has continued to be an important specialty crop in Kentucky, which is presently one of the top producers of sorghum syrup in the U.S (12). Recently, interest in sweet sorghum as a potential source of biofuel has increased in the U.S., especially in the Gulf States, where infrastructure used for the sugar cane industry could be readily converted to the production of ethanol from sorghum juice and biomass (102). Anthracnose is listed as one of the most important diseases of sweet sorghum in Kentucky (13), and it is also common on the hybrids that are increasingly grown in the Gulf States for feedstock (80). If sweet sorghum acreages continue to increase in the southeastern U.S., anthracnose can be expected to grow in importance as a significant limiting factor.

C. sublineola is also reported to be a pathogen of the weed johnsongrass (*S. halepense*), which is related to *S. bicolor* (194), and which is ubiquitous across the southeastern U.S. Johnsongrass is known to be genetically highly diverse (120). This host diversity could lead to the selection of multiple *C. sublineola* pathotypes in the population, and thus johnsongrass could serve as a refuge and an incubator for genetic diversity in the pathogen. Rosewich found that isolates from johnsongrass belonged to a population that was distinct from the isolates from grain sorghum being grown in nearby fields in Texas, suggesting that cross-infection did not occur in that location (144). However, there is currently no information about the population structure of *C. sublineola* on sweet sorghum and johnsongrass in the southeastern U.S.

My goal in this chapter was to characterize the degree of genetic and pathological diversity among the population of *Colletotrichum* on cultivated and wild sorghum in Kentucky and the southeastern U.S. This information will help to evaluate the potential for *Colletotrichum* to cause epidemics in sweet sorghum if it is more widely grown for biofuels in the region in the future. A second goal was to develop and evaluate different types of molecular markers for the evaluation of population diversity in *C. sublineola*.

Materials and Methods

Fungal strains: A total of 400 fungal isolates were collected and archived in a permanent collection for this study (Supplemental Table 1). The isolates were recovered in 2011-2015 from sweet, forage, or grain sorghum (*S. bicolor*), as well as from johnsongrass (*S. halepense*), in Kentucky (KY), Georgia (GA), Alabama (AL), and Florida (FL). A few grain sorghum isolates from Indiana, Texas, Brazil, Burkina Faso, Zambia, South Africa, and Sudan, which were previously collected in the 1980s up through 2013, were also included in parts of the study (Supplemental Table 1).

Isolates were recovered from the margins of anthracnose lesions on leaves or stalks. Tissue pieces were soaked in 70% ethanol for 30 seconds, then 10% sodium hypochlorite (Clorox®) for two minutes, followed by three washes with sterile water. Disinfected tissue pieces were dried on sterile paper towels and placed on potato dextrose agar (PDA, Difco Laboratories, Detroit, MI, USA) containing ampicillin (0.1 mg/ml). Cultures were incubated at 23°C under continuous fluorescent light. Isolates were purified by single-sporing, and preserved on silica gel at –80°C (179). Fungal strains were routinely cultured on PDA at 23°C under continuous fluorescent light.

Genomic DNA extraction: Fungal spores were harvested from 14-day-old PDA cultures and washed three times with sterile water. A volume of 10 ml of potato dextrose broth (PDB; Difco Laboratories, Detroit, MI, USA) was added to a 60 mm Petri plate, and inoculated with a spore suspension to the final concentration of $5x10^4$ spores/ml. Plates were incubated at 23°C under continuous fluorescent light for 7 days, after which mycelial mats were collected, frozen, and lyophilized. Genomic DNA was extracted following the protocol of (174).

Random Amplified Polymorphic DNA (RAPD): Genomic DNA of *Colletotrichum* strains was amplified by using the Random Amplified Polymorphic DNA (RAPD)-Polymerase Chain Reaction (PCR) method (201) with 10 different 10-base random primers (19, 181) (Table 3.1). Reactions were carried out in a final volume of 25µL containing: 10 ng of genomic DNA, 10X PCR buffer, 50 mM $MgCl_2$, 200 µM dNTPs, 5 µM of primer, and one unit (U) of *Taq* DNA polymerase. The PCR cycling protocol consisted of 3 cycles of; 1 min 94ºC denaturation; 1 min annealing at 35ºC; 2 min for extension at 72ºC: followed by 34 cycles of; 1 min denaturation at 94ºC; 20 sec annealing at 40ºC; and 2 min extension at 72ºC: and with a final extension of 72ºC for 10 min. Amplicons were resolved on 1.0% agarose gels, stained with ethidium bromide, and visualized under UV light.

RFLP fingerprints with transposon probes: Proteins containing predicted transposase domains were identified by using the Protein Family Database (Pfam) to screen the predicted proteome of *C. sublineola* strain CgSl1 (see chapter 4 of this dissertation). A putative transposase sequence was identified that was very similar to JMSE01000036, a predicted transposase sequence from the *C. sublineola* isolate CsTX430BB (Figure 1). The reverse complement of the CgSl1 sequence was used to design the transposase primers CsTn1F and CsTn1R (Table 1, figure 1). Genomic DNA from strain CgSl1 was used as a template for the PCR. PCR conditions consisted of 3 min of denaturation at 94°C; followed by 35 cycles of denaturation for 30 sec at 94°C, 60 sec of annealing at 57°C, and 40 seconds of extension at 72°C; and a final extension of 10 min at 72°C. The resulting amplicon was sequenced (Figure 1) and submitted to Genbank as CsTn1 (accession number KY006076). The amplicon was labeled with ^{32}P to produce a probe for Southern hybridization. The Southern blot was performed by using the protocol of (174). Briefly, one µg of the genomic DNA from each fungal strain was digested with the restriction enzyme *Hind*III. The digested DNA was separated on a 0.7% agarose gel and transferred to a positively charged nylon membrane. The membrane was cross-linked with UV light, and probed with the radioactively labeled probes.

RFLP fingerprints with telomere probes: Telomere probes were produced by using a published protocol (65). The resulting amplicons were gel-purified and labeled with ^{32}P. Conditions for doing the Southern blot were as described above.

Identification and characterization of haplotypes: Strains with 100% (RFLP), or at least 95% (RAPD) identical band patterns in each case were visually identified as members of the same haplotype (Supplemental table 1). Polymorphic RAPD bands of 84 representative isolates were scored as presence (1) or absence (0), then Jaccard's coefficients (91) were applied to the data by using the Statistical Analysis Software (SAS) package, in order to group the isolates based on the overall similarity of the band patterns.

Pathogenicity assay: Isolates representing 34 different haplotypes (based on the fingerprinting analyses) from a range of hosts and locations were tested for pathogenicity against the sweet sorghum variety Sugar Drip in the greenhouse (Table 3.2). The experiment consisted of two sets of experiments. The first set was repeated three times (summer and fall of 2014; and summer 2016) and the second one was repeated twice (July and August, 2016). One isolate from maize was used as a negative control: maize isolates do not infect healthy sorghum plants (175). A subset of these isolates, plus one additional representative, (Table 3.2) was also tested on 12 additional sorghum varieties including: Honey, Simon, Bailey, Chinese Amber, Dale, Dwarf Yellow Milo, Cowper, Orange, Planter, N100, Della, and Keller. Sugar Drip seeds were obtained from Dr. Todd Pfeiffer at the University of Kentucky. Seeds of the other varieties were obtained from Dr. Gary Pederson at the Plant Genetic Resources Conservation Unit in Griffin GA. All seed stocks were increased in 2012 at the UK Agronomy Farm.

The fungal strains were cultured on PDA at 23°C under continuous fluorescent light for two weeks. Production of inoculum and disease assessment methods were done according to a modified method of Prom and colleagues (136). Spores were harvested from two-week-old cultures by adding 10 mL of sterile water to each plate and loosening the spores with a mini-pestle. The spores were collected and washed three times with sterile water to remove the mucilage. Finally, a spore suspension ($5x10^6$ spores per mL) was used to inoculate 40 grams of autoclaved Sugar Drip sorghum grains in a Petri plate at a final concentration of $5x10^6$ spores per gram. Grains were previously prepared for inoculation by washing them, soaking them in water for 48 hours, and then autoclaving twice (121°C/30 minutes). The inoculated plates were incubated for 14 days at 23°C in the dark.

The pathogenicity experiments were performed in the greenhouse in pots containing four plants, which were arranged in a randomized complete block design (RCBD) with four blocks, each experimental unit or treatment including eight plants per block. Inoculations of Sugar Drip were repeated three times; inoculations of the other 12 lines were repeated twice. To inoculate the plants, ten colonized grains were dropped into the leaf whorl.

Disease symptoms were evaluated weekly, beginning 14 days after inoculation (dpi), for a period of four weeks. A rating scale with values ranging from 0-5 was used (22, 25, 184). On this scale, 0 indicates no symptoms; 1 is small necrotic dots on the leaves; 2 is small red spots on the leaves; 3 is necrotic lesions without acervuli; 4 is necrotic lesions with acervuli in the centers of the lesions; and 5 is coalescent necrotic lesions and abundant production of acervuli. Two types of interaction were considered, (i) incompatible, including values of 0, 1, 2, and 3; and (ii) compatible, including values of 4 and 5. The disease evaluation was performed based on incidence, expressed as the number of infected plants (compatible interactions).

Incidence values from the combined data sets from the repetitions of each experiment were used to calculate the area under the disease progress curve (AUDPC) by the trapezoidal integration method (20). The AUDPC data were subjected to analysis of variance (ANOVA) after square root transformation before further analysis, to stabilize the variance. Treatment means were compared with the negative control (M30.001) using the Dunnett's test procedure of the SAS software package (SAS Users Guide, SAS Institute, Cary, NC). In addition, the data for the isolates from *S. bicolor* versus *S. halepense* on Sugar Drip were grouped, and the group means were compared to one another by using the general linear models (GLM) procedure of SAS (SAS Users Guide, SAS Institute, Cary, NC). Means were separated with least squares estimates of marginal means (LSMEANS statement). P-values less than 0.05 were considered significant. The analyses were performed using SAS version 9.4.

Results

Molecular marker analysis

RAPD analyses: Strains that shared the same band patterns for all ten primers were considered to belong to the same haplotype. A total of 286 haplotypes were identified among 371 strains that were evaluated by amplification with the RAPD primers (Supplemental table 1). These included 176 haplotypes among the 221 isolates from sweet sorghum; 80 haplotypes among the 91 isolates from johnsongrass; 11 haplotypes among the 16 grain sorghum isolates; and 22 haplotypes among the 42 forage sorghum isolates. Forty-one of the haplotypes encompassed more than one strain (Supplemental table 1): the rest were represented by only a single strain each. Only three of the haplotypes were found on more than one host type (RAPD haplotypes 14, 15, and 16, recovered from both sweet and forage sorghum in AL). With only one exception (RAPD haplotype 17, found in two counties in AL), each RAPD haplotype was recovered from only a single location.

RFLP analyses, transposon probe: A total of 386 isolates were analyzed with the transposon probe. The probe hybridized to multiple polymorphic restriction fragments in all of these isolates. However, hybridization to the johnsongrass and maize isolates was generally weaker, and fewer fragments were recognized, compared with the isolates from forage, sweet, or grain sorghum. Representative results are shown in Figure 3.2 A. A total of 61 different haplotypes were identified with this probe (Supplemental table 1). These included 32 haplotypes among the 229 isolates from sweet sorghum; 15 haplotypes among the 89 isolates from johnsongrass; 15 haplotypes among the 18 grain sorghum isolates; and 3 haplotypes among the 49 forage sorghum isolates.

<u>RFLP analyses, telomere probe:</u> A total of 386 isolates were analyzed with the telomere probe. The telomere probe identified a total of 49 different haplotypes (Supplemental table 1). These included 5 haplotypes among the 229 isolates from sweet sorghum; 33 haplotypes among the 89 isolates from johnsongrass; 12 haplotypes among the 18 grain sorghum isolates; and 1 haplotype among the 49 forage sorghum isolates.

Most of the johnsongrass isolates had diverse patterns of hybridization, consisting of between 4 and 36 individual bands (Figure 3.2 B, Supplemental table 1). For the grain sorghum isolates, the probe hybridized to between one and 24 fragments. None of the patterns of multiple hybridizing fragments matched those from the johnsongrass isolates. Most of the isolates from forage and sweet sorghum had just one major band of similar size when hybridized with the telomere probe, (Figure 3.2 B, Supplemental table 1). For this reason, telomere fingerprinting was not as useful for differentiating among sweet and forage sorghum strains. However it could be used to distinguish between two groups, one mostly from johnsongrass and grain sorghum, and the other mostly from sweet and forage sorghum. There were exceptions to this rule including seven grain sorghum isolates, and three isolates from johnsongrass (JG35.001, JG86.001 and JG88.001), that had the single band, and seven isolates from sweet sorghum representing three haplotypes that had multiple telomere bands. Although these three sweet sorghum haplotypes were relatively rare, they were also cosmopolitan, with representatives from KY, FL, and GA (Supplemental table 1).

Comparisons among the marker types

All three types of markers were evaluated, and the results were compared, for 360 representative isolates (Table 3.3). The RAPD fingerprints included the largest number of polymorphic bands, and as a result differentiated the largest number of haplotypes overall (Table 3.3). Most of these haplotypes were represented by only a single strain. Among the 37 RAPD haplotypes represented by multiple strains (a total of 114 strains that could be evaluated), most were consistent with the telomere and transposon haplotypes of the strains. Only 15 strains, in eight different RAPD haplotypes, were not consistent (Supplemental table 1).

Among the 48 telomere haplotypes, only two included more than one transposon haplotype. One of these was type 33, which had only a single major hybridizing band and included the majority of sweet sorghum and forage sorghum isolates. Among the 59 transposon haplotypes, only seven included more than one telomere haplotype, and only four strains within these seven had telomere haplotypes that were also found in combination with other transposon haplotypes. Thus, the vast majority of strains were consistent for the transposon and telomere haplotypes. Each of the 360 isolates was classified into a group based on the combined transposon-telomere haplotype (Supplemental table 1). The total number of combined haplotypes was 83.

Four cosmopolitan combined haplotypes were found in more than one state. The most common of these was 17:33 (Figure 3.3, Supplemental table 1), which included 80 isolates from sweet sorghum and 13 from forage sorghum (Supplemental table 1), and was found in all four states (KY, AL, GA, and FL) (Figure 3.3, Supplemental table 1). Haplotype 17:33 also included two isolates from KY johnsongrass, and two isolates from grain sorghum, one from Texas and one from South Africa (Figure 3.3, Supplemental table 1).

The two johnsongrass isolates represented the only cases in which the same combined haplotype was found on both *S. bicolor* and *S. halepense*. Two other haplotypes (19:33 and 21:33) were found on both forage and sweet sorghum in GA, FL, and AL. The remaining 80 haplotypes were confined to only a single host type, and 76 were found in only a single location.

Among the johnsongrass isolates, 35 different combined haplotypes could be discerned (Table 3.4, Supplemental table 1). There was a relatively high degree of diversity (richness) in most locations, with even isolates from same plant sometimes belonging to several different haplotypes. Nearly all of the haplotypes were found in only a single location (Figure 3.3). However, there were four combined haplotypes that were more widely dispersed in Kentucky, including haplotypes 42:60 and 44:62, found in both Fayette and Franklin counties (Figure 3.3, Supplemental table 1). None of these more widely dispersed haplotypes was found outside KY.

A cluster analysis of the RAPD fingerprint markers for 84 strains representing the 80 different combined haplotypes revealed that the clusters were mostly consistent with host (*S. bicolor,* highlighted in red in Figure 3.4, versus *S. halepense,* highlighted in gray) and with geographic origin. There were a few exceptions. Thus, strain SS1.001 from Kentucky sweet sorghum clustered with isolates from *S. halepense*, while two KY johnsongrass isolates (JG14.001 and JG16.001) were grouped together with isolates from *S. bicolor*. Several Brazilian grain sorghum isolates clustered together with a Sudanese grain sorghum isolate and several sweet sorghum isolates from Florida (Figure 3.4 G). Two other grain sorghum isolates (from Burkina Faso and Indiana) were grouped with the two atypical KY johnsongrass isolates (Figure 3.4 E). Forage sorghum isolates formed a discreet cluster within the larger cluster of isolates from *S. bicolor* (Figure 3.4 F). A few isolates, e.g. JG1.001 from johnsongrass in Indiana and GS1.001 from grain sorghum in Kentucky, seemed to be relatively unique and didn't cluster closely with any other strains.

Pathogenicity assays

Thirty four representative isolates from johnsongrass, sweet, grain, and forage sorghum were chosen for pathogenicity tests against the susceptible sweet sorghum inbred Sugar Drip (Table 3.2). All the isolates were compared with the negative control, the non-pathogenic maize isolate M30.001. Most of the isolates were able to cause disease (Figures 3.5-3.7). Six JG isolates (JG2.001, JG9.001, JG18.001, JG62.001, JG75.001 and JG102.001), one forage sorghum isolate (FS34.001), one grain sorghum isolate (TX430BB), and two sweet sorghum isolates (SS41.001 and SS100.001), did not differ statistically from the negative control (M30.001) (Figure 3.5-3.7). The other isolates caused moderate to high levels of disease on Sugar Drip in the greenhouse (Figures 3.5-3.7). Most of these isolates were from cultivated sorghum; however there were five johnsongrass isolates (JG7.001, JG34.001, JG42.001, JG86001 and JG88.001), all from Kentucky, that were also pathogenic to Sugar Drip (3.5-3.7).

A statistical analysis comparing aggressiveness of isolates from different hosts revealed that the johnsongrass isolates, as a group, were statistically less aggressive to Sugar Drip than isolates from sweet and forage sorghum (p=0.05).

One additional isolate from sweet sorghum (Table 3.2) and a representative subset of the original 34 isolates were tested on 12 additional sorghum lines in greenhouse assays. Results revealed that most of the johnsongrass isolates could not cause disease on any of the sorghum lines (Figure 3.8-3.11). The exception was JG88.001, which was pathogenic on Honey, Simon, Bailey, Chinese Amber, and Dale. The ability of the pathogenic isolates to cause disease depended on the sorghum line (Figure 3.12). Honey and Simon were susceptible to most of the isolates from *S. bicolor* that were tested, as well as to JG88.001. On the other hand, Della, N100, and Keller were resistant to all of the isolates. Similarly some isolates from *S. bicolor*, e.g. SS63.001, were pathogenic to most of the sorghum lines tested, while others were pathogenic to only a few (e.g. SS240.001). Based on the results of the inoculations performed on the 12 varieties listed above, and on 11 additional varieties inoculated only with SS1.001 and CgSl1 that I have not shown here, I have added colors to the pedigree map to identify the varieties that were susceptible to at least one isolate of *Colletotrichum* versus those that were resistant to all the isolates tested (Figure 3.12).

Discussion

This work revealed that the population of *Colletotrichum* causing lesions on *Sorghum* spp. in the southeastern U.S. is genetically diverse, and comprised mostly of local, genetically distinct sub-populations. Previous studies have suggested that clonal reproduction is of primary importance in population structure of *C. sublineola* on grain sorghum (72, 144). However, results of RAPD analysis suggested that most strains in the current study were not clonal, even when they came from the same fields or the same plants. Although RAPD markers have been widely used to evaluate *C. sublineola* populations (24, 26, 69, 72, 78, 172, 181, 184), they have become less popular as they have developed a reputation for poor reproducibility (72, 144), and as other PCR-based methods (e.g. AFLP) have become available. I found that RAPD groupings based on ten different random primers correlated very well with groups identified by RFLP markers, but they detected about three times as many haplotypes, probably because they identified more polymorphic bands (Table 3.2).

The RAPDs were convenient because they required relatively little DNA and were PCR-based rather than hybridization based. However they were challenging to develop and validate because they had to be carefully screened for reproducibility. I found that only the major bands were reliable and should be included in the analysis. By their nature, PCR-based methods like RAPD and SSR are less stringent than RFLP because band size is the only criterion used for marker identification. It is possible that some bands of similar size actually represent different markers, thus genetic diversity may be underestimated. Use of multiple RAPD primers can alleviate this concern when identifying haplotypes, and given the high degree of reproducibility and correlation with other markers in this study, I conclude that the RAPD markers added useful information because of their greater power to differentiate among individual clones, and the fact that they were equally informative for all strains.

RFLP uses a combination of band size and sequence homology to identify markers, thus providing more reproducibility and reliability than PCR-based methods. I developed and used two repetitive primers, one based on transposon sequences and the other on subtelomeric sequences that often vary among fungal strains e.g. (2, 66, 109, 147). Comparing the RFLP marker types I saw that the telomere markers overall allowed identification of 49 haplotypes, as compared with 61 for the transposon marker. However, the telomere markers were less informative for a relatively large group of isolates, nearly all from sweet and forage sorghum, which had only a single major telomere band of a similar size. Use of several different restriction enzymes in the Southern blots did not change the result. Presumably this group (group 33) has very little diversity of different telomere ends within their genomes. Within the 275 members of group 33, 34 haplotypes could be identified by the addition of the transposon probe. The transposon probe on the other hand was less useful for isolates from johnsongrass, which generally hybridized poorly to this probe, suggesting that they may contain a different related group of transposons. Both probes together provided the most reliable differentiation of all of the strains, and yielded a total of 80 combined haplotypes across the southeast. Most of these were confined to a single location, confirming the RAPD results suggesting high levels of local isolation. A few were more cosmopolitan, with the most common combined haplotypes being 17:33, mostly from sweet sorghum across the southeast, 19:33, from sweet and forage sorghum in AL, FL, and GA, and 21:33, mostly from forage sorghum in AL, but also including a few representatives from GA on sweet sorghum.

RAPD markers provided similar amounts of information for all isolates (in contrast with the telomere and transposon RFLPs, which were more informative for one group versus another). Thus, I used the RAPD markers for an analysis of similarity among strains that were selected to represent all of the combined haplotypes. Results of the cluster analysis suggested that populations from johnsongrass were mostly distinct from those on cultivated sorghum, with a few exceptions. Cluster analysis also indicated that isolates from sweet, forage, and grain sorghum were not separate, suggesting that cross-infection within the cultivated sorghum group is common. This conclusion was also supported by pathogenicity assays, in which most isolates from grain, forage, and sweet sorghum were able to cause disease in Sugar Drip. One grain sorghum isolate, TX430BB was not pathogenic. This isolate was recently subjected to genome sequencing (8). When I tested this isolate on the susceptible line Della, it was able to cause lesions (not shown). Results of this and my other inoculations of multiple strains with multiple isolates provided evidence for significant race structure in the population on sweet sorghum in the SE, as is well known for this pathogen on grain sorghum (25, 27). The presence of races is expected to complicate the task of plant breeders as they develop and deploy new sweet sorghum germplasm in the region in the future.

Considering the two major host families, johnsongrass (*S. halepense*), and cultivated sorghum (*S. bicolor*), the haplotypes recovered from one did not usually appear among the isolates from the other, even when the isolations were made from the same locations. Two johnsongrass isolates (JG86.001 and JG88.001) were rare exceptions to this rule. Both belonged to haplotype 17:33, which was common on sweet and forage sorghum across the region, and was also represented by two grain-sorghum isolates. The two johnsongrass isolates belonging to haplotype 17:33 came from different plants that were in the rows between infected sweet sorghum in Morgan County KY. More than 50 isolates from two varieties of sweet sorghum in those fields also belonged to combined haplotype 17:33, showing that this population was capable of cross-infection of the isolated johnsongrass plants as well as the sweet sorghum. Although they shared the same combined haplotype, these isolates from johnsongrass represented unique RAPD haplotypes and this suggests they were not clonal with one another or with other strains from sweet sorghum plants nearby. Results from pathogenicity assays also suggested the possibility for cross-infection by some, but not all isolates. Four isolates from Kentucky, including JG88.001, showed significant potential for infection of Sugar Drip sorghum, although the majority of isolates from johnsongrass did not differ from the negative controls on Sugar Drip or on 12 other genetically diverse sorghum genotypes. Although cross-infection events seem to be relatively rare, the potential exists nonetheless for johnsongrass to serve as a refuge and a source of inoculum for the pathogen in sweet sorghum fields under rotation conditions.

Some of the same strains that were tested in this chapter on whole plants were also tested in sheath assays on those same plant varieties in chapter 2. In general, results of the two assays were consistent for most of the strains. The notable exception was strain SS1.001, which was pathogenic on Honey, Simon, Bailey, and Planter in whole plant assays but didn't differ from the negative control in sheath assays involving those plant varieties. One possibility is that SS1.001, although overall a highly aggressive strain, may be a "slow starter". If I had waited longer than 60 hours to evaluate the sheath assay, it may have been able to colonize the sheaths effectively. Alternatively, it may react differently compared with the other strains to the different nutritional environments of the sheaths versus the leaf blades (sheaths are sink tissues while leaf blades are source tissues). Whatever the explanation, this demonstrates that the sheath assay should not be relied upon to give accurate predictions of field pathogenicity for all strains on all varieties.

Overall it appeared that there was a higher level of haplotype diversity (richness) among the isolates from johnsongrass than among the isolates from *S. bicolor*. This seems reasonable given the higher level of genetic diversity among the wild johnsongrass versus the domesticated sorghum monocultures. The theory is that the former would promote selection of a variety of genotypes within the pathogen population, while the latter would promote selection of a few individual genotypes. The exception seemed to be the isolates from grain sorghum, which displayed a very high level of richness. Other studies of diversity within grain sorghum populations have also suggested high levels of diversity (30, 135, 184). However, my results may have been an artifact of the fact that these isolates have been collected across many decades and from different continents. The randomness of sampling and the sample size have a big impact on these measures of diversity. Now that I have developed and characterized molecular fingerprinting tools, these can be applied to additional isolates sampled at various levels in the future. Overall this work has revealed that the pathogen in the region has significant genetic potential to overcome individual resistance genes and may pose a serious threat to widespread sorghum monoculture in the future.

Summary

The predictions for chapter 3 were: (1) isolates from sweet sorghum will resemble those on grain and forage sorghum more than they will resemble isolates from johnsongrass; (2) the population of strains recovered from johnsongrass will be more diverse genetically than the population recovered from cultivated sorghum; and (3) isolates from different hosts will vary in their virulence and aggressiveness to sweet sorghum varieties. Predictions 1 and 2 were upheld, although sampling is likely to be a significant complicating factor in my estimates of diversity. It will be important in future to increase the sampling depth in individual locations to properly address this question. Prediction 3 was not upheld. There appeared to be no difference in the ability of isolates from grain, forage, or sweet sorghum to infect Sugar Drip sweet sorghum in the greenhouse. And, although a majority of isolates from johnsongrass were not pathogenic to sweet sorghum, this was not true for all isolates from johnsongrass.

Table 3.1. RAPD and transposon primers used for fingerprint analyses.

Primer	Sequences of Primers	Reference
OPA 1	5’CAGGCCCTTC3’	Vaillancourt & Hanau, 1992
OPA 2	5’TGCCGAGCTG3’	Vaillancourt & Hanau, 1992
OPA 3	5’AGTCAGCCAC3’	Vaillancourt & Hanau, 1992
UBC 356	5’GCGGCCCTCT3’	Browning et al, 1999
UBC 384	5’TGCGCCGCTA3’	Browning et al, 1999
UBC 391	5’GCGAACCTCG3’	Browning et al, 1999
UBC 357	5’AGGCCAAATG3’	Browning et al, 1999
UBC 345	5’GCGTGACCCG3’	Browning et al, 1999
OPA 11	5'CAATCGCCGT3’	This Publication
OPA 18	5'AGGTGACCGT3’	This Publication
CsTn1F	5’AATCCAGGCTGATCTTGGGC3’	This Publication
CsTn1R	5’CCTGGCGAGCCTTCTGATAG3’	This Publication

Table 3.2. The strains of *Colletotrichum* chosen for the pathogenicity assays.

Strain	Host	State	County	Year	Haplotype		
					Transposon	Telomere	RAPD
CgSl1	Grain sorgum	Indiana	Tippecanoe	1975	98	110	300
CsTX430BB	Grain sorgum	Texas	Brazos	1975	17	33	33
FS13.001	Forage sorghum	Alabama	Brewton	2013	17	33	13
FS34.001	Forage sorghum	Alabama	Macon	2013	19	33	12
FS37.001	Forage sorghum	Alabama	Macon	2013	19	33	18
FS41.001	Forage sorghum	Alabama	Cullman	2013	17	33	17
FS5.001	Forage sorghum	Alabama	Brewton	2013	21	33	9
FS8.001	Forage sorghum	Alabama	Brewton	2013	21	33	8
GS1001	Grain sorghum	Kentucky	Hopkins	2013	99	108	298
JG102.001	johnsongrass	Georgia	Tift County	2014	50	84	54
JG18.001	johnsongrass	Kentucky	Fayette	2012	43	61	49
JG2.001	johnsongrass	Kentucky	Russell	2011	38	53	42
JG9.001	johnsongrass	Kentucky	Russell	2012	38	57	41
JG62.001	johnsongrass	Kentucky	Russell	2013	38	73	71
JG34.001	johnsongrass	Kentucky	Frankfort	2013	42	60	109
JG42.001	johnsongrass	Kentucky	Fayette	2013	42	60	35
JG7.001	johnsongrass	Kentucky	Russell	2011	39	56	47
JG75.001	johnsongrass	Florida	Gadsden	2014	47	78	61
JG86.001	johnsongrass	Kentucky	Montgomery	2014	17	33	101
JG88.001	johnsongrass	Kentucky	Montgomery	2014	17	33	102
M30.001	Maize	Kentucky	Fayette	2012	101	112	32
Sweet sorghum		Kentucky	Fayette	2012	5	34	301

Continue Table 3.2.

Strain	Host	State	County	Year	Haplotype		
					Transposon	Telomere	RAPD
SS100.001	Sweet sorghum	Alabama	Macon	2013	17	33	14
SS107.001	Sweet sorghum	Georgia	Tift County	2013	24	33	296
SS124.001	Sweet sorghum	Georgia	Tift County	2013	29	33	208
SS159.001	Sweet sorghum	Georgia	Tift County	2013	19	33	213
SS204.001	johnsongrass	Kentucky	Montgomery	2014	17	33	274
SS240.001	johnsongrass	Georgia	Tift County	2014	19	33	225
SS26.001	Sweet sorghum	Florida	Live Oak	2013	1	33	159
SS29.001	Sweet sorghum	Florida	Live Oak	2013	1	33	166
SS41.001	Sweet sorghum	Florida	Live Oak	2013	1	33	150
SS45.001	Sweet sorghum	Florida	Live Oak	2013	1	33	148
SS63.001	Sweet sorghum	Florida	Live Oak	2013	4	33	178
SS78.001	Sweet sorghum	Alabama	Brewton	2013	17	33	16

Table 3.3. Number of haplotypes identified among 360 strains isolated from johnsongrass, sweet, forage, and grain sorghum by the number of haplotypes identified upon hybridization with the telomere and transposon probes. *The forage sorghum strains all had the same, single band that hybridized with the telomere probe, thus this probe did not reveal any polymorphisms among this group of strains.

Host	Number of individuals	Number of locations	Probes	Polymorphic Bands	Individual Haplotypes
Johnsongrass	86	7	RAPD	1388	79
			Telomere	729	33
			Transposon	228	15
Sweet sorghum	225	6	RAPD	4035	173
			Telomere	44	5
			Transposon	593	31
Forage sorghum	42	3	RAPD	391	22
			Telomere	0*	1
			Transposon	66	3
Grain sorghum	13	7	RAPD	391	11
			Telomere	99	11
			Transposon	266	14

Table 3.4. Combined haplotype richness by host identified from isolates collected in the southeastern US.

Host	Total number of individuals	Location (number of individuals)	Haplotypes*	Total number of Haplotypes* Haplotype Richness **
Johnsongrass	89	Fayette KY (18)	7	35 0.39
		Franklin KY (13)	6	
		Russell KY (35)	14	
		Montgomery KY (2)	1	
		Macon AL (7)	1	
		Tift GA (6)	2	
		Gadson FL (7)	6	
		Tippecanoe IN (1)	1	
Sweet sorghum	229	Brewton AL (24)	1	35 0.15
		Macon AL (8)	1	
		Tift GA (67)	17	
		Live Oak FL (77)	18	
		Gainesville FL (1)	1	
		Montgomery KY (51)	1	
		Fayette KY (1)	1	
Forage sorghum	49	Brewton AL (29)	3	3 0.06
		Cullman AL (8)	2	
		Macon AL (12)	2	
Grain sorghum	17	North America (3)	2	16 0.89
		South America (9)	8	
		Africa (5)	5	

* Total number of haplotypes based on the combined haplotypes identified upon hybridization with telomere and transposon probes.

**Haplotype richness is the number of haplotypes divided by the number of individuals.

Figure 3.1. (A) Transposon probe (CsTn1, accession number KY006076) sequence including the forward and reverse primers highlighted in yellow. (B) Similarity of the putative transposase from TX430BB toCsTn1 peptide sequence.

```
>Cstn1
AATCCAGGCTGATCTTGGGCTCCCTCCAACCCATCAGCAAGTGAGGGAGTTCGCCCAGCGTAT
CCTCTGTATAATGGGTGATCATAAGCCTCTAGGCAAGCGATGGATGGATGGCTTCCTGCGAAG
ATACCCCTCTATCAAGGTCCAGAGAAGTCGTTCTATTGATTCCGGGCGTATTAATGGAGCATC
TACTGAGGTTATTAGGGAATGGTTTAAATACCTCGTGATGCCTCACATTGTTGCCATTAAACC
AGCCAACCGATACAACATGGATGAGACTGGTATCCTTGAGGGCAAGGGAGATAATGGGCTAG
TGCTGGGCAGAGCTGAGACTAAGTCAGTGAGGAAAAAGCAGCCTGGATCACGAGCCTGGGTA
TCTATCATCGAGTGTATCTCAGCTGAGGGCAAGGCACTTCATCCACTGGTTATATATAAGGGC
AAGACGGTCCAGCAGCAGTGGTTTCCTCTAGATCTTGGCCCTCATGATGGATGGGAGTTTACA
GCAACAGAGAATGGATGGACTACTGATGATACTGCTGTTGAGTGGTTGACGAGAGTGTTTCTC
CCTCAGTCTCAGCCTTGCCAGCCTTGCCAGCCTTGCCAGCCTGATCAGGCAAGACTACTTATCC
TGGATGGGCATGGGAGCCATACAACAACTGATTTTATGTGGTTATGCTATACCCATAACGTTT
ATCTGTTATTCTTGCCTCCACATACCTCCCATGTCCTCCAGCCACTAGATCAGTCAGTCTTTAG
CCCTGTGAAGTCTGCTTATCGCAAGGAGCTTGGATATCTCAGCCTTTGGAATGACTCTACTATT
ATCGGCAAAAGGAACTTTATTTCCTGCTATCAGAAGGCTCGCCAG
```

```
Putative transposase [Colletotrichum sublineola]
Sequence ID: KDN72148.1Length: 538Number of Matches: 1
Related Information
Range 1: 70 to 351GenPeptGraphicsNext MatchPrevious Match

Score            Expect  Method          Identities    Positives      Gaps         Frame

540 bits(1390)   0.0     Compositional   266/288(92%)  271/288(94%)   6/288(2%)    +2
                         matrix adjust.

Query  2    IQADLGLPPTHQQVREFAQRILCIMGDHKPLGKRWMDGFLRRYPSIKVQRSRSIDSGRIN  181
            IQADLGLPPTHQQVREFAQRIL  MGDHKPLGKRWMDGFLRR PSIKVQRSRSIDS RIN
Sbjct  70   IQADLGLPPTHQQVREFAQRILRAMGDHKPLGKRWMDGFLRRNPSIKVQRSRSIDSRRIN  129

Query  182  GASTEVIREWFKYLVMPHIVAIKPANRYNMDETGILEGKGDNGLVLGRAETKSVRKKQPG  361
            GASTEVIREWFKYLVMPHI AIKPANRYNMDETGILEGKGDNGLVLGRAETKSVRKKQPG
Sbjct  130  GASTEVIREWFKYLVMPHIAAIKPANRYNMDETGILEGKGDNGLVLGRAETKSVRKKQPG  189

Query  362  SRAWVSIIECISAEGKALHPLVIYKGKTVQQQWFPLDLGPHDGWEFTATENGWTTDDTAV  541
            SRAWVSIIECISAEGKALHPLVIYKGKTVQQQWFPLDLGP+DGWEFTATENGWTTDDTAV
Sbjct  190  SRAWVSIIECISAEGKALHPLVIYKGKTVQQQWFPLDLGPYDGWEFTATENGWTTDDTAV  249

Query  542  EWLTRVFLpqsqpcqpcqpcqpDQARLLILDGHGSHTTTDFMWLCYTHNVYLLFLPPHTS  721
            EWLTRVFL         QPCQPDQARLLI+DGH SHTTTDFMWLCYT+NV+LLFLPPHTS
Sbjct  250  EWLTRVFL------PQSQPCQPDQARLLIVDGHRSHTTTDFMWLCYTYNVHLLFLPPHTS  303

Query  722  HVLQPLDQSVFSPVKSAYRKELGYLSLWNDSTIIGKRNFISCYQKARQ  865
            +VLQPLDQ VFSPVKSAYRKELGYLSLWNDSTIIGKRN ISCYQKARQ
Sbjct  304  YVLQPLDQLVFSPVKSAYRKELGYLSLWNDSTIIGKRNSISCYQKARQ  351
```

Figure 3.2. Representative Southern blot analysis of genomic DNA digested with HindIII. Colletotrichum sp. strain name is indicated on each lane. Blot was hybridized with (A) transposon and (B) telomere probes. Strains marked with the same symbol belong to the same haplotype. The ones without symbols belong to different haplotypes.

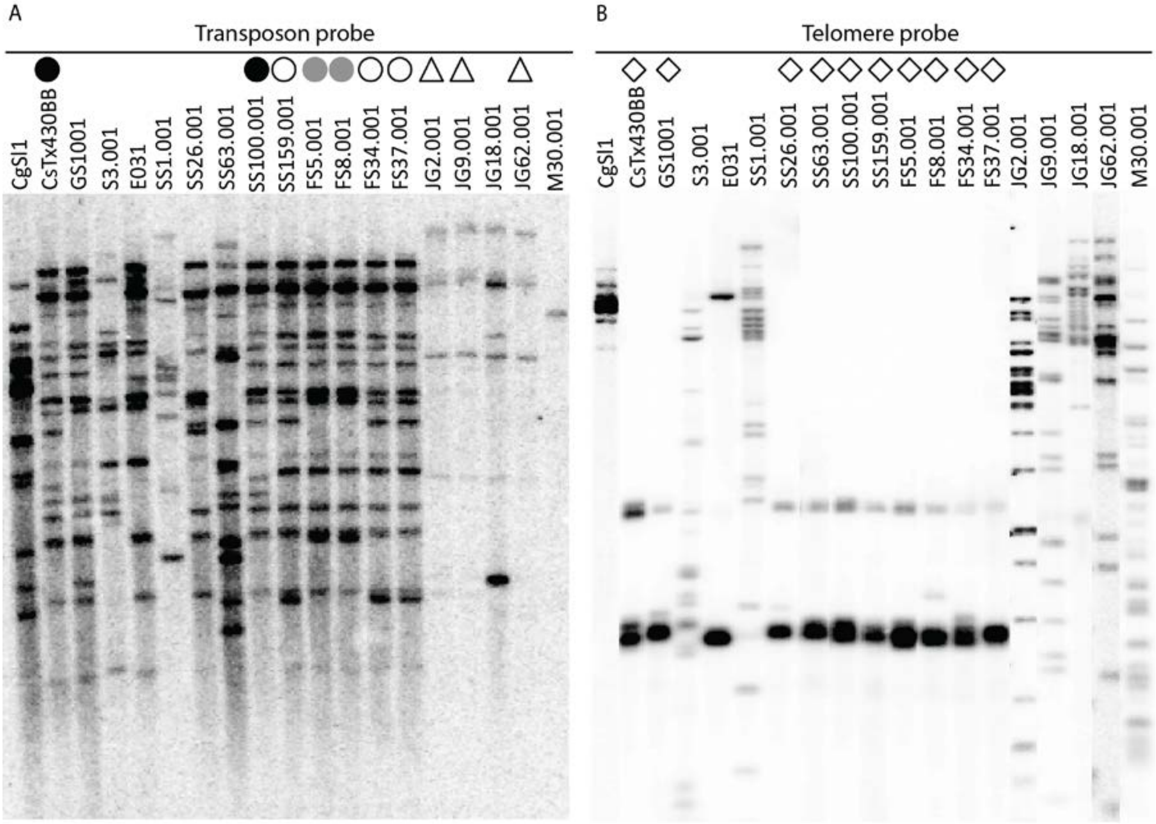


Figure 3.3. Distribution of combined haplotypes. Slices with the same colors represent the same haplotype found in different locations and/or on different host types. Colors match the highlighting in supplemental table 1. Slices in gray tones represent different haplotypes that were found in only that location. The total number of isolates from each location is presented in the center of each pie graph. Empty circles represent locations from which no isolates of that category were obtained.

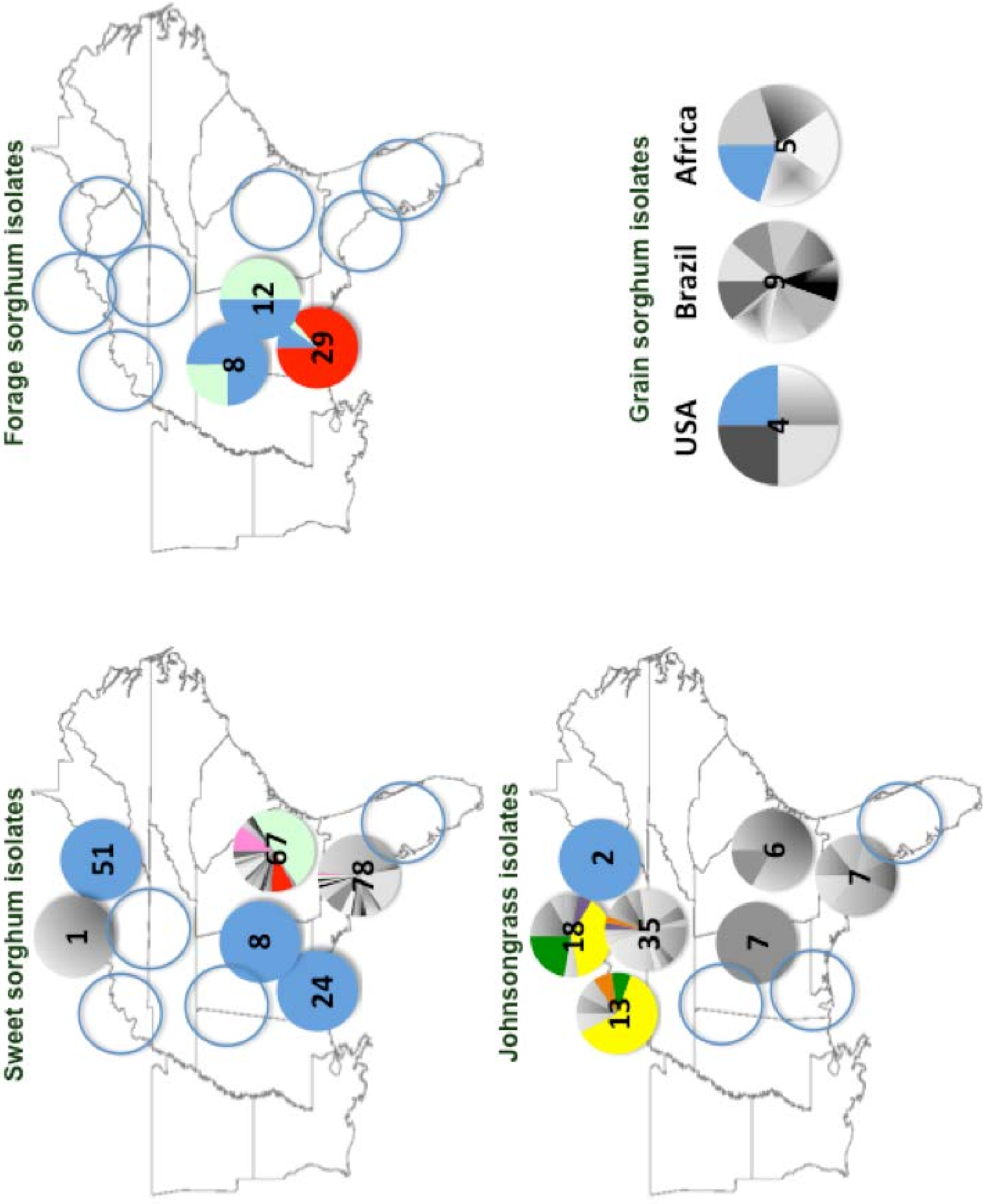


Figure 3.4. Dendrogram illustrating the relatedness of *Sorghum halepense* (JG) (gray boxes), and *S. bicolor* (red boxes) isolates, based on the polymorphic bands revealed by RAPD molecular fingerprinting. The latter group includes isolates from grain (GS1.001, S3.001, S26.001, CgSl1, S17.001 RIII42, and E031), sweet (SS), and forage (FS) sorghum. Isolates from sweet sorghum in the *S. halepense* group (*), and isolates from johnsongrass in the *S. bicolor* group (**). The yellow box (A) includes a representative of a related species *C. graminicola*, which is pathogenic only to maize. Brackets indicate: (A) maize isolate; (B) mostly johnsongrass isolates; (C) johnsongrass isolates only; (D) grain and sweet sorghum isolates from the U.S. and two johnsongrass isolates; (E) mostly sweet sorghum isolates from the U.S., and one grain sorghum isolate; (F) sweet and forage sorghum isolates from the U.S. and (G) grain sorghum isolates from Brazil and Zambia, and sweet sorghum isolates from FL.

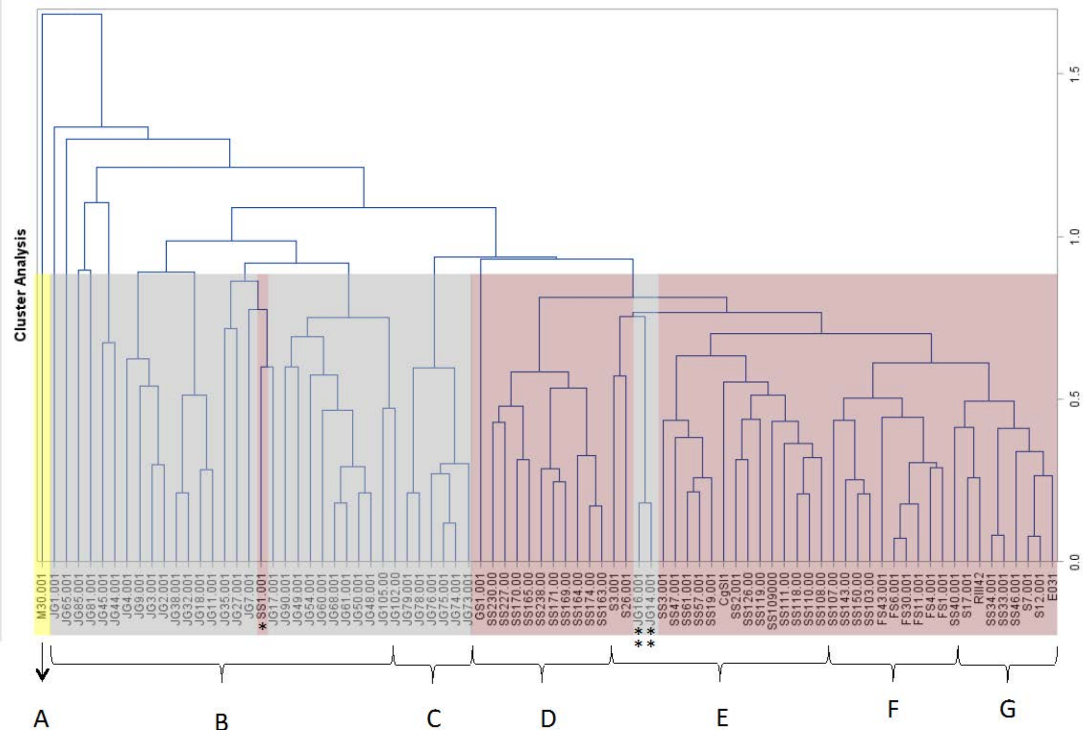


Figure 3.5. Average disease ratings (= the average number of plants rated 0-5) at four weeks post-inoculation in (A) first; (B) second; and (C) third repetition.

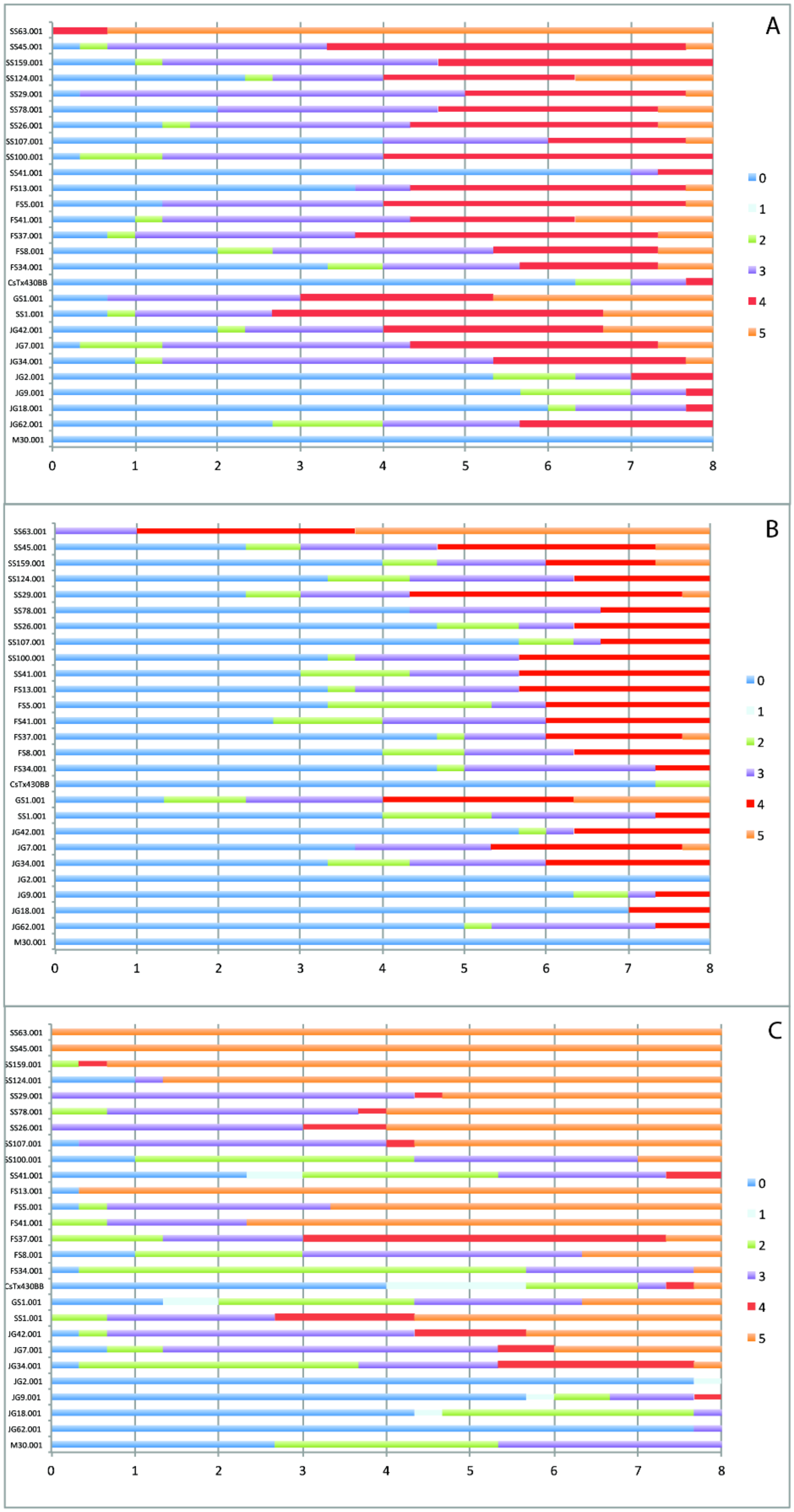


Figure 3.6. Pathogenicity assay of several strains of *Colletotrichum* on Sugar Drip. Incidence values, from the combined data sets of three experiments (summer and fall of 2014; and summer 2016), were used to calculate on the area under the disease progress curve (AUDPC). Black bars are the *S. bicolor* isolates and white bars are the *S. halepense* isolates.

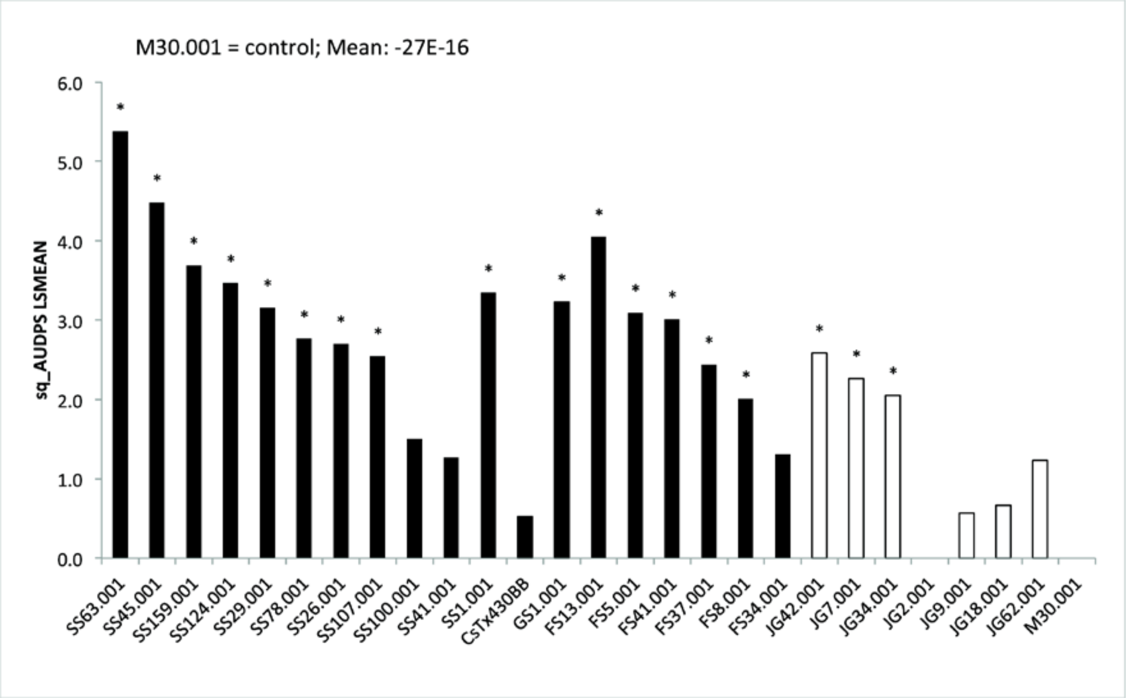


Figure 3.7. Pathogenicity assay of several strains of *Colletotrichum* on Sugar Drip. Incidence values, from the combined data sets of two experiments (July and August of 2016), were used to calculate on the area under the disease progress curve (AUDPC). Black bars are the *S. bicolor* isolates and white bars are the *S. halepense* isolates.

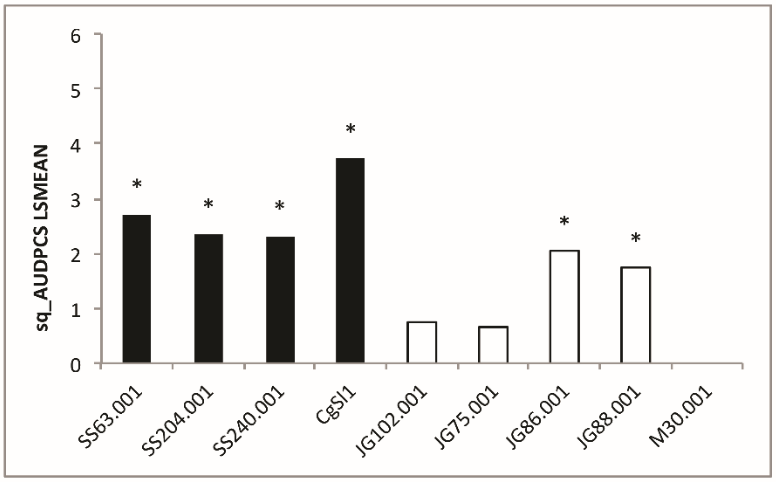


Figure 3.8. Greenhouse experiment comparing 13 fungal strains on Honey, Simon and Bailey. Treatment means were compared with the negative control (M30.001) using the Dunnett's test. An asterisk indicates that the treatment was significantly different from the control at P-value less than 0.05.

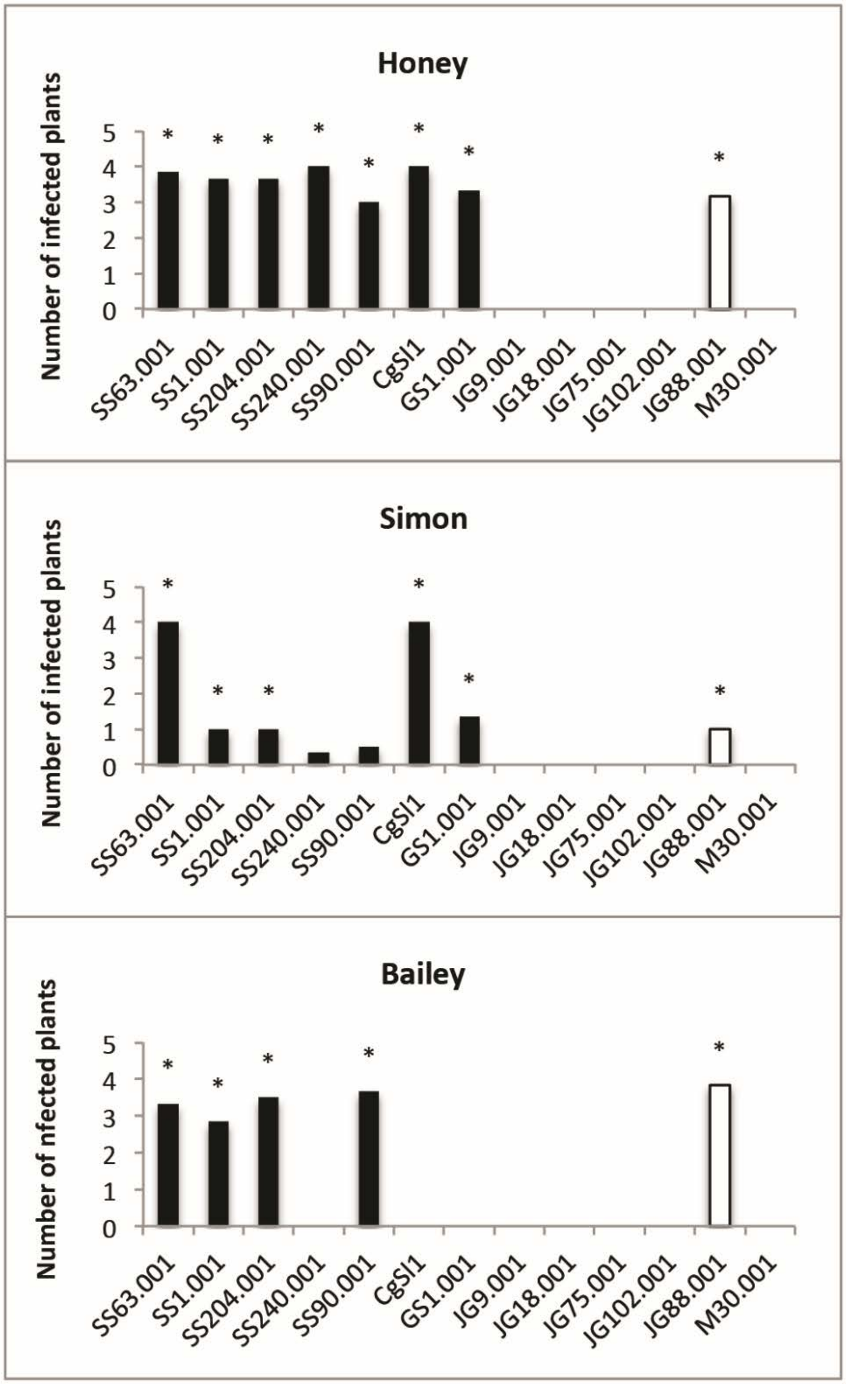


Figure 3.9. Greenhouse experiment comparing 13 fungal strains on Dwarf Yellow Milo, Planter, and Chinese Amber. Treatment means were compared with the negative control (M30.001) using the Dunnett's test. An asterisk indicates that the treatment was significantly different from the control at P-value less than 0.05.

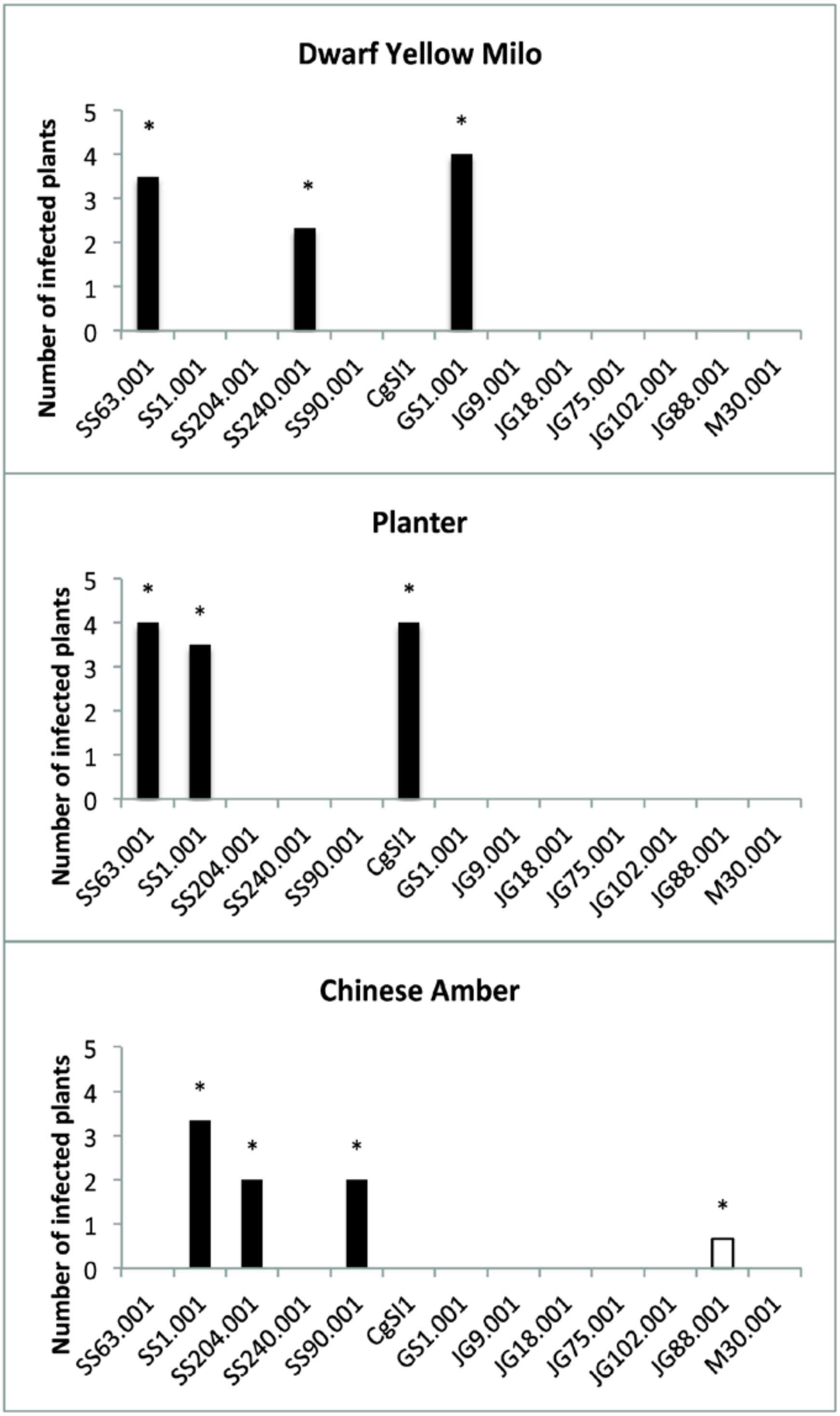


Figure 3.10. Greenhouse experiment comparing 13 fungal strains on Dale, Cowper and Orange. Treatment means were compared with the negative control (M30.001) using the Dunnett's test. An asterisk indicates that the treatment was significantly different from the control at P-value less than 0.05.

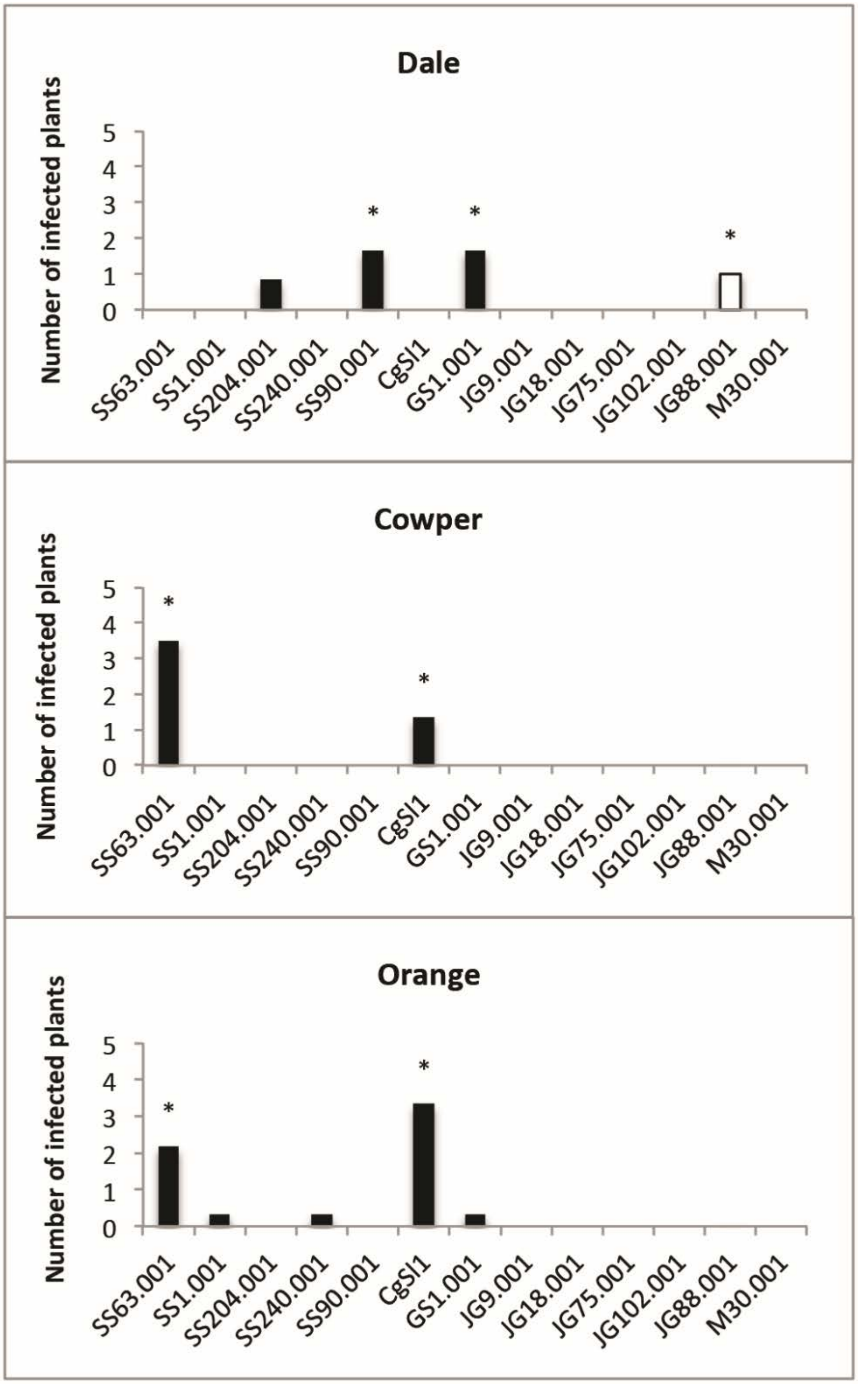


Figure 3.11. Greenhouse experiment comparing 13 fungal strains on N100, Della and Keller. Treatment means were compared with the negative control (M30.001) using the Dunnett's test. An asterisk indicates that the treatment was significantly different from the control at P-value less than 0.05.

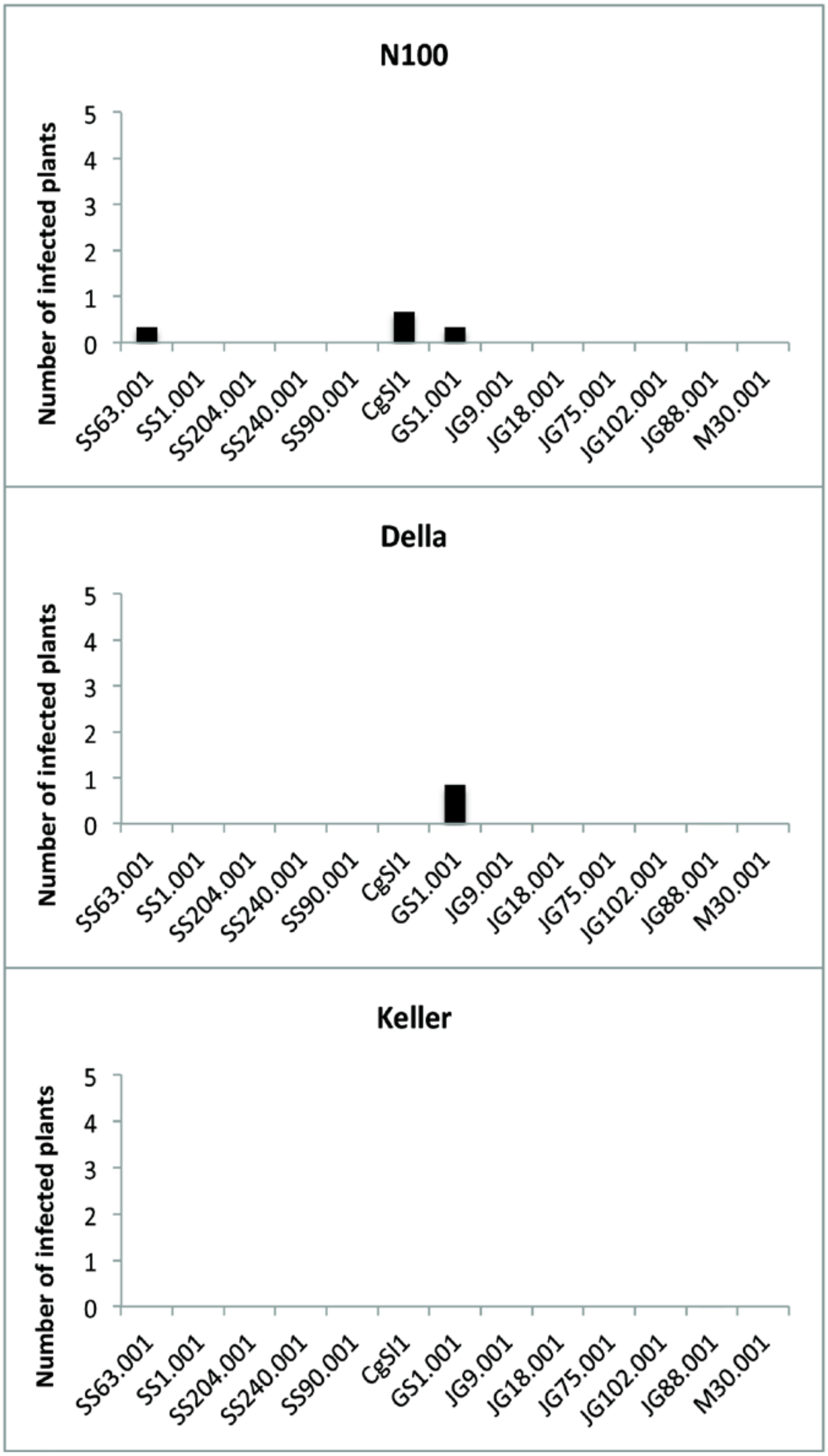
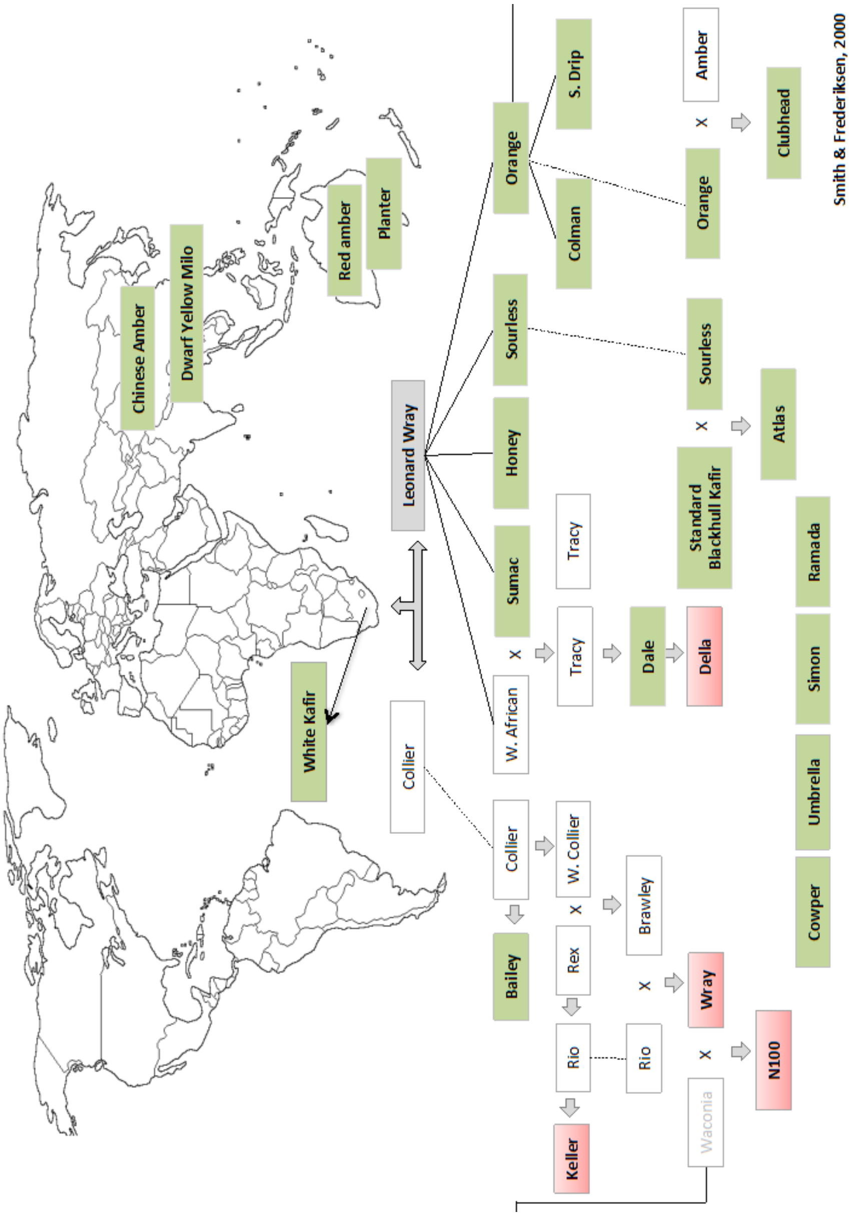


Figure 3.12. Summary of the inoculations performed to evaluate the infection potential of *Colletotrichum* isolates on different varieties of Sweet Sorghum. Table (A) represents the summary of the greenhouse experiment. The green boxes (S) indicate susceptibility and the orange (R) indicate resistance.

Strains	S. Drip	Honey	Simon	Bailey	C. amber	Dale	DYM	Cowper	Orange	Planter	N100	Della	Keller
SS63.001	S	S	S	S	R	R	S	S	S	S	R	R	R
SS1.001	S	S	S	S	S	R	R	R	R	S	R	R	R
SS204.001	S	S	S	S	S	R	R	R	R	R	R	R	R
SS90.001	-	S	R	S	S	S	R	R	R	R	R	R	R
CgSl1	S	S	S	R	R	R	R	S	S	S	R	R	R
GS1.001	S	S	S	R	R	S	S	R	R	R	R	R	R
JG88.001	S	S	S	S	S	S	R	R	R	R	R	R	R
SS240.001	S	S	R	R	R	R	S	R	R	R	R	R	R
JG9.001	R	R	R	R	R	R	R	R	R	R	R	R	R
JG18.001	R	R	R	R	R	R	R	R	R	R	R	R	R
JG75.001	R	R	R	R	R	R	R	R	R	R	R	R	R
JG102.001	R	R	R	R	R	R	R	R	R	R	R	R	R
M30.001	R	R	R	R	R	R	R	R	R	R	R	R	R

Figure 3.13. Pedigree of sweet sorghum varieties summarizing the inoculations performed in this chapter and data not shown. Green boxes, susceptible varieties, in which at least one isolate was able to cause disease. Pink boxes, resistant varieties, which were resistant to all the isolates tested.



Chapter 4 - A comparative genomic analysis of putative pathogenicity genes in the host-specific sibling species *Colletotrichum graminicola* and *Colletotrichum sublineola*

Abstract

The fungal sibling species *Colletotrichum graminicola* and *C. sublineola* cause anthracnose leaf blight and stalk rot of maize and of sorghum, respectively. In spite of their close evolutionary relationship, the two species appear to be completely host-specific. I confirmed the presence of inducible non-host resistance to *C. sublineola* by maize, and to *C. graminicola* by sorghum, in epidermal cells. This resistance was manifested as a rapid deployment of visible resistance responses and cell death. Host specificity in plant-pathogen interactions can often be attributed to the presence or absence of pathogen virulence factors, including small secreted protein (SSP) effectors and specialized secondary metabolites (SSM). I identified numerous non-conserved SSM-associated predicted proteins that could play a role in this non-host recognition. Probes against some of the genes encoding these proteins, and against several putative non-conserved SSP genes, were used to conduct RFLP analysis of a representative group of *Colletotrichum* strains from *Sorghum bicolor* and *S. halepense*. The strains could be assigned to one of three distinct groups based on allelic similarity, and these groups were correlated with the fingerprinting and pathogenicity data from chapter 3. The three groups could be differentiated from one another by the presence or absence of specific alleles of some of these pathogenicity genes (aka apparent allelic fixation).

Introduction

Members of the fungal genus *Colletotrichum* cause anthracnose diseases on nearly every plant species grown for food or fiber worldwide (38, 51). *Colletotrichum graminicola* (Ces.) Wils., and *C. sublineola* Henn., cause economically important anthracnose leaf blight and stalk rot diseases of maize (*Zea mays* L.), and sorghum (*Sorghum bicolor* [L.] Moench), respectively (88, 92, 165, 181). These two fungal sibling species are morphologically very similar, but reproductively isolated (181). Results of molecular phylogenetic analyses suggest that they diverged from a common ancestor relatively recently, perhaps at the same time as the split between maize and sorghum (thought to be approximately 12 million years ago) (39, 40, 61, 157, 165, 167, 181). There are no reports in the literature of *C. graminicola* infecting sorghum or of *C. sublineola* infecting maize in the field, and most studies agree that the two species are host-specific (45, 92, 105, 202). We have found that *C. sublineola* can infect maize stalk epidermal cells, and maize leaf sheath cells that are dead or dying (175, 187). This ability of *C. sublineola* to conditionally infect some maize tissues might explain two earlier papers that reported that maize was susceptible to isolates of *Colletotrichum* from sorghum (31, 198). It also suggests that host range is determined by active recognition of and response to the non-pathogen by most healthy tissues of the non-host, rather than structural barriers or the absence of some vital nutrient or other factor.

The determination of host range in plant pathogens is often attributed to the presence or absence of pathogen virulence factors, particularly specialized secondary metabolites (SSMs), and small secreted proteins (SSPs) (aka. effectors) (7, 36, 48, 59, 90, 103, 180).

The presence of particular SSMs has been associated with the determination of host range in some phytopathogenic fungi including *Alternaria* spp. (90) and *Cochliobolus* spp. (36). The major classes of fungal SSMs include polyketides, peptides, terpenes, and indole alkaloids (10, 11, 108). Each of these classes is associated with a specific family of proteins. These SSM-associated proteins are: polyketide synthases (PKS); nonribosomal peptide synthetases (NRPS); terpene synthases (TS); and dimethylallyl transferases (DMAT), respectively. Genes encoding these enzymes and other proteins involved in the production of the SSMs are often found physically associated in transcriptionally co-regulated gene clusters (113, 115).

Fungal effectors have been defined as SSPs that alter the structure or modulate the function of host cells to facilitate infection (62, 186). Some effectors are translocated and operate in the host cytoplasm (53, 60, 99, 100). Others function in the plant cell apoplast (95). Some effectors act as host specific toxins and induce apoptosis only in certain plant genotypes, conferring host specificity in several important necrotrophic pathogens (185, 189). Examples of known effector categories include serine proteases, necrosis and ethylene-inducing protein 1-like proteins (NEP1-like proteins), and small cysteine-rich proteins (48, 122, 141).

Some plants have evolved an ability to recognize and respond to certain effectors by activating defense pathways via specific resistance (R) proteins, a phenomenon known as effector-triggered immunity (ETI). In these cases, the effectors act as avirulence (avr) factors. Multiple rounds of mutation and selection of R and avr genes during a co-evolutionary “arms-race” leads to the presence of multiple pathogenic races expressing different combinations of avr genes within the pathogen population (93). Recent evidence suggests that inducible non-host resistance in many agriculturally-important pathosystems, particularly involving closely related hosts, is due to ETI. In these cases all members of the non-host plant species contain the same R gene(s), whereas all members of the nonpathogenic microbial species contain the corresponding avr gene(s) (33, 96, 114, 125, 153, 154, 168, 177, 178, 183).

A number of recent comparative genomics studies have confirmed that genes encoding SSM-associated proteins and SSPs show evidence of rapid evolution in related pathogens with different host ranges (7, 17, 21, 36, 50, 58, 74, 107, 127, 137, 138, 149, 152, 163, 203).

My goal in this chapter was to see if the differences in host specificity and molecular fingerprints among strains from *S. bicolor* versus *S. halepense* that I reported in chapter 3 were associated with differences in potential pathogenicity genes. I had three predictions: (1) that the sibling species *C. sublineola* and *C. graminicola* would differ in the presence of one or more genes predicted to encode SSPs and SSM-associated proteins (aka pathogenicity genes); (2) that pathogenicity genes found in *C. sublineola* but not in *C. graminicola* would be conserved among all members of the population infecting *Sorghum bicolor* in Kentucky and the southeastern U.S. (suggesting the possibility that they were involved in triggering ETI in maize); and (3) that populations affecting *S. bicolor* would be distinguishable from populations affecting *S. halepense* by allelic differences in individual pathogenicity genes, ie. allelic fixation (suggesting that the two populations are genetically isolated).

Material and Methods

Plant and fungal growth and inoculation: Fungal strains were preserved on silica gel at -80°C (179). Strains were cultured on potato dextrose agar (PDA, BD Difco, Franklin Lakes, NJ) under continuous fluorescent lights at 23°C. Spores were harvested from 2-week-old culture plates by gently scraping them from the surface, and washed three times before use.

Seeds of sweet sorghum (Sugar Drip) or maize (Mo17) were sown in a mixture of two parts sterile topsoil and three parts of Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada). Seedlings were maintained in the greenhouse with 14 hours of light, watered every other day to saturation using an automated overhead irrigation system, and fertilized beginning one week after emergence two or three times per month as needed with a solution of 150 ppm of Peters 20-10-20 (Scotts-Sierra Horticultural Product Co., Marysville, OH).

Maize and sorghum seedlings at the V6 stage were inoculated by applying a suspension of $5x10^6$ spores per ml with a compressed-air sprayer (Preval Model 267 Paint Spray Gun). After inoculation, the plants were incubated for 18 h in the dark at 25 °C in a dew chamber at 100% relative humidity before being returned to the greenhouse bench. Maize leaf sheaths were inoculated with a suspension of $5x10^5$ spores per ml as described previously in this dissertation.

Sequencing, assembly, and annotation of fungal genomes: The genomes of *C. graminicola* strain M5.001 and *C. sublineola* strain CgSl1 were sequenced to 29X and 43X, respectively, in the Advanced Genetic Technologies Center (AGTC) at the University of Kentucky. Genome assembly was done by using Newbler version 2.9. Annotation was done by using MAKER. The genome assemblies and predicted proteomes for *C.*

graminicola strain M1.001 and for *C. sublineola* strain TX430BB were downloaded from the NCBI BioProjects database (accession numbers PRJNA37879 and PRJNA246670, respectively). Genome assemblies and predicted proteomes for *C. sublineola* strain S3.001 and for *C. falcatum, C. somersetensis, C. caudatum, C. eremochloae,* and *C. zoysiae* were downloaded from the Joint Genomes Institute Genome Portal (http://genome.jgi.doe.gov/). The comparative genome analysis and general protein characterization (other than specifically mentioned in this dissertation chapter) was done by AGTC staff, Dr. Neil Moore, and Dr. Ester Buiate (2016).

Identification of orthologous and unique genes: Fungal protein sequences were downloaded from the Broad Institute (*C. graminicola, C. higginsianum, Fusarium graminearum, F. oxysporum, Verticillium dahliae, Aspergillus flavus*) and the Joint Genome Institute Genome Portal (*Trichoderma reesei, C. falcatum, C. somersetensis, C. caudatum, C. eremochloae, C. zoysiae*, *C. sublineola* strain S3.001). Protein sequences from *Epichloë festucae* were the FGENESH gene predictions previously used in the Clavicipitaceae analysis (151). Putative orthologs were identified by using two methods. The first method was application of Ortho-MCL and COCO-CL (COrrelation COefficient-based CLustering) to the annotations (94, 111), following a procedure previously used for ortholog identification within the Clavicipitaceae (151). The species included for comparison in the Ortho-MCL/COCO-CL analysis were: *C. graminicola*; *C. higginsianum*; *C. sublineola* CgSl1; *M. oryzae*; *E. festucae*; *F. graminearum*; *F. oxysporum*; *T. reesei*; *V. dahlia*; and *A. flavus*. The second method used for ortholog identification was Reciprocal Best Hit (RBH) with an expect-value cutoff of 1e-5 (119, 190). This method was used to compare proteins from *C. graminicola*, *C. sublineola*, and *C. higginsianum*.

Identification of SSM-associated proteins in *Colletotrichum sublineola*: I identified the five classes of candidate SSM-associated genes (PKS, NRPS, PKS-NRPS hybrid, DMAT, and TS) from *C. sublineola* by applying a process that included Pfam and Ortho-MCL/COCO-CL analysis; followed by manual annotation and domain validation in the Conserved Domain Database (CDD) (http://www.ncbi.nlm.nih.gov/cdd/); blastp comparisons with the NCBI nr database; and InterproScan analysis. This protocol has been described in more detail previously (131).

I manually annotated *C. sublineola* SSM gene clusters by evaluating Ortho-MCL/COCO-CL results for the genes that were located upstream and downstream of the SSM-associated backbone genes. Genes that had no or few orthologs were considered to belong to the clusters, while genes that were conserved in most or all of the ten species included in the analysis defined the outside boundaries of the clusters.

Phylogenetic analysis of *Colletotrichum* SSM-associated proteins: I performed a phylogenetic analysis of SM genes by using the tools on the phylogeny.fr website (http://www.phylogeny.fr/index.cgi). The A and KS N-terminal and C-terminal domains of the NRPS, PKS, and NRPS-PKS hybrids were identified by using the NCBI CCD. Amino acid sequences were aligned by using MUSCLE version 3.7, and phylogenies were inferred by maximum-likelihood using PhyML version 3.0. Statistical branch support was provided by an approximation to the standard likelihood ratio test, aLRT (52).

Preparation of probes and Southern blot: Genomic DNA extraction was performed following the protocol of (174). Genomic DNA from the *C. sublineola* strain CgSl1 was used as a template for PCR. Primers used for amplification of pathogenicity gene probes are listed in Table 4.1. PCR conditions consisted of 3 min of denaturation at 94°C; followed by 35 cycles of denaturation for 30 sec at 94°C, 60 sec of annealing at 59°C, and 40 seconds of extension at 72°C; and a final extension of 10 min at 72°C. The identities of representative amplicons of the expected size were confirmed by sequencing.

A total of 88 isolates were selected to represent the diversity among the 400 individuals in my collection, based on the molecular fingerprinting analyses described in chapter 3 of this dissertation (Supplemental Table 1, Supplemental Table 2). The Southern blot was performed as described previously. Each of the strains was hybridized with each of the probes (Table 4.1).

Analysis of allelic diversity among populations: Polymorphic RFLP bands were scored as present (1) or absent (0) (Supplemental Table 2). Jaccard's coefficients (91) were applied to the data by using the Statistical Analysis Software (SAS) package, in order to group the isolates in a dendrogram based on the overall similarity of the band patterns.

Results

Characterization of strain CgSl1, and confirmation of host specificity

Partial sequences of four genes have been used previously for multigene phylogenetic analysis of *Colletotrichum* (131). These included portions of the actin (*Act)* gene; the chitin synthase 1 (*Chs-1)* gene; the histone3 (*His3)* gene; and the beta-tubulin (*Tub2)* gene. These sequences from CgSl1 shared 100% identity with those of strain S.3001, the designated epitype specimen for *C. sublineola* (131) (Figure 4.1). The 5.8S nuclear ribosomal gene and the two flanking internal transcribed spacer (ITS) sequences from CgSl1 also shared 99.6% identity with the ITS sequence of S3.001. This indicates that CgSl1 belongs to the *C. sublineola* species as it is presently defined (Figure 4.1).

Colletotrichum graminicola strain M1.001 was isolated from maize in Missouri in the late 1970s (70). *Colletotrichum sublineola* strain CgSl1 was isolated in the early 1980s from grain sorghum in Indiana (92). I confirmed that these isolates were host-specific, and I also determined that the non-host reaction was associated with rapid cell death, typical of ETI (33, 96, 114, 125, 153, 154, 168, 177, 178, 183). Strain M1.001 caused typical, sporulating anthracnose lesions on maize leaves (cv. Mo17) within three days post inoculation (dpi), but on leaves of sorghum (cv. Sugar Drip) it produced only small reddish flecks, which failed to expand or sporulate even up to 7 dpi (Figure 4.2 A, D). Strain CgSl1 caused large, sporulating anthracnose lesions on sorghum, but not on maize leaves (Figure 4.2 B, C). Strain M1.001 readily infected and colonized multiple cells of detached leaf sheaths of maize by 48 hr after inoculation (hpi) and strain CgSl1 did the same in sorghum sheaths by 72 hpi (Figure 4.3 A, B). In contrast, *C. graminicola* failed to infect leaf sheath cells of sorghum, and *C. sublineola* failed to infect maize leaf sheath cells, even up to 6 dpi (Figure 4.3 C, D). Sorghum responded within 48 hpi to *C. graminicola* appressoria by an accumulation of numerous vesicles containing red pigments, and maize responded to *C. sublineola* appressoria by the formation of iridescent papillae (Figure 4.3 C, D). Previous studies have determined that the red pigments consist of various anthocyanin phytoalexins (162). The maize papillae are composed primarily of callose (117). Visible primary hyphae were always very small, and were produced in fewer than 1% of infection attempts in both non-host combinations. Unpenetrated cells beneath *C. sublineola* appressoria in maize leaf sheaths typically retained their ability to plasmolyze even up to 48 hpi, but cells containing rare penetration hyphae appeared granulated, and did not plasmolyze normally (Figure 4.4 A, B). Sorghum cells beneath *C. graminicola* appressoria usually plasmolyzed at 24 hpi, but by 48 hpi most of the cells had lost the ability to plasmolyze, whether they contained infection hyphae or not (Figure 4.4 C, D, Figure 4.5). Most of the cells in the mock-inoculated maize and sorghum controls still plasmolyzed normally up to 72 hpi (Figure 4.6). *Colletotrichum sublineola* and *C. graminicola* were able to colonize non-host leaf sheaths readily if the cells were killed first by a localized application of liquid nitrogen (Figure 4.7). These observations suggest that host specificity is based on active recognition of the non-pathogen by living non-host plant cells, followed by rapid deployment of defense responses targeting the infection sites, and ultimately plant cell death prior to, or coincident with, penetration.

Identification of SSM-associated genes in *C. sublineola* strain CgSl1

The program Ortho-MCL and the refiner COCO-CL were used to identify genes in *C. sublineola* that were orthologous to the previously identified SSM-associated genes of *C. graminicola* and *C. higginsianum* (131). Using this approach, combined with manual annotation, I identified 31 PKS genes, eight NRPS genes, six PKS-NRPS hybrid genes, 14 TS genes, and eight DMAT genes in *C. sublineola* (Table 4.2). Pfam analysis of the *C. sublineola* protein predictions identified 172 putative SSM domains. All of the SSM-associated genes that were identified by Ortho-MCL and COCO-CL (above) were included among the SSM genes identified after manual annotation of the Pfam domains. However, the Pfam analysis identified additional genes in each class (three TSs and one DMAT) encoded by *C. sublineola* that were not found in either *C. graminicola* or *C. higginsianum* (Table 4.2).

Phylogenetic analysis of the SSM-associated proteins

A phylogenetic analysis was performed to address the relationships among the putative SSM-associated proteins in *C. graminicola* and *C. sublineola*. The more distantly-related species *C. higginsianum*, a pathogen of Brassicaceae, was also included for comparison. SSM-associated genes in *C. graminicola* and *C. higginsianum* were previously published (131). After manual annotation and identification of overlapping gene models, I reduced the 58 PKS genes that were previously identified in *C. higginsianum* (131) to 36 complete genes for analysis (Table 4.2). The amino acid sequences of the adenylation domain (A domain) of NRPS proteins and PKS-NRPS hybrids (16, 164); the keto-synthase (KS) N-terminal and C-terminal domains of PKS proteins and PKS-NRPS hybrids (101); and the entire DMAT and TS protein sequences, were used for the phylogenetic analyses. The alignments are provided in Appendix 3 of this dissertation.

Results of the analysis revealed a high degree of diversity, with relatively few SSM-associated protein ortholog families that were conserved across all three *Colletotrichum* species (Figures 4.8-4.12). As expected, *C. graminicola* and *C. sublineola* shared more ortholog families than either shared with *C. higginsianum*, consistent with a more recent common ancestor. The presence of some ortholog families only in *C. higginsianum* and *C. graminicola*, or only in *C. higginsianum* and *C. sublineola*, suggested that some members of these families may have been lost since the divergence of *C. higginsianum* from the other two species. The PKS proteins were the largest and most diverse group of SSM-associated proteins, with 79 proteins or protein ortholog families across the three species.

The NRPS proteins comprised the smallest group, with only 15 different proteins or ortholog families. *Colletotrichum graminicola* and *C. sublineola* shared about half of their PKS proteins, and also about half of their PKS-NRPS hybrid and TS proteins. The DMAT and NRPS proteins were more highly conserved, with about two thirds represented in both species. Searches of the NCBI nr database, and of the predicted proteomes of five close relatives in the JGI database, revealed that there were no SSM-associated protein genes in either *C. sublineola* or in *C. graminicola* that were unique to either species (Supplemental Table 3).

Conservation of gene clusters: I identified gene clusters in *C. sublineola* by manual analysis of the genes located on either side of the “backbone” SSM-associated genes (i.e. the genes encoding PKS, NRPS, TS, DMAT, and PKS-NRPS hybrids) that had been identified by using Ortho-MCL/COCO-CL and Pfam. A total of 67 putative SSM-associated gene clusters in the *C. sublineola* genome (Supplemental Table 2), were compared with the 42 clusters that were previously identified from *C. graminicola* (131). There were 25 PKS gene clusters that appeared to be shared (with more than 50% of the genes in common) between *C. sublineola* and *C. graminicola*. One of these is the melanin cluster (Figure 4.13) (131), and another is likely to be responsible for the production of monorden because it is identical in gene structure and content with the RADS cluster of *Pochonia chlamydospora* (Figure 4.14) (176). *Colletotrichum sublineola* and *C. graminicola* also shared five DMATs clusters, six NRPSs gene clusters, four hybrids and fifteen TSs gene clusters (Supplemental table 2). One of these conserved TS clusters is probably involved in the production of carotenoids (131).

I found five and four PKS clusters that were not shared in *C. sublineola* and *C. graminicola*, respectively. Examples of non-conserved PKS gene clusters are shown in (Figures 4.15-4.16). *C. sublineola* also has four DMATs, two TSs, two NRPSs, and two hybrid gene clusters that are unique (Supplemental Table 2). An example of a unique *C. sublineola* NRPS cluster is shown in Figure 4.17.

Comparison of the *C. graminicola* and *C. sublineola* SSP genes

Colletotrichum graminicola M1.001 encodes 143 predicted SSPs that are not found in *C. sublineola* strain CgSl1, while *C. sublineola* has 301 that are not shared with *C. graminicola* (Buiate, 2016). After comparisons with the NCBI nr database and available genome data from a group of five close relatives of *C. graminicola* and *C. sublineola* (http://genome.jgi.doe.gov/), there appeared to be 21 SSPs that were specific only to *C. sublineola*, and there were no lineage specific SSM-associated genes. One gene, CSUB_000060, appears to be shared only by *C. graminicola* and *C. sublineola*, and is not present in any other species currently in the databases.

Genomic analysis of SSP and SSM diversity among different isolates

Two other genome assemblies are available for *C. sublineola*. The TX430BB strain was isolated from grain sorghum in Texas in the late 1980s, and was sequenced by Baroncelli et al., (2016) (7). The S3.001 strain is the epitype for the species (40, 41), and its genome assembly can be accessed from JGI (http://genome.jgi.doe.gov/). This strain was isolated in the late 1980s in Burkina Faso (181). *C. sublineola* isolate CgSl1 shared 99% of its predicted gene models with TX430BB and S3.001. Only 23 CgSl1 SSP genes were not found in the TX430BB assembly and only seven were not found in S3.001. Only 39 gene sequences were not found in either of the other two other strains, including 2 SSPs. All of the SSM-associated genes in CgSl1 appeared to be conserved in both other strains of *C. sublineola*.

RFLP analysis: putative effector proteins and secondary metabolites

Two secondary metabolite-associated genes from CgSl1, CSUB_020320 (PKS14) and CSUB_060040 (PKS40), were selected for analysis. PKS14 is conserved with *C. graminicola* and *C. higginsianum*, while PKS40 is found only in *C. sublineola* (Supplemental Table 3). PCR primers were designed to amplify the sequences encoding the KS domains for each gene (Table 4.1). Twelve putative SSP genes, seven of which appeared to be unique to *C. sublineola,* and not found in *C. graminicola* and *C. higginsianum* (Supplemental Table 4), were also chosen for use as probes. Two SSP genes, (CSE21 and CSE24) that were shared by *C. sublineola* and *C. graminicola* but not *C. higginsianum*, and three that were conserved in all three species (CSE33, CSE53, and CSE56) were also chosen for comparison (Supplemental Table 4).

All of the probes identified restriction fragment length polymorphisms among the isolates that were analyzed. Some isolates did not hybridize at all to some of the probes. Representative results are shown in (Figure 4.18). A dendrogram was built to illustrate the relatedness of 88 representative fungal strains (Supplemental Table 4) isolated from johnsongrass (JG), grain sorghum (GS1.001, S3.001, E031, CgSl1, TX430BB), sweet sorghum (SS), and forage sorghum (FS), based on RFLP allelic variation for the 14 putative effector and SMM-associated probes. This dendogram separated most of the isolates from cultivated sorghum from most of the isolates from johnsongrass (Figure 4.19). However, there were some exceptions. For example, strain SS1.001 from sweet sorghum grouped with johnsongrass isolates, while JG88.001, JG14.001, and JG35.001 clustered with the cultivated sorghum isolates. Similar results were found in the fingerprinting analysis that was reported in chapter 3 of this dissertation. Among the isolates from johnsongrass (and SS1.001), there were two distinct clusters, marked on the dendrogram by orange and blue boxes. Isolates within the blue box that were tested were pathogenic to *Sorghum bicolor*, while isolates in the orange box were not (chapter 3 of this dissertation). The isolates in the blue box could be differentiated from the others by several of the probes. Thus, none of the isolates in the blue box hybridized to CSSM3. All of them had unique alleles of CSE4, CSE17, CSE21, CSE53, and CSE56. The johnsongrass isolates in the orange box could be differentiated from the other two groups by the absence of CSE6. Furthermore, JG14.001, JG88.001, and JG35.001 shared alleles of CSE23 and CSE33 that were otherwise found only in isolates from cultivated sorghum, confirming their close relationship as revealed in the dendrogram.

Conclusions

This work was part of a larger project in which our team compared gene models from two contemporaneous, co-occurring strains of the sibling species *C. graminicola* and *C. sublineola*, and identified those that do not appear to be conserved as potential candidates for involvement in host specificity. This approach was based on previous studies that have shown that gene gain and loss is associated with host range in many plant pathogens, including *Colletotrichum* (7, 74).

The analysis confirmed that the genomes of the *C. graminicola* and *C. sublineola* strains were very similar to one another in both gene content and gene order, consistent with a relatively recent common ancestor. I confirmed that each strain was able to successfully colonize its own living host (maize and sorghum, respectively), while the closely related non-host underwent an apparent hypersensitive response upon challenge.

We found that 14% of the *C. graminicola* gene models, and 22% of the *C. sublineola* gene models, were not conserved in the other species. Certain categories of genes were especially likely to be non-conserved (NC) including genes that were predicted to encode SSPs and SSM-associated proteins that may play important roles in early events related to host recognition and the induction of compatibility. Some of the NC SSP gene sequences were also not conserved among different strains within each species, which suggested the possibility of selection within the population and a potential avr function. Races of both *C. sublineola* and *C. graminicola* have been reported to occur (3, 14, 30, 44, 128, 143, 184, 199). There were relatively few strain-specific SSP gene sequences, however, and no strain specific SSM sequences. This differs from some other fungal species, e.g. *Magnaporthe oryzae,* where the deletion of effector genes seems to be common, and to play an important role in the rapid evolution of new races (205, 207). However, comparisons with genome assemblies of the five closely related species within the graminicolous clade, accessed from JGI (http://genome.jgi.doe.gov/), suggests a more important role for deletion of SSP and SSM genes in speciation and host species adaptation, a finding that has also been reported by others based on comparative analyses of a wider range of *Colletotrichum* genera (7, 74).

Within the population of *Colletotrichum* strains from *Sorghum* spp., RFLP analysis revealed the presence of three distinct groups, based on allelic variations of several of the putative pathogenicity loci that were identified in this chapter. These groups were highly correlated with the fingerprinting data from chapter 3. Thus, one group consisted primarily of isolates from cultivated sorghum, while the other consisted primarily of isolates from johnsongrass. Exceptions to the rule in the RFLP analysis, were also exceptions in the fingerprinting analysis (e.g. JG88.001 and SS1.001), additional evidence that these represent occasional cross-infections of *S. bicolor* by members of the *S. halepense* population, and *vice versa*.

The isolates from johnsongrass were further divided into at least two different groups (the orange and blue groups) based on the RFLP data. These groups were consistent with pathogenicity data from chapter 3. Thus, the orange group contained isolates that appear to be host-specific to johnsongrass, while the members of the blue group (including SS1.001) are able to infect both *S. bicolor* and *S. halepense*. All the members of the orange group appear to lack one SSP gene (CSE6) that is present in all of the isolates that are able to infect *S. bicolor*. The orange group also differs from nearly all of the rest of the strains in having unique alleles of CSE56. Only JG35.001 outside of the orange group also shares one of these alleles.

All members of the blue group share the same allele of CSE6, which differs from the allele that is shared by all of the members of the green group of isolates that come mostly from *S. bicolor*. Other fixed allelic differences between the blue group and the other strains can be seen for CSE4, CSE17, CSE21, CSE53, and CSE56. Members of the blue group also lack the CSSM3 gene that is found in all of the other strains affecting *Sorghum* spp.

The allelic differences revealed by the RFLP analysis support the presence of three non-recombining sub-groups within the population of *Colletotrichum* strains affecting *Sorghum* spp. in the southeastern U.S.

Summary

I had three predictions for this chapter: (1) that the sibling species *C. sublineola* and *C. graminicola* would differ in the presence of one or more genes predicted to encode SSPs and SSM-associated proteins (aka. pathogenicity genes); (2) that pathogenicity genes found in *C. sublineola* but not in *C. graminicola* would be conserved among all members of the population infecting *Sorghum bicolor* in Kentucky and the southeastern U.S. (suggesting the possibility that they could be involved in triggering ETI in maize); and (3) that populations affecting *S. bicolor* would be distinguishable from populations affecting *S. halepense* by allelic differences in individual pathogenicity genes, ie. allelic fixation (suggesting that the populations are genetically isolated).

Prediction 1 was upheld. *Colletotrichum graminicola* and *C. sublineola* differed in the presence or absence of a large number of putative SSP and SSM genes and gene clusters. Prediction 2 was mostly supported. Analysis of RFLPs using probes against pathogenicity genes that were shared between *C. sublineola* and *C. graminicola* (CSE33, CSE53, CSE56, CSSM2, CSE21 and CSE24) revealed that all of them were present in all isolates. Among the genes that were not conserved with *C. graminicola* (CSE1, CSE4, CSE6, CSE12, CSE23, CSE29, and CSSM3), most were also shared among all the strains. However, one of the genes (CSE6) was absent from the orange group of strains from johnsongrass, and another (CSSM3) was not found among the members of the blue group. Prediction three was upheld, with the caveat that allelic variation differentiates two groups within the johnsongrass isolates, rather than one, and that one of those groups also includes a strain from sweet sorghum (SS1.001). These groups are consistent with pathogenicity assays in which one (the orange group) appears to be host-specific for johnsongrass while the other (the blue group) can affect either johnsongrass or sweet sorghum. The evidence points to these three groups being genetically isolated from one another.

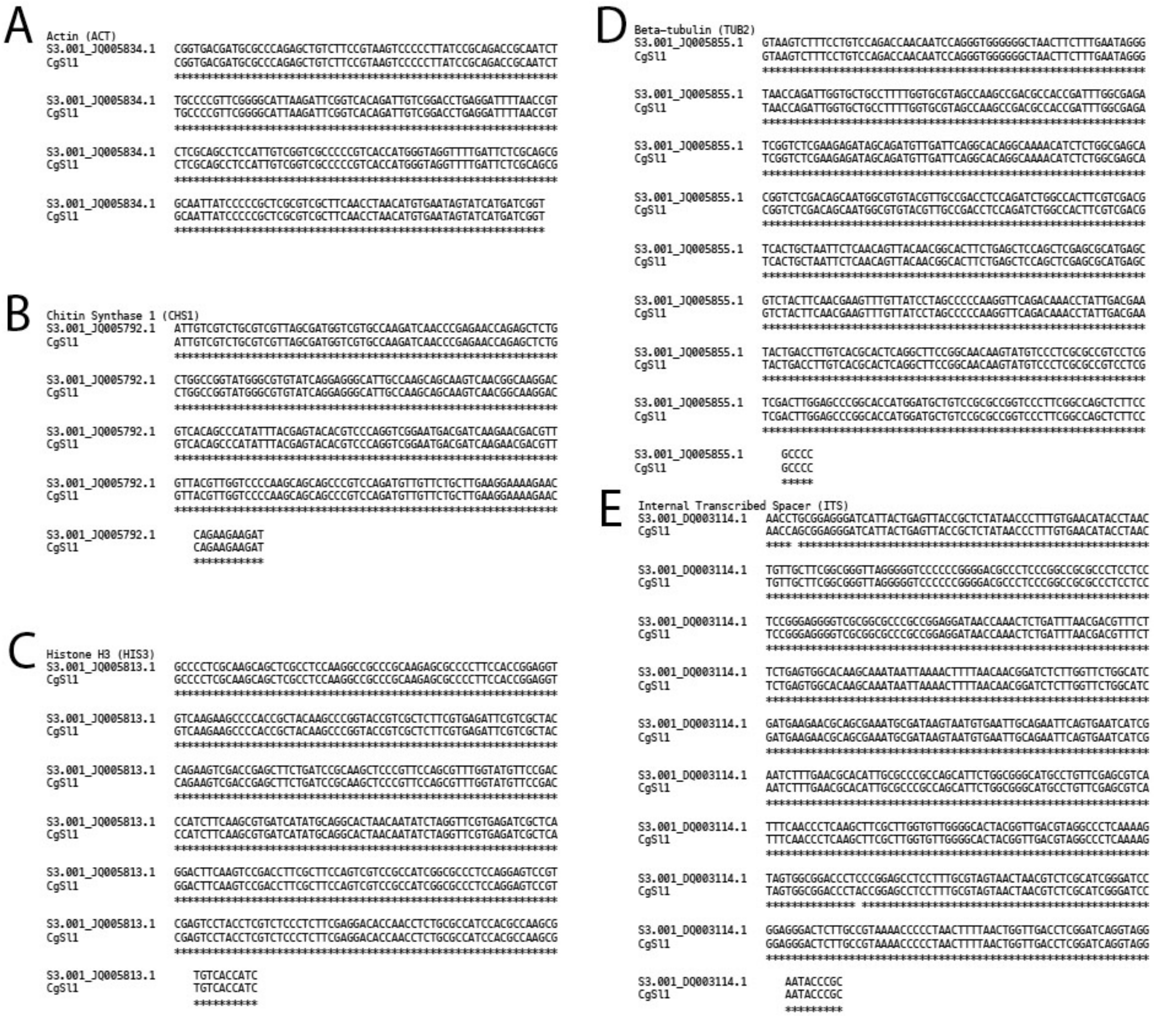
Table 4.1. Primers used as probes

Gene	Primer name	Sequences of Primers	Notes
CSUB_000030	CSE1	5’GGACAACGACAGAGCATCCC3’ 5’TGTCTGCAGTGGAGGAGGAG3’	UNIQUE Csub
CSUB_000060	CSE4	5’TGTGTAGCTGAACGTCGCAT3’ 5’GTTTGCTGTCGGGACACACT3’	UNIQUE Csub
CSUB_000110	CSE6	5’GCTATGGGAGCCCTAGCAGA3’ 5’CTAGCATTTTCCAACAGCGGC3’	UNIQUE Csub
CSUB_000120	CSE12	5’ATTATTGCTACCCTCGCGGC3’ 5’GAGAGTGTCCTCAGGGTGGG3’	UNIQUE Csub
CSUB_000150	CSE17	5’AGCTTCTGGCTTTAGTCGCC3’ 5’CTATCAGTCGTGCCCCCGTA3’	UNIQUE Csub
CSUB_054660	CSE21	5’GAAGGCCGTCACTATCCTCA3’ 5’CAGGCGAAGGATCGGAAG3’	SHARED CsubCgram
CSUB_014250	CSE23	5’CACCTAGGCCAGACCAAATG3’ 5’TCACTACTTCCTCCGTCTTGG3’	UNIQUE Csub
CSUB_101250	CSE24	5’GTCCCACCAATTTGGACTTG3’ 5’CTCGTCGTTCTAGGCTCACC3’	SHARED CsubCgram
CSUB_133010	CSE29	5’TGCGTCCTCAATGTGTTTTG3’ 5’CGAGACTTGGAGCGGAGATA3’	UNIQUE Csub
CSUB_020960	CSE33	5’TGCAGGTCGTAACAAAGTGG3’ 5’CTCATTATGCACGCCATCAC3’	CsubCgramChig
CSUB_054650	CSE53	5’CTTTTCCGCCCACATCCT3’ 5’ACAAAATTTTCCGGGTCGAT3’	CsubCgramChig
CSUB_075920	CSE56	5’TTGGCTTCTTTGGGAATACG3’ 5’GGACAATCCGGTCAGACCTA3’	CsubCgramChig
CSUB_020320 aka PKS14	CSSM2	5’GGCCTACCACGAGACTACCA3’ 5’TTGGCTGTGTGATACCTTGC 3’	CsubCgramChig
CSUB_060040 aka PKS40	CSSM3	5’GTCACTTTGATGCCCGATTT3’ 5’GGTCTCGCTAGGACATTGGA3’	UNIQUE Csub

Table 4.2. Ortho-MCL prediction of shared SSM genes for the three species of *Colletotrichum*. PKS, polyketide synthases; NRPSs, non-ribosomal peptide synthetases; PKS-NRPS hybrids that contain at least one PKS and at least one NRPS domain; DMAT, dimethylallyl transferases; and TS, terpene synthases. The numbers in parentheses represent the total number of genes in each category based on Pfam predictions for *C. sublineola*. *Manual annotation was performed on data retrieved from O'Connell et al, 2012.

Type	*C. sublineola*	*C. graminicola**	*C. higginsianum**
PKS	31 (31)	36	36
NRPS	8 (8)	7	12
PKS-NRPS	6 (6)	4	3
DMAT	8 (9)	7	10
TS	14 (17)	14	17

Figure 4.1. Alignments of sequences of CgSl1 with species type S3.001. A: actin, B: chitin synthase, C: histone H3, D: beta-tubulin, E: ITS. Alignments were performed with MUSCLE version 3.7.



```
A
Actin (ACT)
S3.001_JQ005834.1   CGGTGACGATGCGCCCAGAGCTGTCTTCCGTAAGTCCCCCTTATCCGCAGACCGCAATCT
CgSl1               CGGTGACGATGCGCCCAGAGCTGTCTTCCGTAAGTCCCCCTTATCCGCAGACCGCAATCT
                    ************************************************************

S3.001_JQ005834.1   TGCCCCGTTCGGGGCATTAAGATTCGGTCACAGATTGTCGGACCTGAGGATTTTAACCGT
CgSl1               TGCCCCGTTCGGGGCATTAAGATTCGGTCACAGATTGTCGGACCTGAGGATTTTAACCGT
                    ************************************************************

S3.001_JQ005834.1   CTCGCAGCCTCCATTGTCGGTCGCCCCCGTCACCATGGGTAGGTTTTGATTCTCGCAGCG
CgSl1               CTCGCAGCCTCCATTGTCGGTCGCCCCCGTCACCATGGGTAGGTTTTGATTCTCGCAGCG
                    ************************************************************

S3.001_JQ005834.1   GCAATTATCCCCCGCTCGCGTCGCTTCAACCTAACATGTGAATAGTATCATGATCGGT
CgSl1               GCAATTATCCCCCGCTCGCGTCGCTTCAACCTAACATGTGAATAGTATCATGATCGGT
                    **********************************************************

B
Chitin Synthase 1 (CHS1)
S3.001_JQ005792.1   ATTGTCGTCTGCGTCGTTAGCGATGGTCGTGCCAAGATCAACCCGAGAACCAGAGCTCTG
CgSl1               ATTGTCGTCTGCGTCGTTAGCGATGGTCGTGCCAAGATCAACCCGAGAACCAGAGCTCTG
                    ************************************************************

S3.001_JQ005792.1   CTGGCCGGTATGGGCGTGTATCAGGAGGGCATTGCCAAGCAGCAAGTCAACGGCAAGGAC
CgSl1               CTGGCCGGTATGGGCGTGTATCAGGAGGGCATTGCCAAGCAGCAAGTCAACGGCAAGGAC
                    ************************************************************

S3.001_JQ005792.1   GTCACAGCCCATATTTACGAGTACACGTCCCAGGTCGGAATGACGATCAAGAACGACGTT
CgSl1               GTCACAGCCCATATTTACGAGTACACGTCCCAGGTCGGAATGACGATCAAGAACGACGTT
                    ************************************************************

S3.001_JQ005792.1   GTTACGTTGGTCCCCAAGCAGCAGCCCGTCCAGATGTTGTTCTGCTTGAAGGAAAAGAAC
CgSl1               GTTACGTTGGTCCCCAAGCAGCAGCCCGTCCAGATGTTGTTCTGCTTGAAGGAAAAGAAC
                    ************************************************************

S3.001_JQ005792.1      CAGAAGAAGAT
CgSl1                  CAGAAGAAGAT
                       ***********

C
Histone H3 (HIS3)
S3.001_JQ005813.1   GCCCCTCGCAAGCAGCTCGCCTCCAAGGCCGCCCGCAAGAGCGCCCCTTCCACCGGAGGT
CgSl1               GCCCCTCGCAAGCAGCTCGCCTCCAAGGCCGCCCGCAAGAGCGCCCCTTCCACCGGAGGT
                    ************************************************************

S3.001_JQ005813.1   GTCAAGAAGCCCCACCGCTACAAGCCCGGTACCGTCGCTCTTCGTGAGATTCGTCGCTAC
CgSl1               GTCAAGAAGCCCCACCGCTACAAGCCCGGTACCGTCGCTCTTCGTGAGATTCGTCGCTAC
                    ************************************************************

S3.001_JQ005813.1   CAGAAGTCGACCGAGCTTCTGATCCGCAAGCTCCCGTTCCAGCGTTTGGTATGTTCCGAC
CgSl1               CAGAAGTCGACCGAGCTTCTGATCCGCAAGCTCCCGTTCCAGCGTTTGGTATGTTCCGAC
                    ************************************************************

S3.001_JQ005813.1   CCATCTTCAAGCGTGATCATATGCAGGCACTAACAATATCTAGGTTCGTGAGATCGCTCA
CgSl1               CCATCTTCAAGCGTGATCATATGCAGGCACTAACAATATCTAGGTTCGTGAGATCGCTCA
                    ************************************************************

S3.001_JQ005813.1   GGACTTCAAGTCCGACCTTCGCTTCCAGTCGTCCGCCATCGGCGCCCTCCAGGAGTCCGT
CgSl1               GGACTTCAAGTCCGACCTTCGCTTCCAGTCGTCCGCCATCGGCGCCCTCCAGGAGTCCGT
                    ************************************************************

S3.001_JQ005813.1   CGAGTCCTACCTCGTCTCCCTCTTCGAGGACACCAACCTCTGCGCCATCCACGCCAAGCG
CgSl1               CGAGTCCTACCTCGTCTCCCTCTTCGAGGACACCAACCTCTGCGCCATCCACGCCAAGCG
                    ************************************************************

S3.001_JQ005813.1      TGTCACCATC
CgSl1                  TGTCACCATC
                       **********

D
Beta-tubulin (TUB2)
S3.001_JQ005855.1   GTAAGTCTTTCCTGTCCAGACCAACAATCCAGGGTGGGGGCTAACTTCTTTGAATAGGG
CgSl1               GTAAGTCTTTCCTGTCCAGACCAACAATCCAGGGTGGGGGCTAACTTCTTTGAATAGGG
                    ************************************************************

S3.001_JQ005855.1   TAACCAGATTGGTGCTGCCTTTTGGTGCGTAGCCAAGCCGACGCCACCGATTTGGCGAGA
CgSl1               TAACCAGATTGGTGCTGCCTTTTGGTGCGTAGCCAAGCCGACGCCACCGATTTGGCGAGA
                    ************************************************************

S3.001_JQ005855.1   TCGGTCTCGAAGAGATAGCAGATGTTGATTCAGGCACAGGCAAAACATCTCTGGCGAGCA
CgSl1               TCGGTCTCGAAGAGATAGCAGATGTTGATTCAGGCACAGGCAAAACATCTCTGGCGAGCA
                    ************************************************************

S3.001_JQ005855.1   CGGTCTCGACAGCAATGGCGTGTACGTTGCCGACCTCCAGATCTGGCCACTTCGTCGACG
CgSl1               CGGTCTCGACAGCAATGGCGTGTACGTTGCCGACCTCCAGATCTGGCCACTTCGTCGACG
                    ************************************************************

S3.001_JQ005855.1   TCACTGCTAATTCTCAACAGTTACAACGGCACTTCTGAGCTCCAGCTCGAGCGCATGAGC
CgSl1               TCACTGCTAATTCTCAACAGTTACAACGGCACTTCTGAGCTCCAGCTCGAGCGCATGAGC
                    ************************************************************

S3.001_JQ005855.1   GTCTACTTCAACGAAGTTTGTTATCCTAGCCCCCAAGGTTCAGACAAACCTATTGACGAA
CgSl1               GTCTACTTCAACGAAGTTTGTTATCCTAGCCCCCAAGGTTCAGACAAACCTATTGACGAA
                    ************************************************************

S3.001_JQ005855.1   TACTGACCTTGTCACGCACTCAGGCTTCCGGCAACAAGTATGTCCCTCGCGCCGTCCTCG
CgSl1               TACTGACCTTGTCACGCACTCAGGCTTCCGGCAACAAGTATGTCCCTCGCGCCGTCCTCG
                    ************************************************************

S3.001_JQ005855.1   TCGACTTGGAGCCCGGCACCATGGATGCTGTCCGCGCCGGTCCCTTCGGCCAGCTCTTCC
CgSl1               TCGACTTGGAGCCCGGCACCATGGATGCTGTCCGCGCCGGTCCCTTCGGCCAGCTCTTCC
                    ************************************************************

S3.001_JQ005855.1      GCCCC
CgSl1                  GCCCC
                       *****

E
Internal Transcribed Spacer (ITS)
S3.001_DQ003114.1   AACCTGCGGAGGGATCATTACTGAGTTACCGCTCTATAACCCTTTGTGAACATACCTAAC
CgSl1               AACCAGCGGAGGGATCATTACTGAGTTACCGCTCTATAACCCTTTGTGAACATACCTAAC
                    **** *******************************************************

S3.001_DQ003114.1   TGTTGCTTCGGCGGGTTAGGGGGTCCCCCCGGGGACGCCCTCCCGGCCGCGCCCTCCTCC
CgSl1               TGTTGCTTCGGCGGGTTAGGGGGTCCCCCCGGGGACGCCCTCCCGGCCGCGCCCTCCTCC
                    ************************************************************

S3.001_DQ003114.1   TCCGGGAGGGGTCGCGGCGCCCGCCGGAGGATAACCAAACTCTGATTTAACGACGTTTCT
CgSl1               TCCGGGAGGGGTCGCGGCGCCCGCCGGAGGATAACCAAACTCTGATTTAACGACGTTTCT
                    ************************************************************

S3.001_DQ003114.1   TCTGAGTGGCACAAGCAAATAATTAAAACTTTTAACAACGGATCTCTTGGTTCTGGCATC
CgSl1               TCTGAGTGGCACAAGCAAATAATTAAAACTTTTAACAACGGATCTCTTGGTTCTGGCATC
                    ************************************************************

S3.001_DQ003114.1   GATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCATCG
CgSl1               GATGAAGAACGCAGCGAAATGCGATAAGTAATGTGAATTGCAGAATTCAGTGAATCATCG
                    ************************************************************

S3.001_DQ003114.1   AATCTTTGAACGCACATTGCGCCCGCCAGCATTCTGGCGGGCATGCCTGTTCGAGCGTCA
CgSl1               AATCTTTGAACGCACATTGCGCCCGCCAGCATTCTGGCGGGCATGCCTGTTCGAGCGTCA
                    ************************************************************

S3.001_DQ003114.1   TTTCAACCCTCAAGCTTCGCTTGGTGTTGGGGCACTACGGTTGACGTAGGCCCTCAAAAG
CgSl1               TTTCAACCCTCAAGCTTCGCTTGGTGTTGGGGCACTACGGTTGACGTAGGCCCTCAAAAG
                    ************************************************************

S3.001_DQ003114.1   TAGTGGCGGACCCTCCCGGAGCCTCCTTTGCGTAGTAACTAACGTCTCGCATCGGGATCC
CgSl1               TAGTGGCGGACCCTACCGGAGCCTCCTTTGCGTAGTAACTAACGTCTCGCATCGGGATCC
                    ************** *********************************************

S3.001_DQ003114.1   GGAGGGACTCTTGCCGTAAAACCCCCTAACTTTTAACTGGTTGACCTCGGATCAGGTAGG
CgSl1               GGAGGGACTCTTGCCGTAAAACCCCCTAACTTTTAACTGGTTGACCTCGGATCAGGTAGG
                    ************************************************************

S3.001_DQ003114.1      AATACCCGC
CgSl1                  AATACCCGC
                       *********
```

Figure 4.2. A, maize inoculated with *C. graminicola*, B, sorghum inoculated with *C. sublineola*, C, maize inoculated with *C. sublineola*, and D, sorghum inoculated with *C. graminicola*, E, maize mock inoculated with water, and F, sorghum mock inoculated with water. All at 7 dpi.

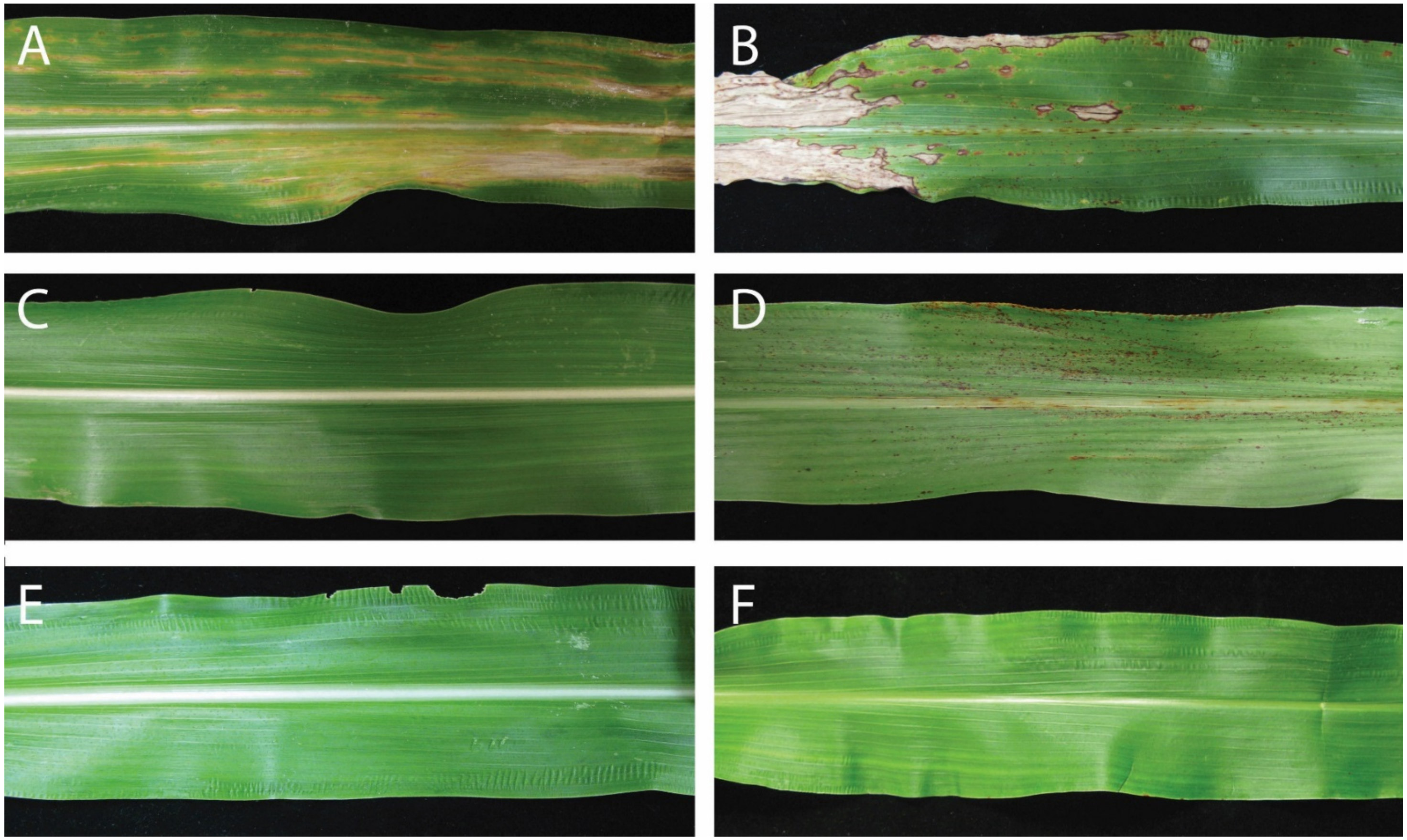


Figure 4.3. A, *C. graminicola* hyphae in maize leaves, 48 dpi. B. *C. sublineola* hyphae in sorghum leaves, 72 dpi. C. *C. graminicola* on sorghum, 48 dpi, white arrow indicates red vesicles surrounding the appressorium. D. *C. sublineola* on maize, 48 dpi, white arrow indicates an iridescent papillum beneath a melanized appressorium. Scale bars equal to 50 μm.

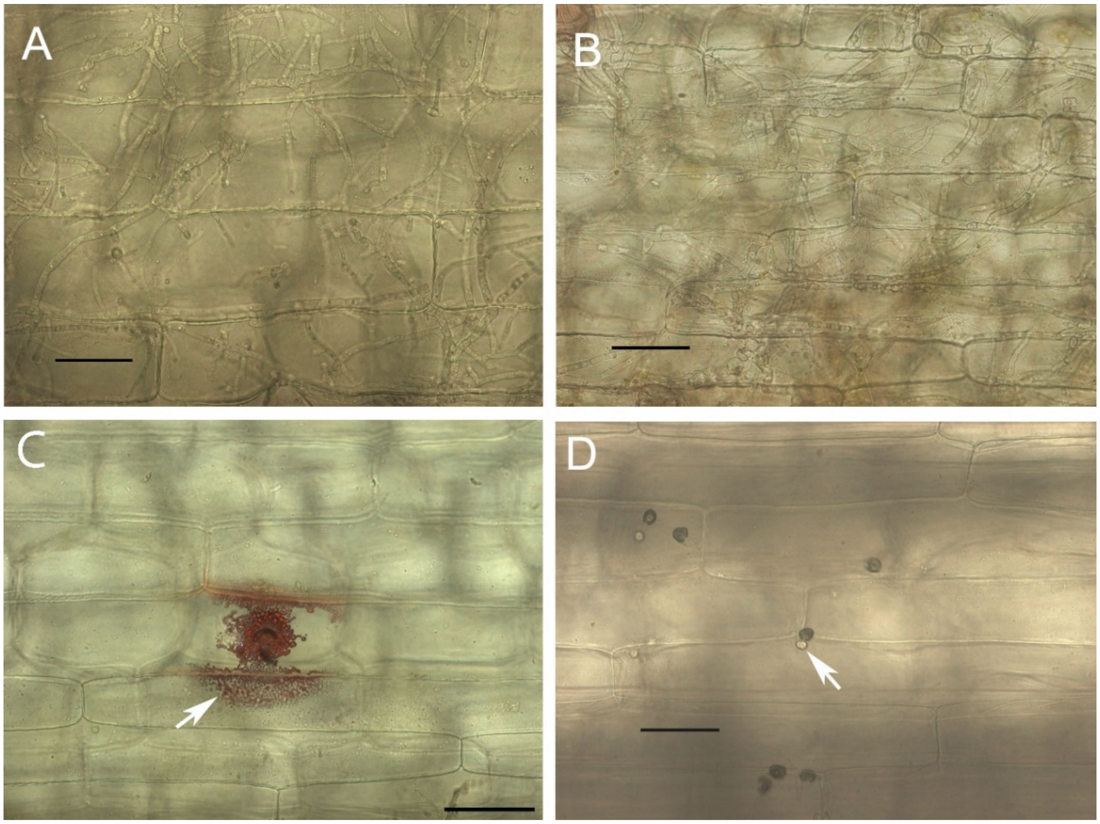


Figure 4.4. A. CgSl1 on maize sheath, 48hpi. Cell beneath appressorium (white arrow) plasmolyzes normally. B. CgSl1 on maize sheath, small penetration hypha (white arrow) 48hpi. Adjacent cell (black arrow) plasmolyzes normally. Cell containing penetration hypha appears granulated, plasma membrane visible but appears abnormal. C. M1.001 on sorghum sheath, 24 hpi., cell beneath appressorium (white arrow) plasmolyzes normally. D. M1.001 on sorghum sheath, 48 hpi. Treated with sucrose but no plasmolysis evident in any of the cells in the vicinity of the appressoria (white arrows). Scale bars equal to 50 μm.

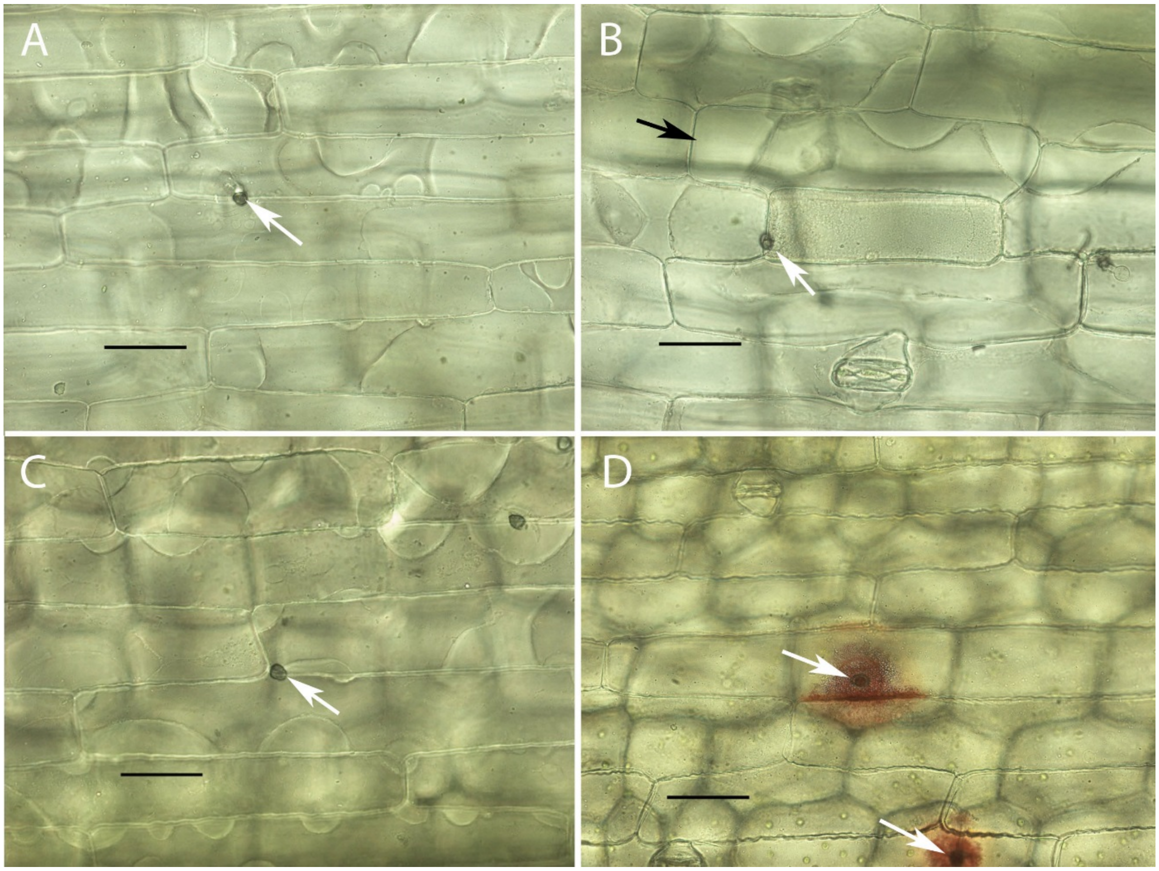


Figure 4.5. M1.001 on Sugar Drip sorghum, 48 hpi, cells beneath appressoria (white arrows) plasmolyzed (result not typical). Scale bars equal to 50 µm.

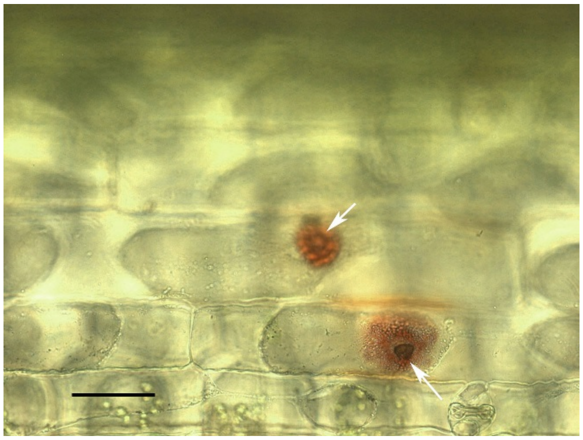


Figure 4.6. Plasmolysis controls. A: Maize leaf sheath, 72 hours after mock inoculation, most cells still plasmolyze. B: Sugar Drip leaf sheath, 72 hours after mock inoculation, most cells still plasmolyze. Scale bars equal to 50 μm.

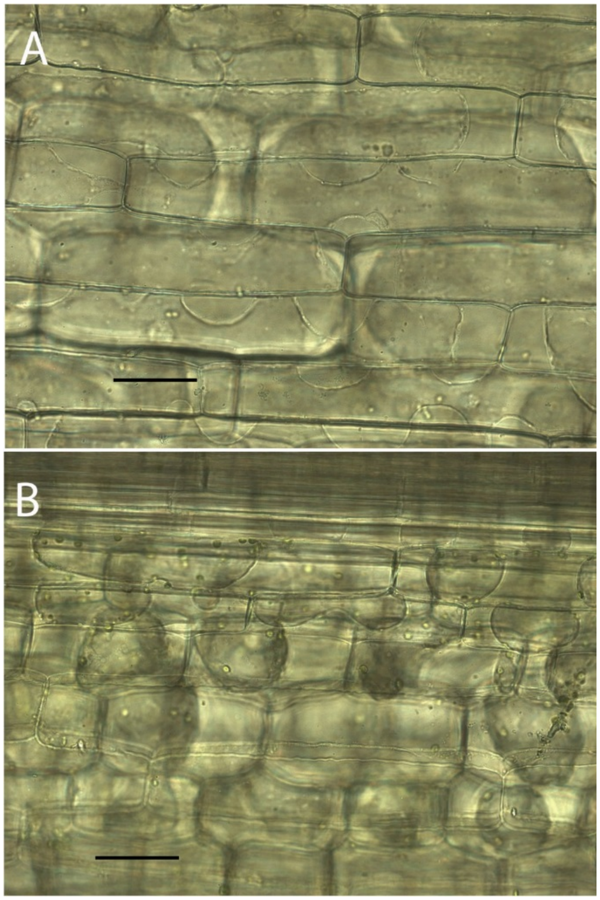
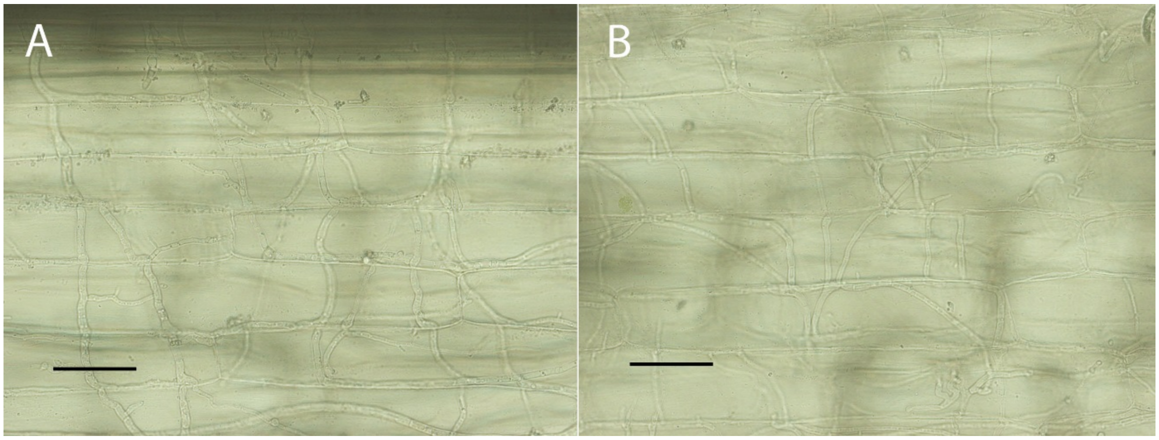


Figure 4.7. A. CgSl1 growing in cells of maize sheaths killed by liquid nitrogen, 48 hpi. B. M1.001 growing in cells of sorghum sheaths killed by liquid nitrogen, 48 hpi. Scale bars equal to 50 µm.



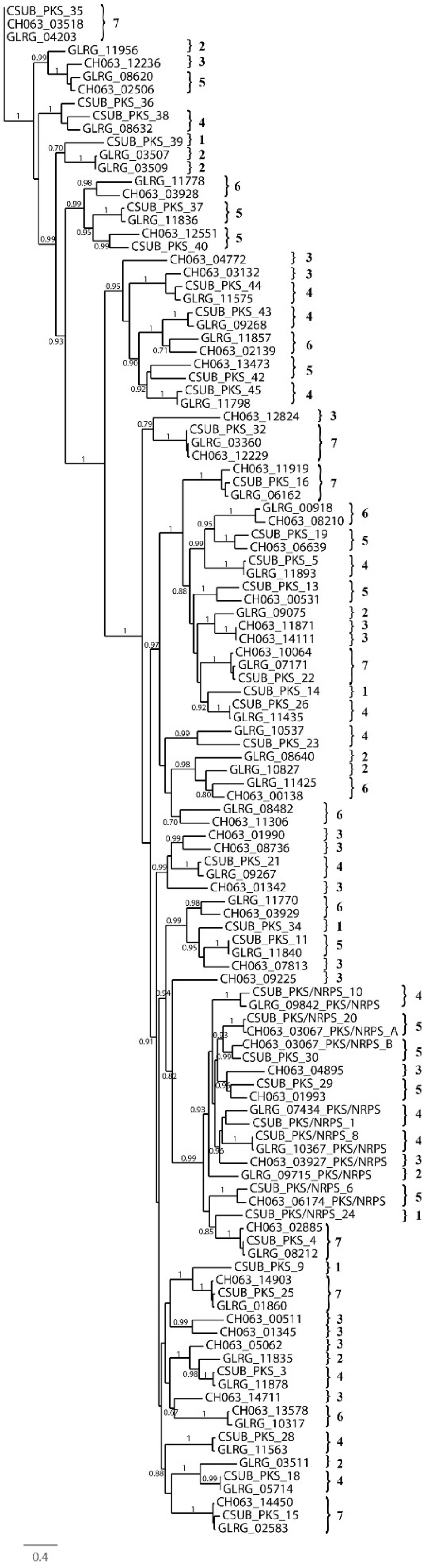


Figure 4.8. Phylogenetic tree of the ketoacyl CoA synthetase domain sequences of putative PKSs and PKS-NRPS hybrids. Sequences were aligned by using MUSCLE version 3.7, and phylogenies were inferred by maximum-likelihood using PhyML version 3-0 Statistical. The numbers on the branch nodes indicate support values above 50%, calculated by an approximation to the standard Likelihood Ratio Test, aLRT. Brackets refer to proteins found only in (1) *C. sublineola*, (2) *C. graminicola*, (3) *C. higginsianum*; (4) *C. sublineola* and *C. graminicola*, (5) *C. sublineola* and *C. higginsianum*, (6) *C. graminicola* and *C. higginsianum*, (7) *C. sublineola*, *C. graminicola* and *C. higginsianum***.**

Figure 4.9. A. Phylogenetic tree of the terpene synthase sequences. Sequences were aligned by using MUSCLE version 3.7, and phylogenies were inferred by maximum-likelihood using PhyML version 3-0 Statistical. The numbers on the branch nodes indicate support values above 50%, calculated by an approximation to the standard Likelihood Ratio Test, aLRT. Brackets refer to proteins found only in (1) *C. sublineola*, (2) *C. graminicola*, (3) *C. higginsianum*; (4) *C. sublineola* and *C. graminicola*, (5) *C. sublineola* and *C. higginsianum*, (6) *C. graminicola* and *C. higginsianum*, (7) *C. sublineola*, *C. graminicola* and *C. higginsianum***.**

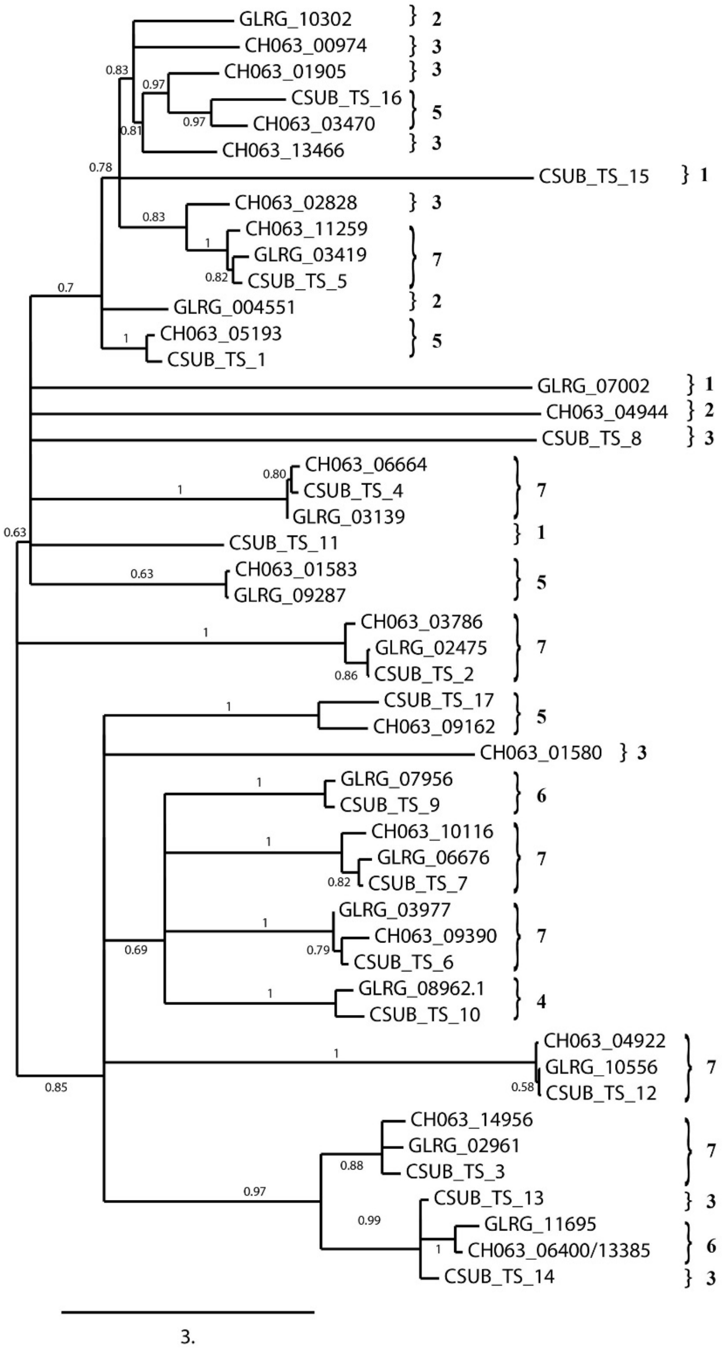
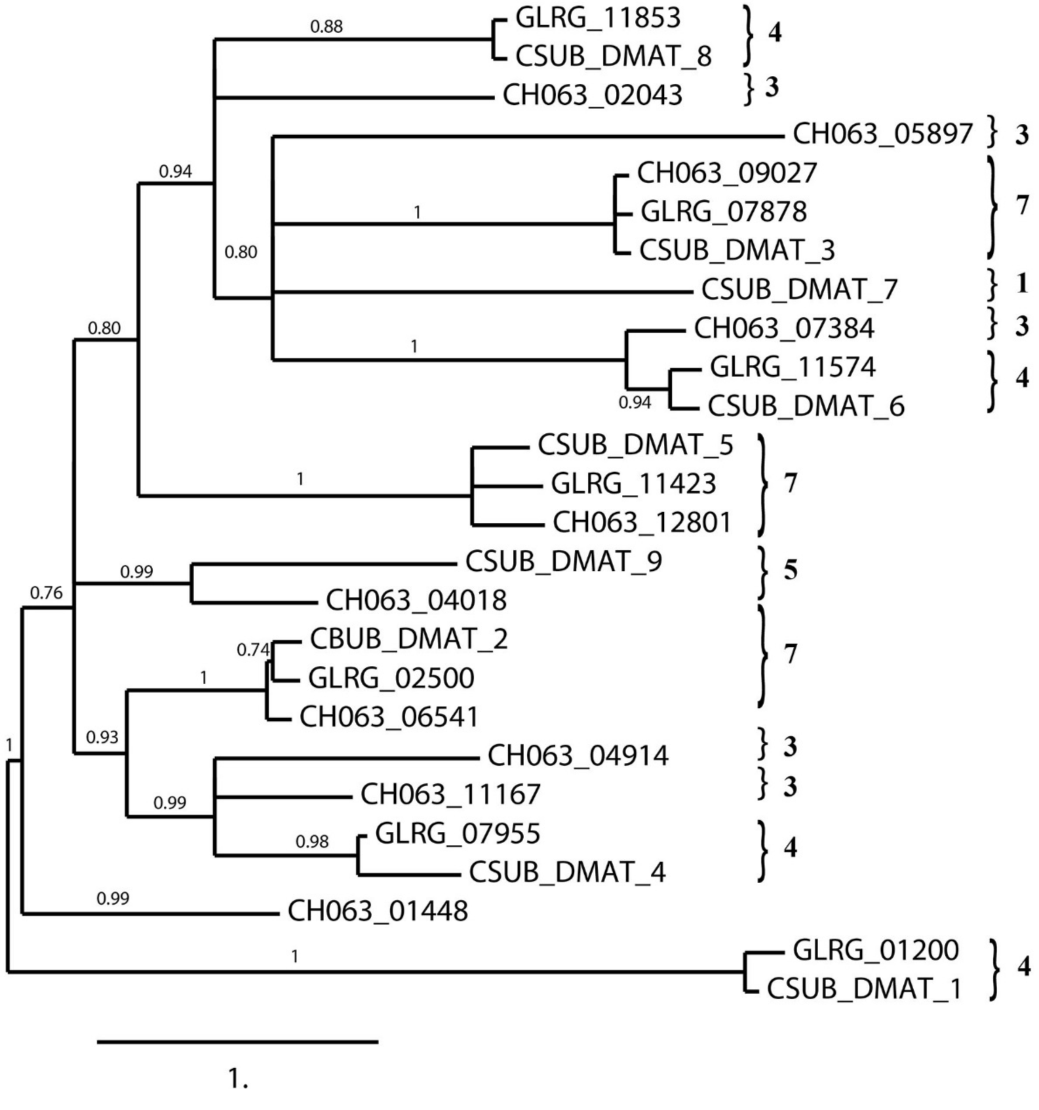


Figure 4.10. Phylogenetic tree of the dimethylallyl transferase sequences. Sequences were aligned by using MUSCLE version 3.7, and phylogenies were inferred by maximum-likelihood using PhyML version 3-0 Statistical. The numbers on the branch nodes indicate support values above 50%, calculated by an approximation to the standard Likelihood Ratio Test, aLRT. Brackets refer to proteins found only in (1) *C. sublineola*, (2) *C. graminicola*, (3) *C. higginsianum*; (4) *C. sublineola* and *C. graminicola*, (5) *C. sublineola* and *C. higginsianum*, (6) *C. graminicola* and *C. higginsianum*, (7) *C. sublineola*, *C. graminicola* and *C. higginsianum***.**



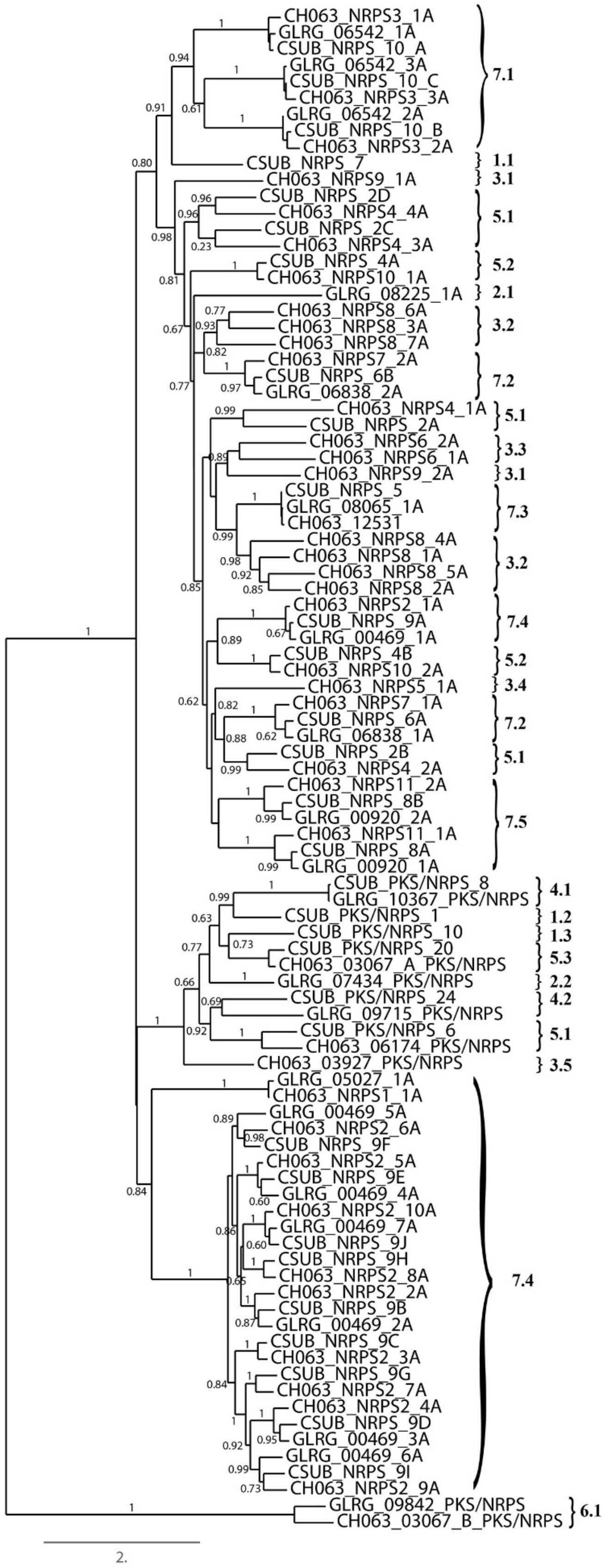
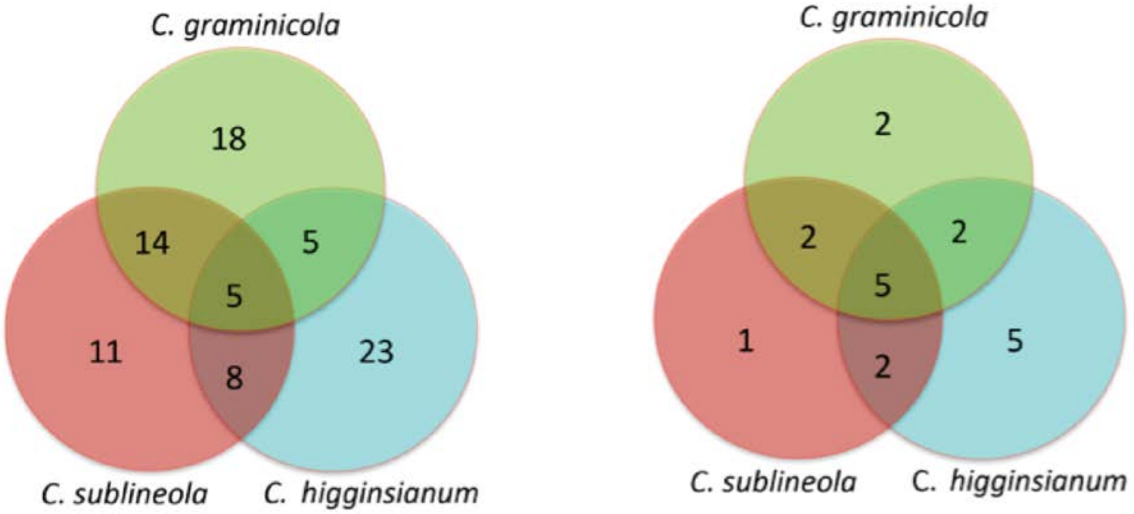
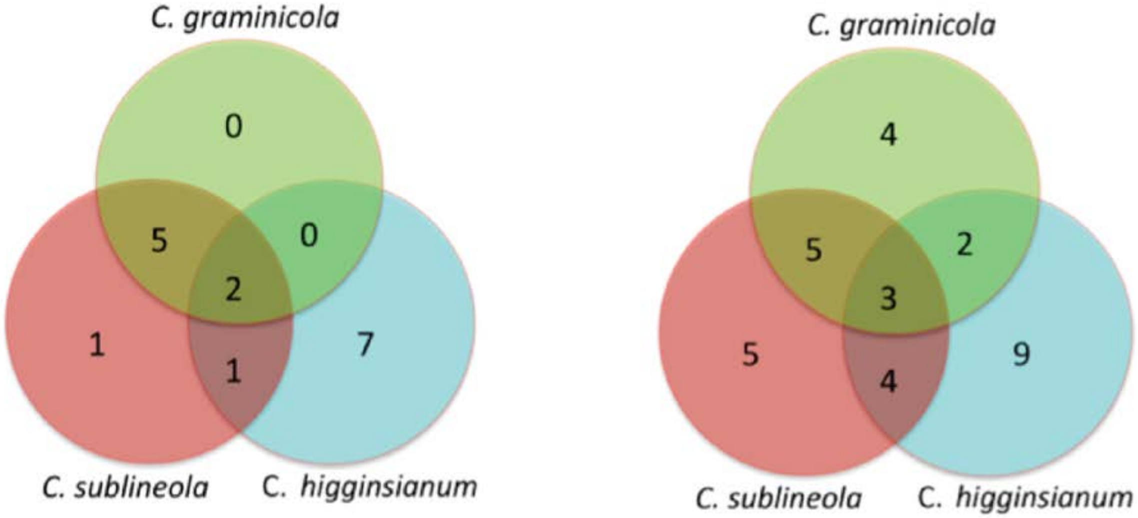


Figure 4.11. Phylogenetic tree of the AMP binding domains of putative NRPS and PKS-NRPS hybrids. Sequences were aligned by using MUSCLE version 3.7, and phylogenies were inferred by maximum-likelihood using PhyML version 3-0 Statistical. The numbers on the branch nodes indicate support values above 50%, calculated by an approximation to the standard Likelihood Ratio Test, aLRT. Brackets refer to proteins found only in (1) *C. sublineola*, (2) *C. graminicola*, (3) *C. higginsianum*; (4) *C. sublineola* and *C. graminicola*, (5) *C. sublineola* and *C. higginsianum*, (6) *C. graminicola* and *C. higginsianum*, (7) *C. sublineola*, *C. graminicola* and *C. higginsianum***.** The number after the decimal point refers to different domains of the same gene.

Figure 4.12. Venn diagram summarizing the numbers of shared and unique sequences among the three species.



PKS/Hybrid NRPS/Hybrid



DMAT TS

Figure 4.13. The organization of the conserved melanin gene clusters from *Colletotrichum sublineola, C. graminicola* and *C. orbiculare* are shown, with related genes depicted in the same color. The gray genes encode hypothetical proteins. Microsynteny is indicated by the gray bars.

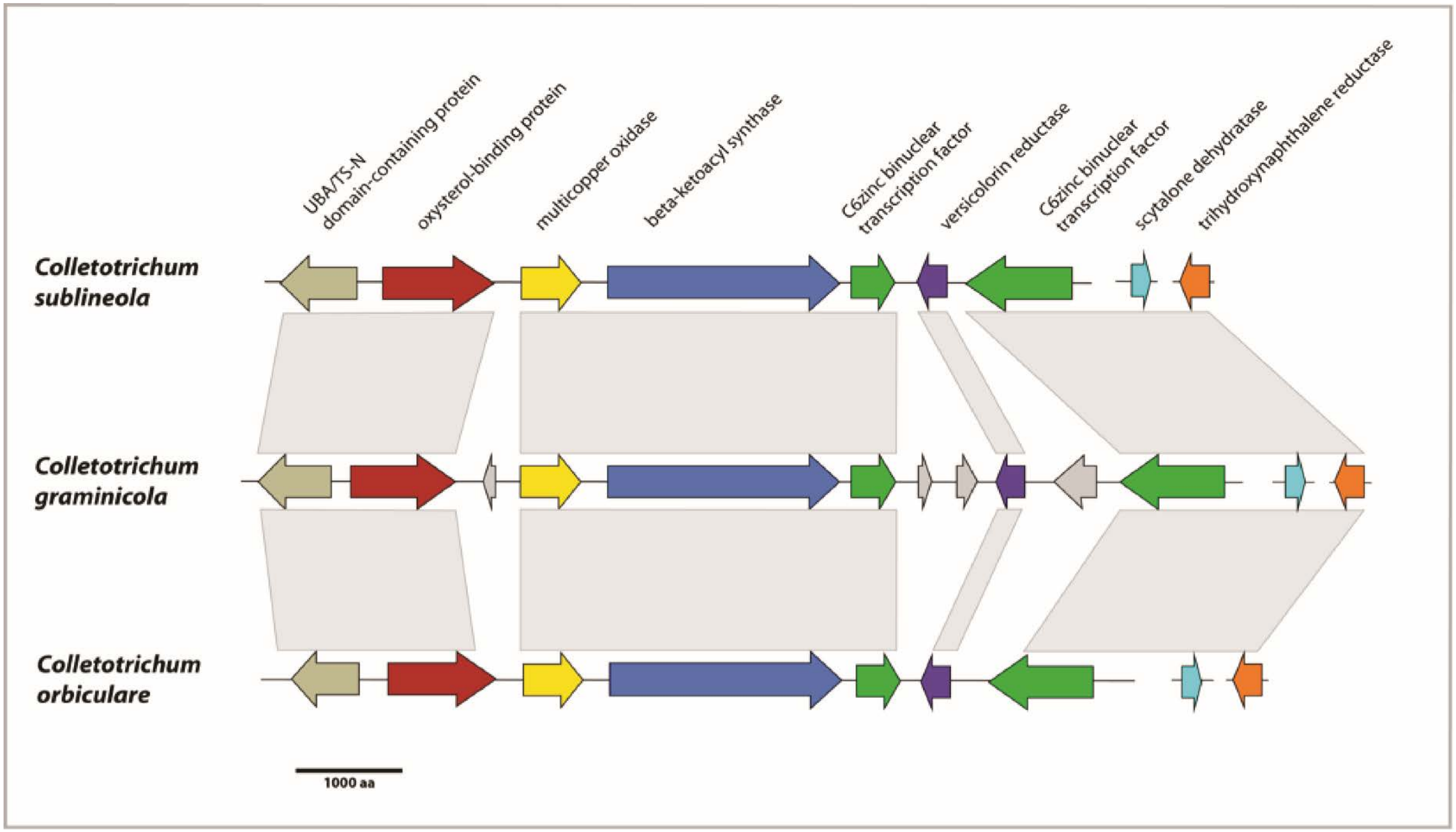


Figure 4.14. The organization of radicicol (RADS) gene clusters from *Pochonia chlamydospora*, *Colletotrichum graminicola and C. sublineola* are shown, with related genes depicted in the same color. Microsynteny is indicated by the gray bars.

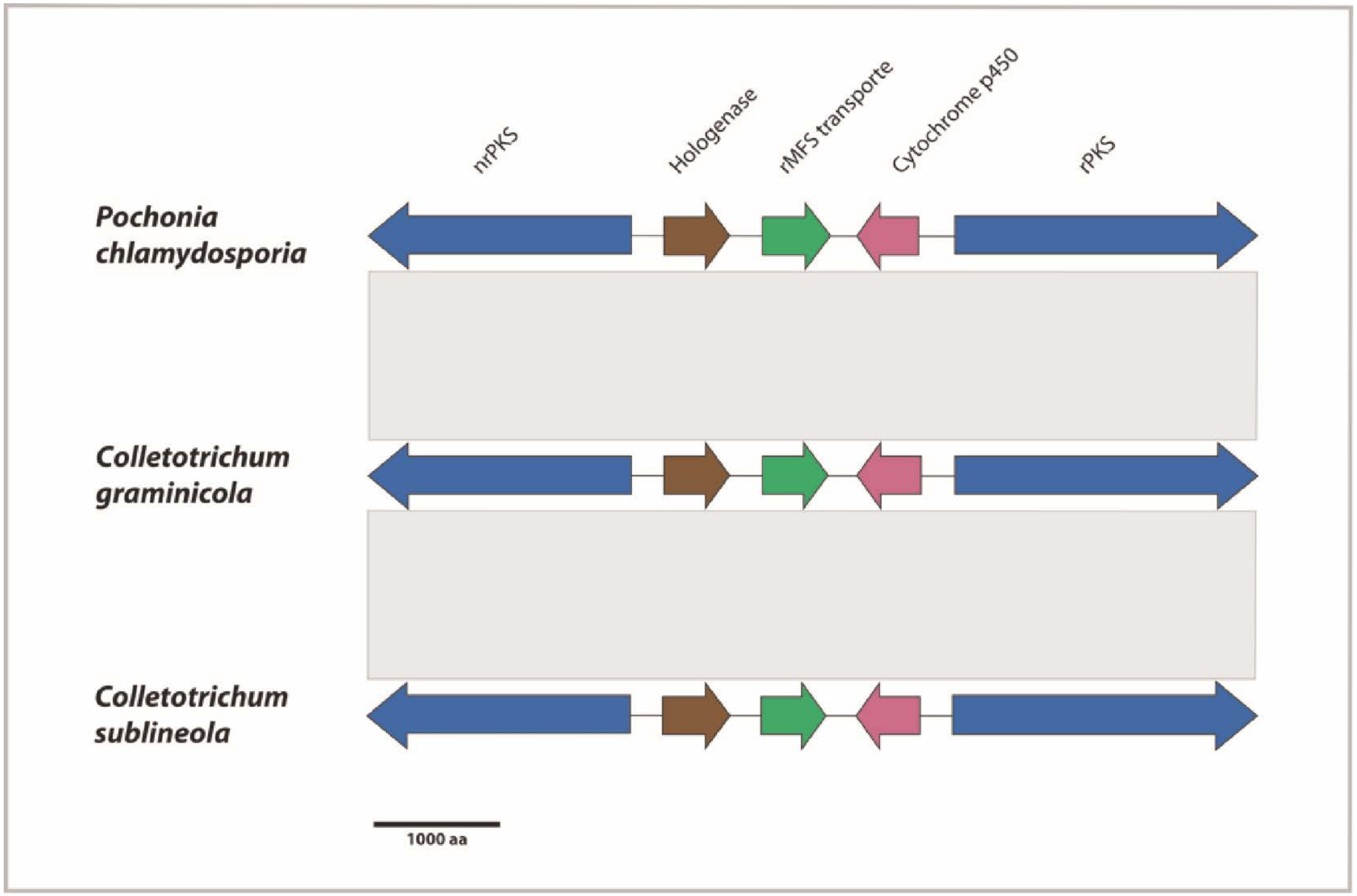


Figure 4.15. Diagrams of representative SM gene cluster (PKS 42) that appears to be unique to *C. sublineola*. Microsynteny is indicated by the gray bars.

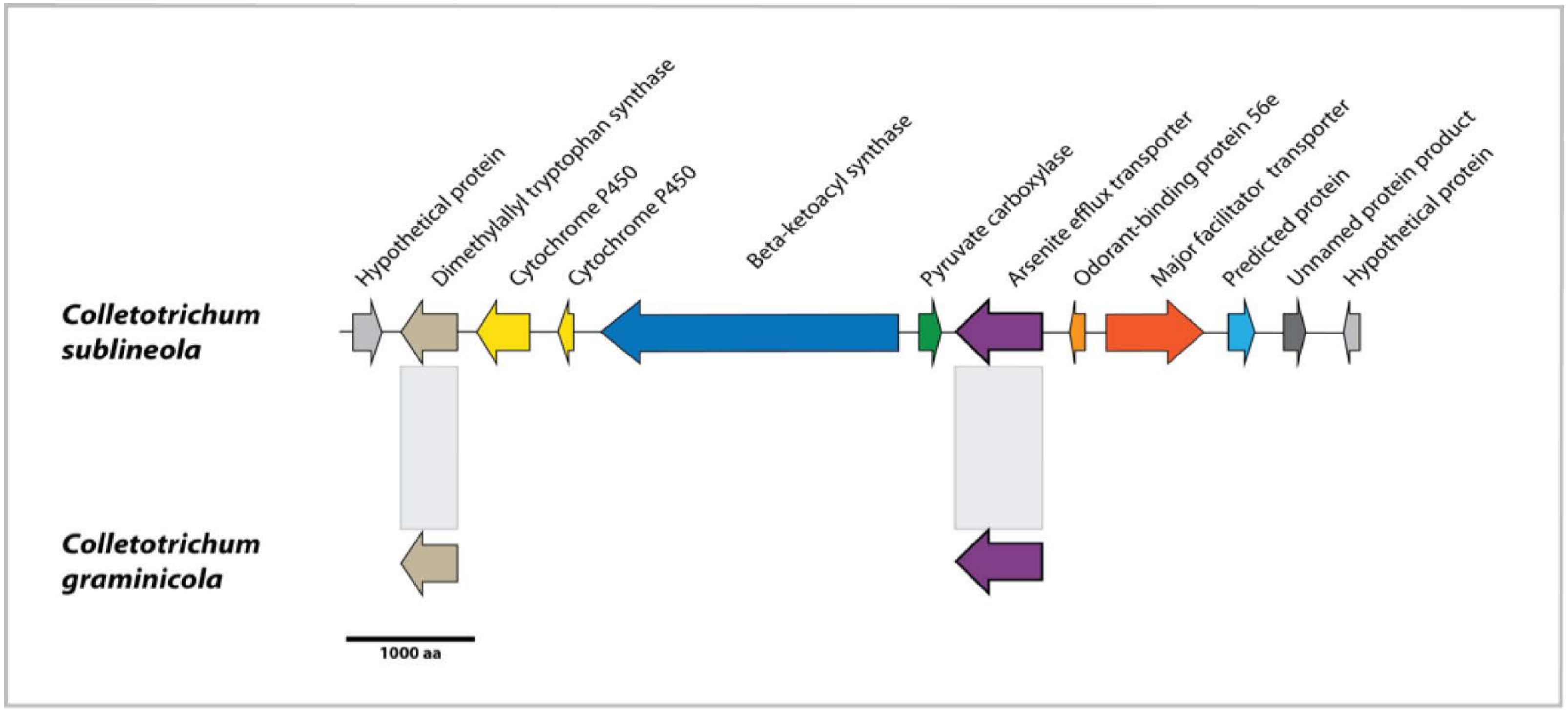


Figure 4.16. Diagrams of representative PKS gene cluster that appears to be present in *C. graminicola* but not in *C. sublineola.*

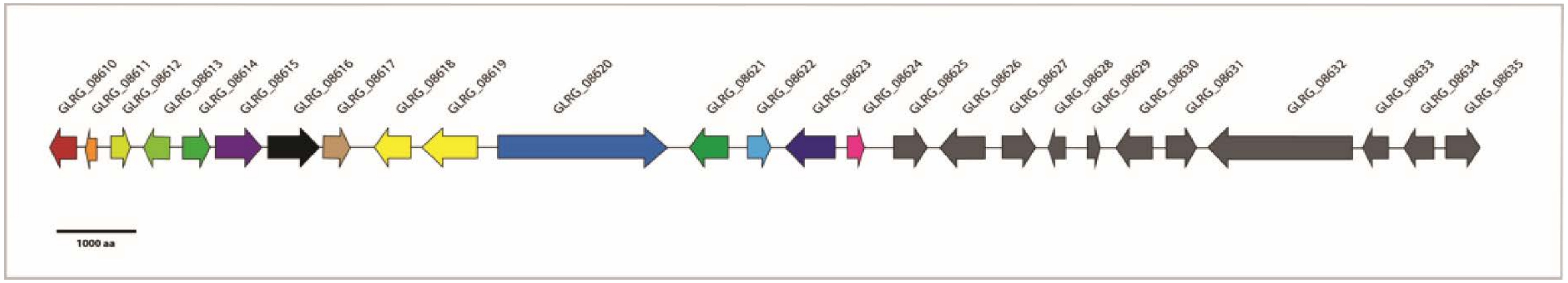


Figure 4.17. An example of a NRPS (NRPS 7) gene cluster unique *Colletotrichum sublineola* in comparison with *C. graminicola*, with related genes depicted in the same color. The gray genes encode hypothetical proteins. Microsynteny is indicated by gray area.

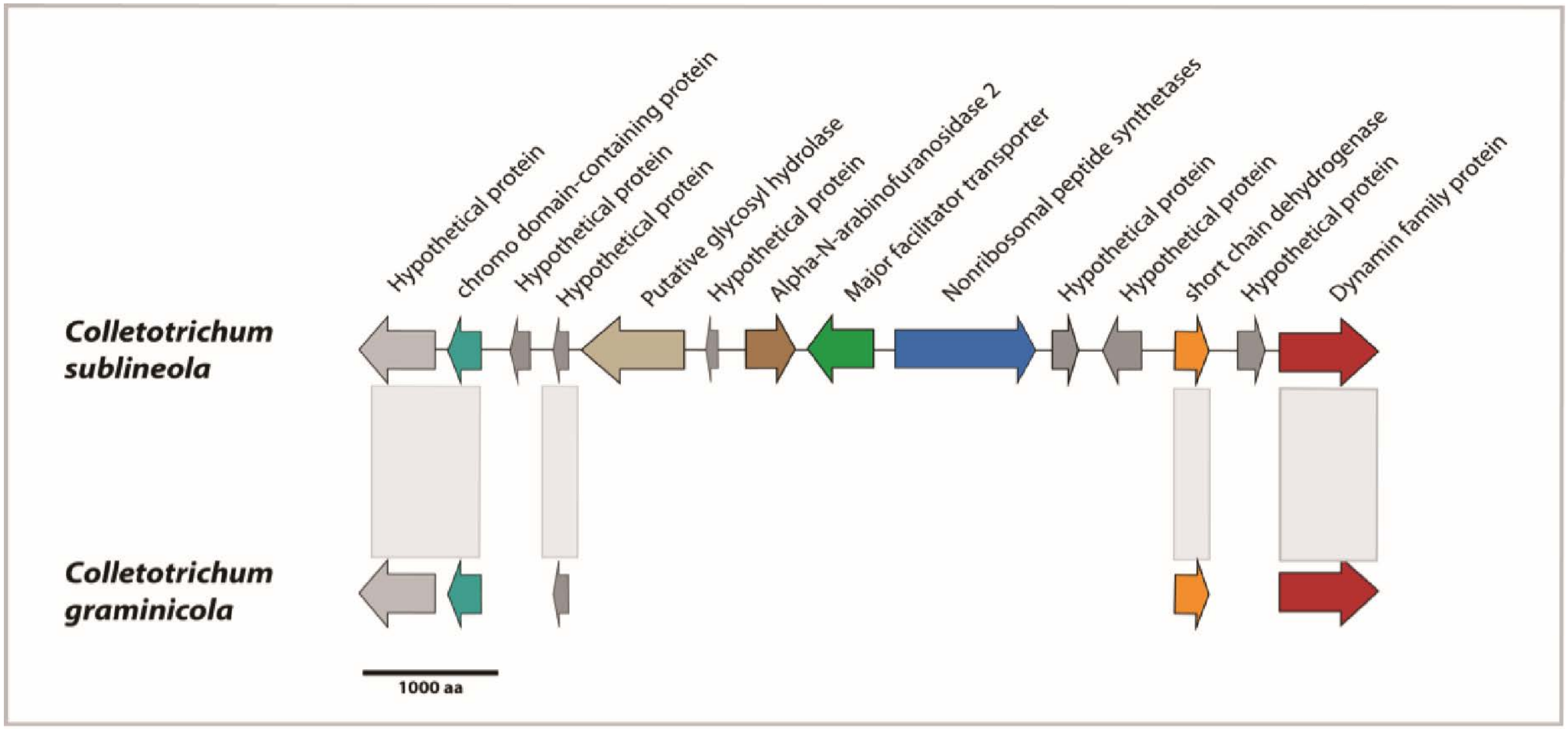


Figure 4.18. Southern blot analysis of genomic DNA digested with HindIII. *Colletotrichum sp.* strain name is indicated on each lane. Blot was hybridized with (A) CSE6, (B), CSE21, (C) CSSM3, and (D) actin probe control.

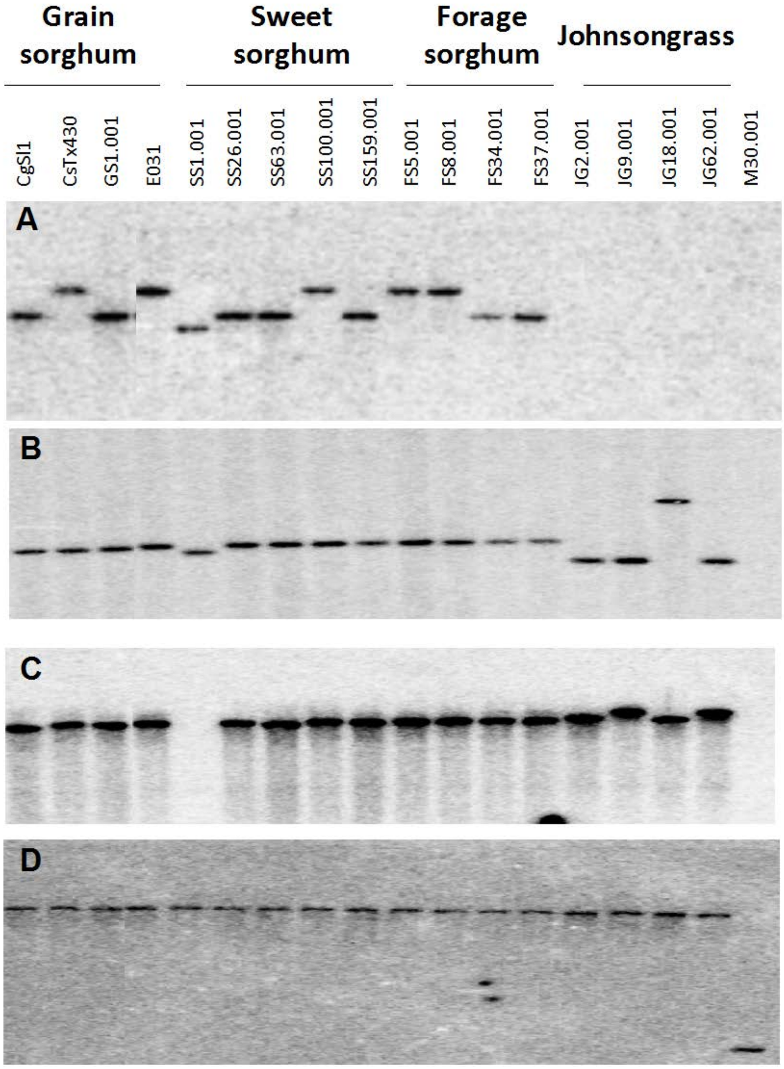
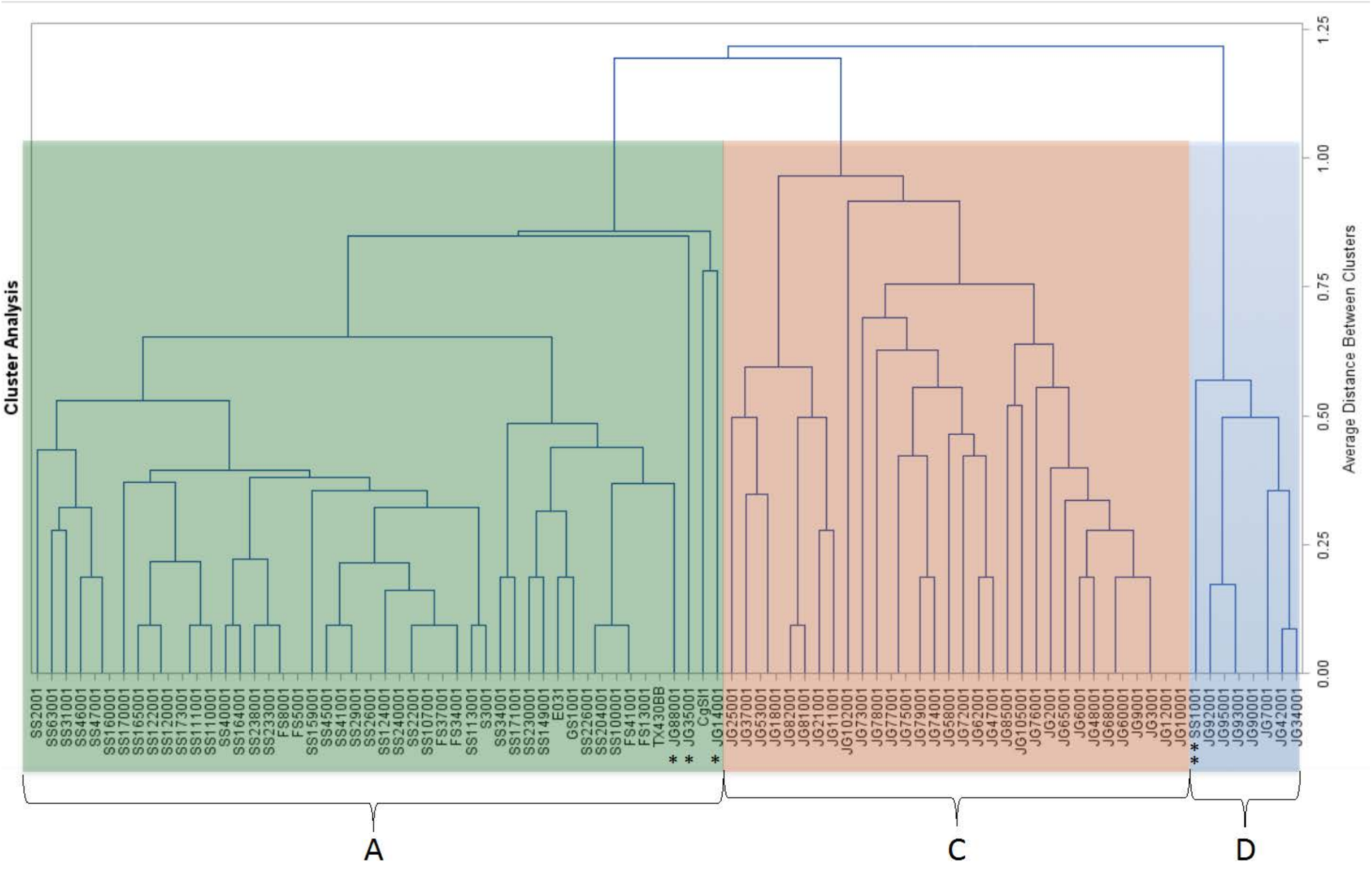


Figure 4.19: Dendrogram illustrating the relatedness of isolates from *Sorghum halepense* (JG) (blue box and orange box), and *S. bicolor* (green box), which includes isolates from grain (GS1.001, S3.001, CgSl1, and E031), sweet (SS) and forage (FS) sorghum, based on single locus probes using RFLP. There were some exceptions, marked with (*) isolates from *S. halepense* in the *S. bicolor*



Chapter 5 - Identification of two new species of *Colletotrichum* infecting *Sorghum bicolor* and *S. halepense* in the Southeastern United States

Abstract

Colletotrichum sublineola reportedly causes anthracnose disease on grain, forage and sweet sorghum (*Sorghum bicolor*), and also on the weed johnsongrass (*S. halepense*). My goal in this chapter was to analyze populations of *Colletotrichum* from cultivated and wild sorghum by evaluating individual genetic markers that are relevant to phylogenetic species determinations within this group of *Colletotrichum* fungi. Phylogenetic analysis was performed using 32 representative *Colletotrichum* strains that had been isolated from johnsongrass, and from forage, grain, and sweet sorghum. Phylogenetic trees were inferred based on portions of the DNA lyase gene (*Apn2*); the manganese superoxide dismutase gene (*Sod2*); and a region between the *Apn2* and the *Mat1* genes (*Mat1/Apn2*). The trees distinguished *C. sublineola* from two novel species. One of these new species was associated only with *S. halepense* throughout the southeastern U.S. Accordingly, I named this species *Colletotrichum halepense* (Xavier & Vaillancourt). There was evidence for cross-infection of *S. bicolor* and *S. halepense* by *C. sublineola* and by the second new species, which I named *C. caselae* (Xavier & Vaillancourt). These findings have significant implications for the development and deployment of resistant sweet sorghum varieties in areas where johnsongrass is common. This information will help to evaluate the potential for *Colletotrichum* spp. to cause epidemics in sweet sorghum if acreages continue to expand in the SE in the future.

Taxonomic novelties: Description of new species: *C. halepense* (Xavier & Vaillancourt) and *C. caselae* (Xavier & Vaillancourt).

Introduction

Anthracnose, caused by the fungus *Colletotrichum sublineola*, is the most important disease of sorghum (*Sorghum bicolor*) worldwide (3, 29, 71, 92, 193). The primary control for anthracnose in grain sorghum is the use of resistant cultivars, although no cultivar is immune to the disease (28). Furthermore, resistance frequently fails, which is thought to be related to the high degree of genetic variability in the population of *C. sublineola* affecting grain sorghum (25, 27).

Sweet sorghum has been grown in the Southeastern U.S. for more than 150 years, but it has been on a limited scale, primarily for forage and for the production of syrup. Recently, interest in sweet sorghum as a potential source of biofuel has increased in the U.S., especially in the Gulf States, where infrastructure used for the sugar cane industry could be converted to the production of ethanol from sorghum juice and biomass (102). Anthracnose is becoming common on the hybrids that are increasingly grown across the region (80). However, relatively little is known the populations of *Colletotrichum* that affect sweet sorghum in the region.

C. sublineola is also reported to be a pathogen of the weed johnsongrass (*Sorghum halepense*), which is related to *Sorghum bicolor* (194), and which is ubiquitous across the southeastern U.S., often found within and alongside fields of cultivated sorghum. Johnsongrass is known to be genetically highly diverse (120), and this diversity could result in balancing selection that maintains multiple pathotypes within the pathogen population.

Thus, johnsongrass could serve as a refuge and an incubator for genetic diversity, which could impact the durability of resistance genes deployed in nearby cultivated sorghum crops. Rosewich (1996) found that isolates from johnsongrass belonged to a population that was distinct from isolates from the grain sorghum varieties grown in nearby fields in Texas, suggesting that cross-infection did not occur in that location (144). My own analyses of genetic diversity using fingerprinting markers confirmed that most isolates from johnsongrass were genetically distinct from most isolates on cultivated sorghum in the southeastern U.S. I also found that most johnsongrass isolates were not aggressive to *S. bicolor* (chapter 3 of this dissertation). However, I did find evidence suggesting that a minority of isolates were capable of cross-infection of both hosts (chapter 3 of this dissertation). Analysis of allelic diversity at several putative pathogenicity loci showed that the populations of *Colletotrichum* from *S. bicolor* and from *S. halepense* could be separated into three distinct groups (chapter 4 of this dissertation). These groups were differentiated from one another by the presence or absence of specific alleles of several of these pathogenicity genes (aka apparent allelic fixation) (chapter 4 of this dissertation).

My goal in this final chapter of my dissertation was to test the hypothesis that the three groups identified in chapter 4 represent different species. Sequences from the *Sod2* gene, the *Apn2* gene, and from the intergenic region between *Apn2* and *Mat1*, have been used previously to identify new species within the *C. graminicola* species complex that affects graminaceous hosts (40). To test my hypothesis, I analyzed these sequences, as well as measuring various morphological traits in culture, for several representatives of each of the three groups of *Colletotrichum* isolates.

Material and Methods

Fungal Strains: A total of 32 isolates from sweet, forage, and grain sorghum (*S. bicolor*), and from johnsongrass (*S. halapense*), from a variety of locations were used for this study (Table 5.1). The strains were chosen to represent the genetic diversity that was revealed by the fingerprinting, pathogenicity, and RFLP analyses (chapters 3 and 4 of this dissertation). Isolates were purified by single-sporing, and preserved on silica gel at -80°C (179). Fungal strains were routinely cultured on PDA at 23°C under continuous fluorescent light.

Multigene Phylogenetic Analysis: A multigene phylogenetic analysis of the 32 representative strains of *Colletotrichum* was performed. The primer pair ITS4 and ITS5 was used to amplify the internal transcribed spacer (ITS) sequence of the rDNA repeats (200). The primer pair SOD507F and SOD507F/R was used to amplify a 533bp fragment of the manganese superoxide dismutase gene *Sod2* (39). In order to amplify the DNA lyase gene (*Apn2*) the primers Apn1W1F and Apn1W1R were used, generating a 799bp fragment (40). Primers MAT1/APN2F and MAT1/APN2R were used in an attempt to amplify the region between the *Apn2* gene and the *Mat1* gene (*Mat1/Apn2*) (40). However, these primers failed to amplify a product from any of the *Colletotrichum* strains used in my study. The *Mat1/Apn2* sequence from *C. graminicola* (accession number FJ377991) was then used to search for similar sequences in the genome of the CgSl1 strain of *C. sublineola* by blastn. A matching sequence from *C. sublineola* was retrieved, and used to design a new *Mat1/Apn2* primer pair (MAT1/APN2F and MAT1/APN2R), which amplified a product of 860bp from all 32 strains. Primer sequences are shown in Table 5.2.

PCR reactions were performed in a final volume of 25µL containing: 25 ng of genomic DNA, 10X PCR buffer, 50 mM $MgCl_2$, 200 µM dNTPs, one U of *Taq* DNA polymerase, and 5 µM of each primer. The PCR cycling conditions were as follows: initial denaturation at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 1 min, annealing for 30 sec at (i) 53°C for *Sod2*, (ii) 51°C for ITS, (iii) 55°C for *Mat1/Apn2*, and (iv) 55°C for *Apn2*; extension at 72°C for 1 min; with a final extension at 72°C for 10 min.

The amplified fragments were visualized on 0.7% agarose gels, and then excised and purified by using the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). The purified PCR amplicons were submitted to the University of Kentucky Advanced Genetic Technologies Center (UK-AGTC) for sequencing. Forward and reverse primers were used to sequence the amplicons with the BigDye Terminator cycle-sequencing chemistry (Applied Biosystems, Foster City, CA) on an ABI 3730 DNA Capillary Sequencer following the manufacturer's protocol, except in a reaction volume of 10 µL.

I also included sequences from the *C. subineola* epitype strain CBS131301, aka S3.001, as a control. *Colletotrichum eremochloae*, the closest known relative of *C. sublineola* (41), was included as an outgroup in all the trees. The ITS, *Sod2*, *Apn2*, and *Mat1/Apn2* sequences of these strains were identified by blastn searches of the genome databases that are available from the Joint Genome Institute Genome Portal (http://genome.jgi.doe.gov).

The sequences were aligned by using MUSCLE version 3.7. Alignments were checked in Geneious Pro (version 5.4.7) and Mesquite (version 3.1), followed by trimming and manual editing and optimization of the alignments as required. Phylogenetic analysis was performed by using phylogeny.fr (http://wwwphylogeny.fr/) with default parameters. Phylogenies were inferred by maximum-likelihood using PhyML version 3.0. Statistical branch support was provided by an approximation to the standard Likelihood Ratio Test-aLRT (52).

Morphological analysis: Fungal cultures were grown on PDA at 23°C under continuous fluorescent light for 14 days. Fungal spores were harvested from the PDA cultures and washed three times with sterile water. A conidial suspension was filtered through a layer of cheesecloth and resuspended in water. The spores were mounted on microscope slides in distilled H_2O. For the production of appressoria, spores were germinated in drops of deionized water on plastic cover slips in a moist chamber overnight. Hyphopodial appressoria were produced by overlaying small blocks of carrot agar containing mycelia with glass cover slips. The mycelia and cover slips were then placed in Petri dishes containing filter paper moistened with distilled sterile H_2O. The Petri dishes were sealed with Parafilm and incubated at 23°C under continuous fluorescent light for 4-7 days. The cover slips were removed from the agar and then placed onto drops of water on fresh microscope slides. Lengths and widths of 50 randomly chosen conidia, appressoria, or hyphopodia per strain were measured on a Zeiss Axioscop at 400x magnification with the measurements module of the Openlab computer program.

Results

Multigene sequencing

Four sequences (ITS, *Sod2*, *Mat1/Apn2* and *Apn2*) were amplified from 32 isolates, sequenced, and aligned (Alignments are provided in Appendix 4 of this dissertation). Two clades were well-separated by the ITS sequences (Figure 5.1). One of the clades contained the epitype for *C. sublineola*, S3.001, while the other included five isolates that had previously been grouped together based on fingerprints, belonging to the “blue group” identified by RFLP analysis of putative pathogenicity genes (chapters 3 and 4 of this dissertation). Strain CgSl1 belonged to the same clade as S3.001, but it was located on a long branch due to the presence of a SNP. The closely related species *C. eremochloae* was not separated from *C. sublineola* in the ITS tree.

A phylogenetic tree constructed by using an alignment of the concatenated *Sod2*, *Apn2*, and *Mat1/Apn2* sequences resolved three major clades (Figure 5.2). The alignment of the concatenated gene sequences included 2018bp of which 1917bp were phylogenetically informative. One clade contained the epitype S3.001, and so I concluded that members of this clade belong to the species *C. sublineola*. S3.001 differs all from the rest of the members of that group by one SNP. The strain CgSl1 differed from the rest by the presence of a 15bp indel, representing one copy of a sequence that is repeated twice. All the other strains in the clade had only one copy of this sequence. Two other clades were consistent with the two other groups (the “orange” and the “blue” groups) previously identified based on RFLP analysis (see chapter 4 of this dissertation).

Separate trees constructed by using each of three gene sequences (*Sod2*, *Apn2*, and *Mat1/Apn2*) all grouped the isolates into the same three major clades with a high degree of confidence (Figure 5.3-5.5). However, within the largest clade I observed reticulation, with the positions of CgSl1, GS1.001, SS63.001, and S3.001 varying relative to one another and to the other members of the clade in each tree. This indicates the possibility of recombination among the genes, and it suggests that this large clade (highlighted in green) represents a single species, according to the multilocus genealogical concordance method described by (169). Because this clade includes the epitype S3.001, I conclude that it represents the previously described species *C. sublineola*. The clade included isolates from both *S. bicolor* and *S. halepense*, confirming that *C. sublineola* has the ability to infect both. The other two clades did not exhibit reticulation and were consistent in all three trees. One of these clades (highlighted in orange) included only strains isolated from johnsongrass, while the other, highlighted in blue, included strains from both hosts. The strains in this blue clade were also separated from the rest by the ITS sequence (Figure 5.1). The three clades were consistent with the pathogenicity data, and with the three groups identified by the results of RFLP analysis using single-gene probes corresponding to putative pathogenicity genes (chapters 3 and 4 of this dissertation).

Morphology

Measurements were made of spores, hyphopodia, and appressoria for representative isolates from each of the phylogenetic groups (Figures 5.6-5.8). The average lengths and widths of conidia, appressoria, and hyphopodia and the ranges, are shown in table 5.3. The ranges overlapped broadly among the three groups, and the averages were similar (Table 5.3).

Colony morphology (upper surface) varied among the different isolates but most produced salmon spores masses in the center of the colony, surrounded by white-to-gray mycelial growth (Figures 5.6-5.8). The lower surfaces also varied among the isolates, with color ranging from orange to black. Isolates produced sclerotia on the bottom of the plates. Isolates in the orange clade produced more aerial mycelia than the other groups (Figures 5.6-5.8).

Spores were similar in shape for the three groups, ranging from mostly falcate to fusiform. Appressoria were also similar, ranging from nearly circular to oval. The hyphopodia were more irregular and lobed. Overall, strains within the group varied quite a lot in appearance and there was no consistent character that could be reliably used to differentiate the clades based on morphology alone.

Taxonomy

I propose that these 32 isolates belong to three different *Colletotrichum* species, based on fingerprinting, RFLP, pathogenicity, and DNA sequence data. I propose that two of these (the orange and blue clades) are new species, which I have named *Colletotrichum halepense*, associated only with the weed *S. halepense*, and *C. caselae* (in honor of Dr. Carlos Casela, retired from EMBRAPA Milho e Sorgo), pathogenic to both *S. halepense* and *S. bicolor*. The third group (green clade) conforms to the canonical *C. sublineola* species, based on gene sequence similarity with the epitype for the species S3.001. It should be noted that the *Sod2* and *Mat1/Apn2* sequences of S3.001 that are deposited in Genbank (42) do not match the genome sequence from the same strain at JGI. I confirmed the JGI sequences (CBS131301 v1.0/581817 and CBS131301 v1.0/ 572845, respectively) by personally sequencing *Sod2* and *Mat1/Apn2* in this isolate, which we have in our laboratory strain collection. Alignments of the S3001 sequences from JGI, from GenBank, and from the strain in our lab are shown in figure 5.9A-B. It appears that there are some sequencing errors in the Genbank accessions.

I confirmed that *Colletotrichum sublineola* is pathogenic to both *S. bicolor* and *S. halepense*. (39). I also confirmed that *C. caselae* (represented by SS1.001) is also pathogenic to both *Sorghum* spp. In contrast to the other two species, *C. halepense* appears to be confined to a single host, *S. halepense*.

Colletotrichum halepense K. V. Xavier & L. J. Vaillancourt, sp. nov. Figures 5.6 (A- P).

The colonies vary in appearance on potato dextrose agar medium, from pale yellowish to dark gray with black sclerotia visible from the lower surface. Colonies produce abundant aerial mycelium. The salmon spore masses are restricted to the center of the colonies

(Figure 5.6A-C). The conidial shape for *C. halepense* is falcate or fusiform, apices acute, 24.15-42.11 (av. 28.75 μm) by 4-6 μm (av. 5.44 μm). Hyphopodial appressoria round, globose to lobate, apices cylindrical to lobate, 8.54-26.47 (av. 13.90 μm) by 7.6-1- 24.92 μm (av. 10.83 μm). Appressoria are round to irregular, 4.61-12.54 μm (av. 8.56 μm) by 5.08-11.12 (av. 6.66 μm).

Morphologically similar to other species of *Colletotrichum* isolated from gramineaceous hosts, but differs based on molecular phylogenetic analyses of *Sod2, Mat1/Apn2* and *Apn2* genes, and unique association with the host *Sorghum halepense*. Type strain JG18.001 from Russell County KY.

<u>*Colletotrichum caselae*</u> K. V. Xavier & L. J. Vaillancourt, sp. nov. Figures 5.7 (A- P).

The colonies vary from orange to dark gray in color on potato dextrose agar medium from the lower surface. Salmon spore masses were produced in the centers of the colony, surrounded by gray mycelia. Alternatively, producing mycelia all over the plate, with no conidial masses visible.

The conidial shape for *C. caselae* is falcate or fusiform, apices acute, 23.08-37.5 (av. 28.72 μm) by 4.5-7.02 μm (av. 5.69 μm). Hyphopodial appressoria are round, globose to lobate, apices cylindrical to lobate, or flattened, 9.47-19.65 (av. 14.01 μm) by 6.25-13.12 μm (av. 9.73 μm). Appressoria are round to irregular, 5,4-10.11 (av. 7.87 μm) by 4.2-8.62 μm (av. 6.04 μm).

Morphologically similar to other species of *Colletotrichum* isolated from gramineaceous hosts but differs based on molecular phylogenetic analyses of ITS, *Sod2*, *Mat1/Apn2* and *Apn2* genes. *Colletotrichum caselae* is associated with the hosts *Sorghum halepense* and *Sorghum bicolor*. Type strain SS1.001 from Fayette County, Kentucky.

Discussion

In this chapter, I tested the prediction that three genetically distinct groups that I had observed within the populations of *Colletotrichum* affecting cultivated sorghum and johnsongrass actually represent different species. Sequences from the *Sod2* gene, the *Apn2* gene, and from the intergenic region between *Apn2* and *Mat1*, have been used previously to identify new species within the *C. graminicola* species complex that affects graminaceous hosts (40). To test my prediction, I analyzed these sequences together with the internal transcribed spacer (ITS) of the ribosomal DNA, as well as measuring various morphological traits in culture, for several representative *Colletotrichum* isolates.

Morphological characters are usually not sufficient to distinguish among closely related *Colletotrichum* species. For example, based on morphology *C. sublineola*, infecting *S. bicolor*, was erroneously called *C. graminicola* up until relatively recently. Molecular phylogenetic analyses finally established that these were two sibling species with specific host ranges (39, 40, 61, 157, 165, 167, 181). Colony appearance on PDA and the shapes and sizes of conidia, appressoria and hyphopodia among the three genetically distinct groups I identified were all very similar. There was no consistent character that distinguished the three groups morphologically.

However, results from the multigene phylogenetic analyses strongly supported the hypothesis that *Colletotrichum* isolates that infect *Sorghum* spp. in the southeastern U.S. belong to three different, closely related species. One is the previously described *C. sublineola*, and the other two are undescribed. There was evidence for reticulation among individual gene trees within *C. sublineola*, suggesting the possibility of recombination. A telemorph has been described for *C. sublineola* (181), although it has not been observed in nature. The *C. sublineola* clade included strains from both *S. bicolor* and *S. halepense*. One of the two new species appears to be limited to johnsongrass, while the other, like *C. sublineola*, apparently has the ability to infect either *S. halepense* or *S. bicolor* (see chapter 3 of this dissertation). I have named these new species *C. halapense*, and *C. caselae*, respectively. The ability of *C. sublineola* and *C. caselae* to infect both *Sorghum* species has significant implications for sorghum breeding and anthracnose disease management in regions where johnsongrass is common. The presence of three presumably genetically isolated species also has implications for gene flow, including of pathogenicity genes.

Table 5.1. Strains of *Colletotrichum sublineola*, *C. halepense* and *C. caselea* studied in this chapter, with collection details.

Species	Strain	Host	Country	State	County	Year
C. sublineola	CgSl1	Grain sorghum	United States	Indiana	Tippecanoe	1975
	E031	Grain sorghum	Brazil	Minas Gerais	Indianapolis	2009
	FS13.001	Forage sorghum	United States	Alabama	Brewton	2013
	FS34.001	Forage sorghum	United States	Alabama	Macon	2013
	FS37.001	Forage sorghum	United States	Alabama	Macon	2013
	FS5.001	Forage sorghum	United States	Alabama	Brewton	2013
	FS8.001	Forage sorghum	United States	Alabama	Brewton	2013
	GS1.001	Grain sorghum	United States	Kentucky	Hopkins	2013
	JG88.001	johnsongrass	United States	Kentucky	Montgomery	2014
	SS100.001	Sweet sorghum	United States	Alabama	Macon	2013
	SS107.001	Sweet sorghum	United States	Georgia	Tift County	2013
	SS124.001	Sweet sorghum	United States	Georgia	Tift County	2013
	SS159.001	Sweet sorghum	United States	Georgia	Tift County	2013
	SS171.001	Sweet sorghum	United States	Florida	Live Oak	2014
	SS204.001	Sweet sorghum	United States	Kentucky	Montgomery	2014
	SS26.001	Sweet sorghum	United States	Florida	Live Oak	2013
	SS29.001	Sweet sorghum	United States	Florida	Live Oak	2013
	SS41.001	Sweet sorghum	United States	Florida	Live Oak	2013
	SS45.001	Sweet sorghum	United States	Florida	Live Oak	2013
	SS63.001	Sweet sorghum	United States	Florida	Live Oak	2013

Table 5.1. (continued)

Species	Strain	Host	Country	State	County	Year
C. sublineola	SS63.001	Sweet sorghum	United States	Florida	Live Oak	2013
	SS78.001	Sweet sorghum	United States	Alabama	Brewton	2013
	SS240.001	Sweet sorghum	United States	Georgia	Tift County	2014
C. caselae	JG34.001	johnsongrass	United States	Kentucky	Frankfort	2013
	JG42.001	johnsongrass	United States	Kentucky	Fayette	2013
	JG7.001	johnsongrass	United States	Kentucky	Russell	2011
	SS1.001	Sweet sorghum	United States	Kentucky	Fayette	2012
C. halepense	JG102.001	Sweet sorghum	United States	Georgia	Tift County	2014
	JG18.001	johnsongrass	United States	Kentucky	Fayette	2012
	JG2.001	johnsongrass	United States	Kentucky	Russell	2011
	JG62.001	johnsongrass	United States	Kentucky	Russell	2013
	JG75.001	johnsongrass	United States	Florida	Gadsden	2014
	JG9.001	johnsongrass	United States	Kentucky	Russell	2012

Table 5.2. Primers used for PCR amplification and DNA sequencing of specific genes

Gene	Primer	Sequences of Primers	Reference
ITS	ITS5	5′ GGAAGTAAAAGTCGTAACAAGG 3′	White et al., 1990
	ITS4	5′ TCCTCCGCTTATTGATATGC 3′	White et al., 1990
SOD2	SOD507F	5′ ATGGCAGCCTTTCCGTTGAGATAC 3′	Crouch et al, 2006
	SOD507F/R	5′ AGTTGACATGAAGCCACCTACAGC 3′	Crouch et al, 2006
APN2	Apn1W1F	5′ ATGGAGCACAAAAACGAACA 3′	Crouch et al, 2009b
	Apn1W1R	5′ GCGGAGCAGAGGATGTAGTC 3′	Crouch et al, 2009b
MAT/APN	MAT1/APN2F	5' CCCGAGTTCGACCGGAGACG 3'	accession #
	MAT1/APN2F	5' TTCCGAGAGGCCGGGGACAA 3'	accession #

Table 5.3. Characteristics of the spores, appressoria, and hyphopodia of the three phylogenetic groups.

Species		Spore		Hyphopodium		Appressorium	
		Length	Width	Length	Width	Length	Width
Green Clade (C. sublineola)	Range	19.82-36.78	4.11-7.68	7.03-20.85	6.57-16.76	6.44-11.64	4.85-9.02
	Average	28.10	5.54	12.34	10.00	8.57	6.70
Orange Clade	Range	24.15-42.11	4-6.65	8.54-26.47	7.6-24.92	4.61-12.54	5.08-11.12
	Average	28.75	5.44	13.90	10.83	8.56	6.66
Blue Clade	Range	23.08-37.5	4.5-7.02	9.47-19.65	6.25-13.12	5.4-10.11	4.2-8.64
	Average	28.72	5.69	14.01	9.73	7.87	6.04

Figure 5.1. Phylogenies of *C. sublineola* isolates inferred from ITS gene. Isolates are labeled as johnsongrass (JG), grain (GS1001, E031, CgSl1), sweet (SS) and forage (FS) sorghum isolates. Clades are highlighted in green and blue. Outgroup isolate, *C. eremochloae* is highlighted in yellow. Isolates highlighted in orange belong to the "orange" group of isolates identified by RFLP analysis in chapter 4 of this dissertation.

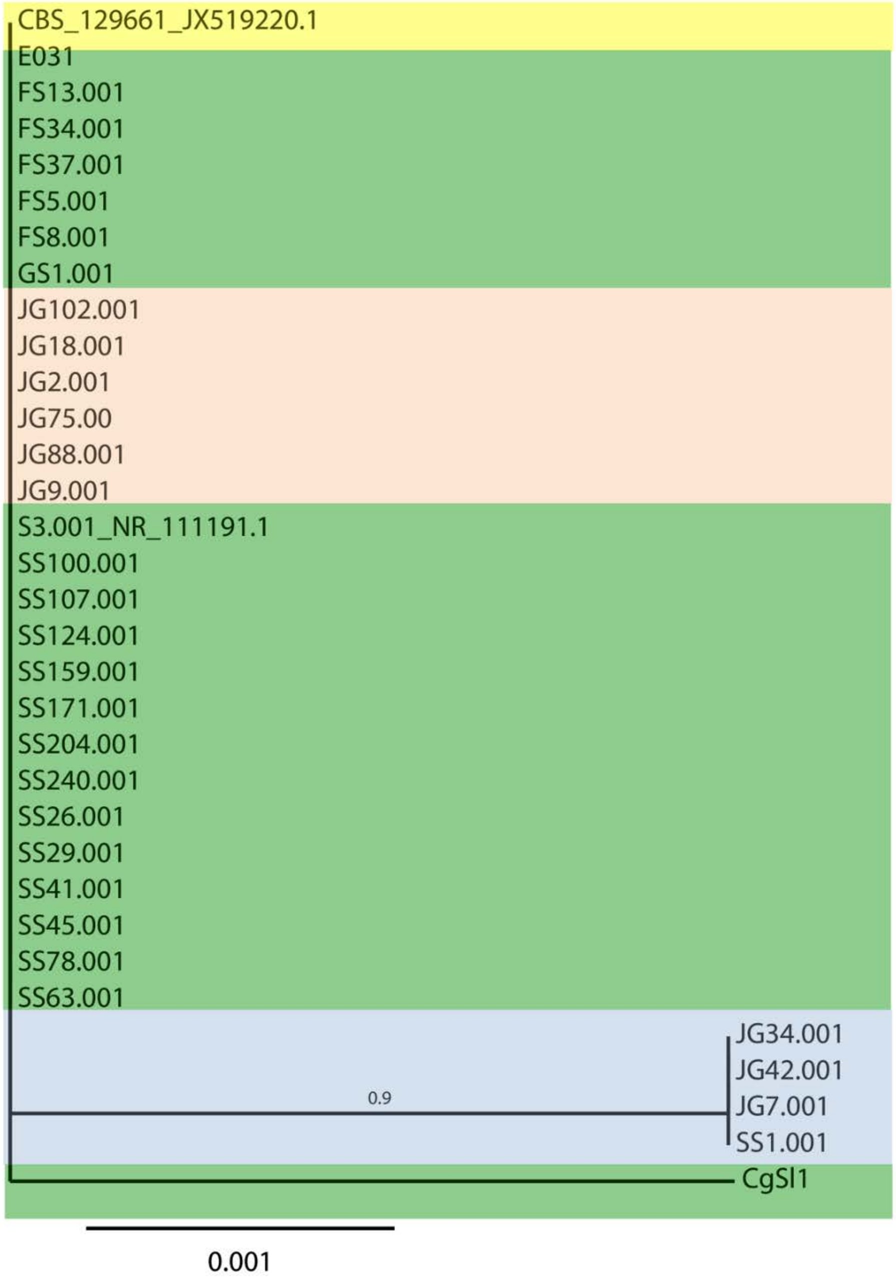


Figure 5.2. Phylogenies of *C. sublineola* isolates inferred from three concatenated gene sequences (*Sod2*, *Apn2*, and *Mat1/Apn2*). Isolates are labeled as johnsongrass (JG), grain (GS1001, E031, CgSl1), sweet (SS) and forage (FS) sorghum isolates. Clades are highlighted in green, orange, and blue. Outgroup isolate, *C. eremochloae* is highlighted in yellow.

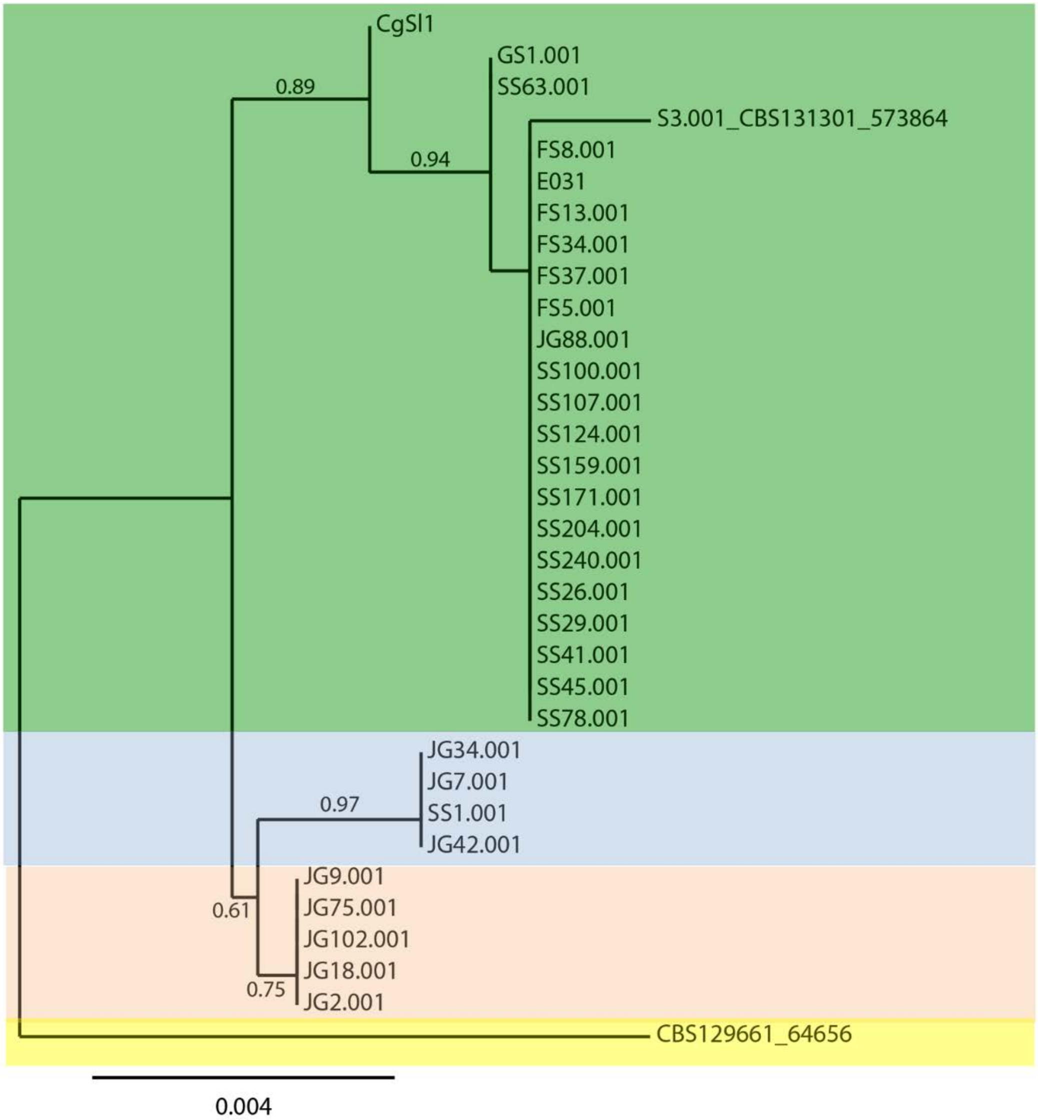


Figure 5.3. Phylogenies of *C. sublineola* isolates inferred from the *Apn2* gene. Isolates are labeled as johnsongrass (JG), grain (GS1001, E031, CgSl1), sweet (SS) and forage (FS) sorghum isolates. Clades are highlighted in green, orange, and blue. Outgroup isolate, *C. eremochloae* is highlighted in yellow.

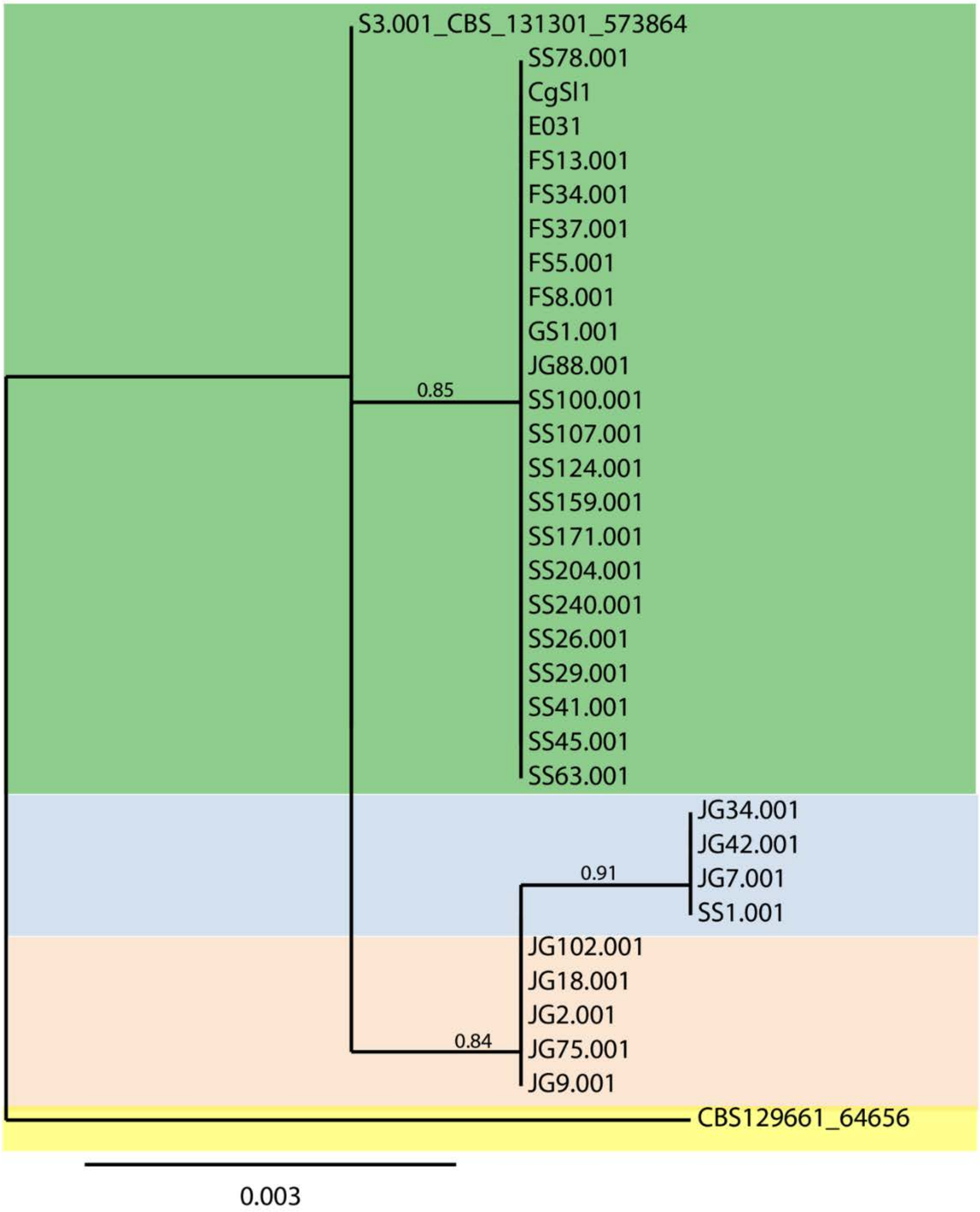


Figure 5.4. Phylogenies of *C. sublineola* isolates inferred from the *Mat1/Apn2* sequence. Isolates are labeled as johnsongrass (JG), grain (GS1001, E031, CgSl1), sweet (SS) and forage (FS) sorghum isolates. Clades are highlighted in green, orange, and blue. Outgroup isolate, *C. eremochloae* is highlighted in yellow.

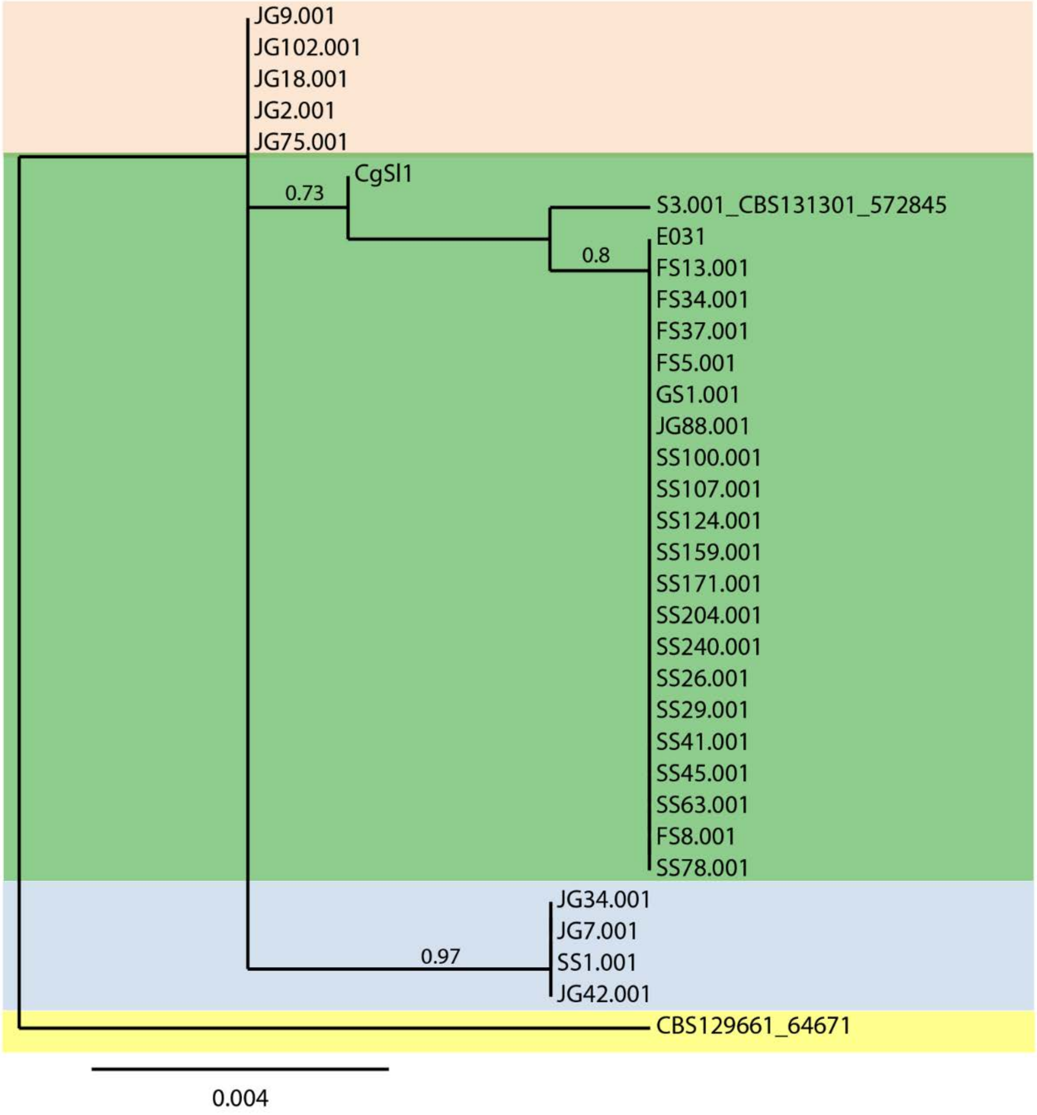


Figure 5.5. Phylogenies of *C. sublineola* isolates inferred from the *Sod2* gene. Isolates are labeled as johnsongrass (JG), grain (GS1001, E031, CgSl1), sweet (SS) and forage (FS) sorghum isolates. Clades are highlighted in green, orange, and blue. Outgroup isolate, *C. eremochloae* is highlighted in yellow.

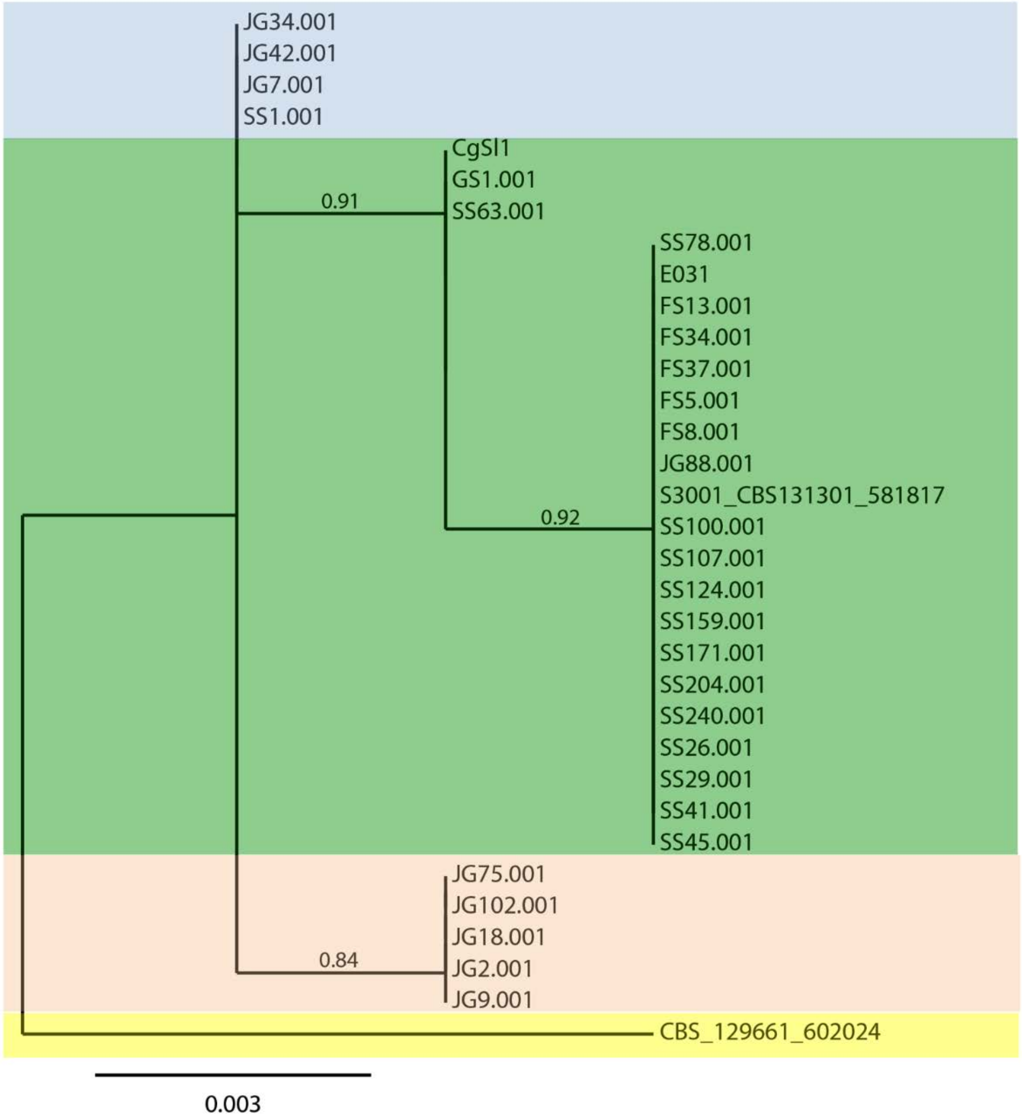


Figure 5. 6. Morphological characters, Orange Clade (*Colletotrichum halepense*) (A- D) Cultures on PDA, 15 d growth, from above and below; (E–H) Hyphopodia; (I- L) Appressoria; and (M- P) Conidia. Scale bars = 20 μm. Isolates (A, E, I, M) JG2.001, (B, F, J, N) JG9.001, (C, G, K, O) JG75.001, (D, H, L, P).

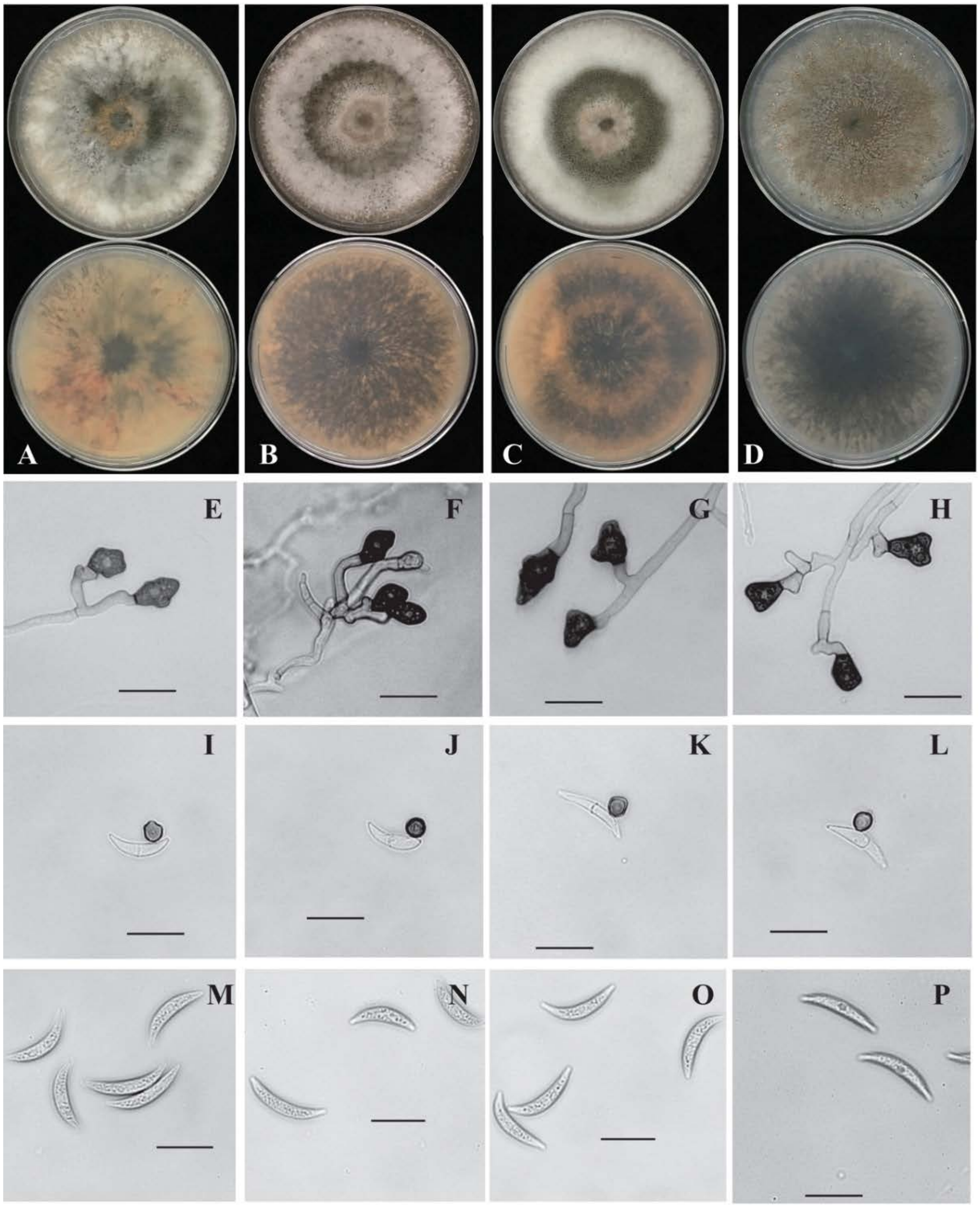


Figure 5.7. Morphological features, Blue Clade, (*Colletotrichum caselae*). (A- D) Cultures on PDA, 15 d growth, from above and below; (E–H) Hyphopodia; (I- L) Appressoria; and (M- P) Conidia. Scale bars = 20 µm. Isolates (A, E, I, M) JG34.00, (B, F, J, N) SS1.001, (C, G, K, O) JG7.001, (D, H, L, P) JG42.001.

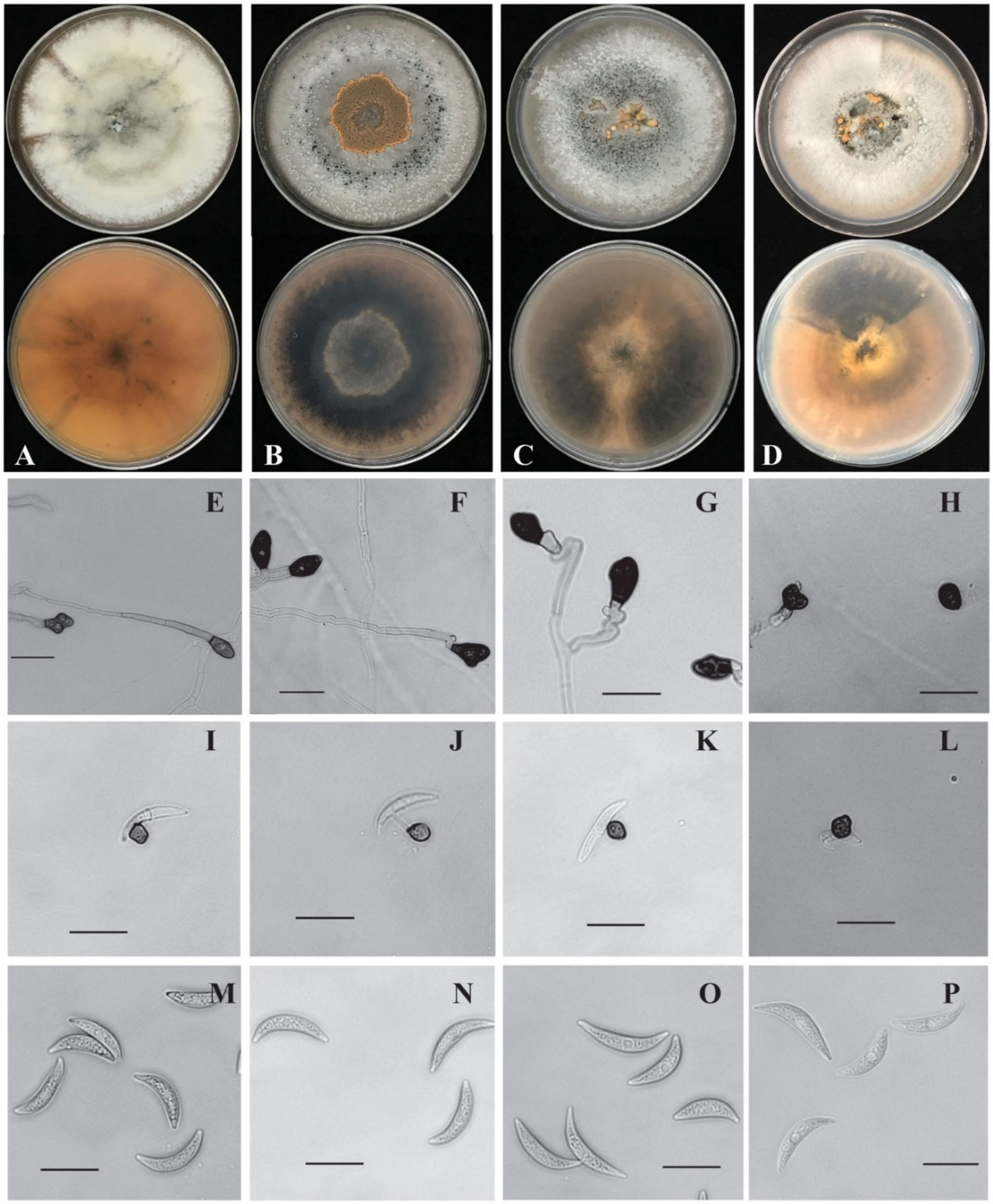


Figure 5.8. Morphological Features, Green clade (*Colletotrichum sublineola*). (A- D) Cultures on PDA, 15 d growth, from above and below; (E–H) Hyphopodia; (I- L) Appressoria; and (M- P) Conidia. Scale bars = 20 μm. Isolates (A, E, I, M) JG34.00, (B, F, J, N) SS1.001, (C, G, K, O) JG7.001, (D, H, L, P) JG42.001.

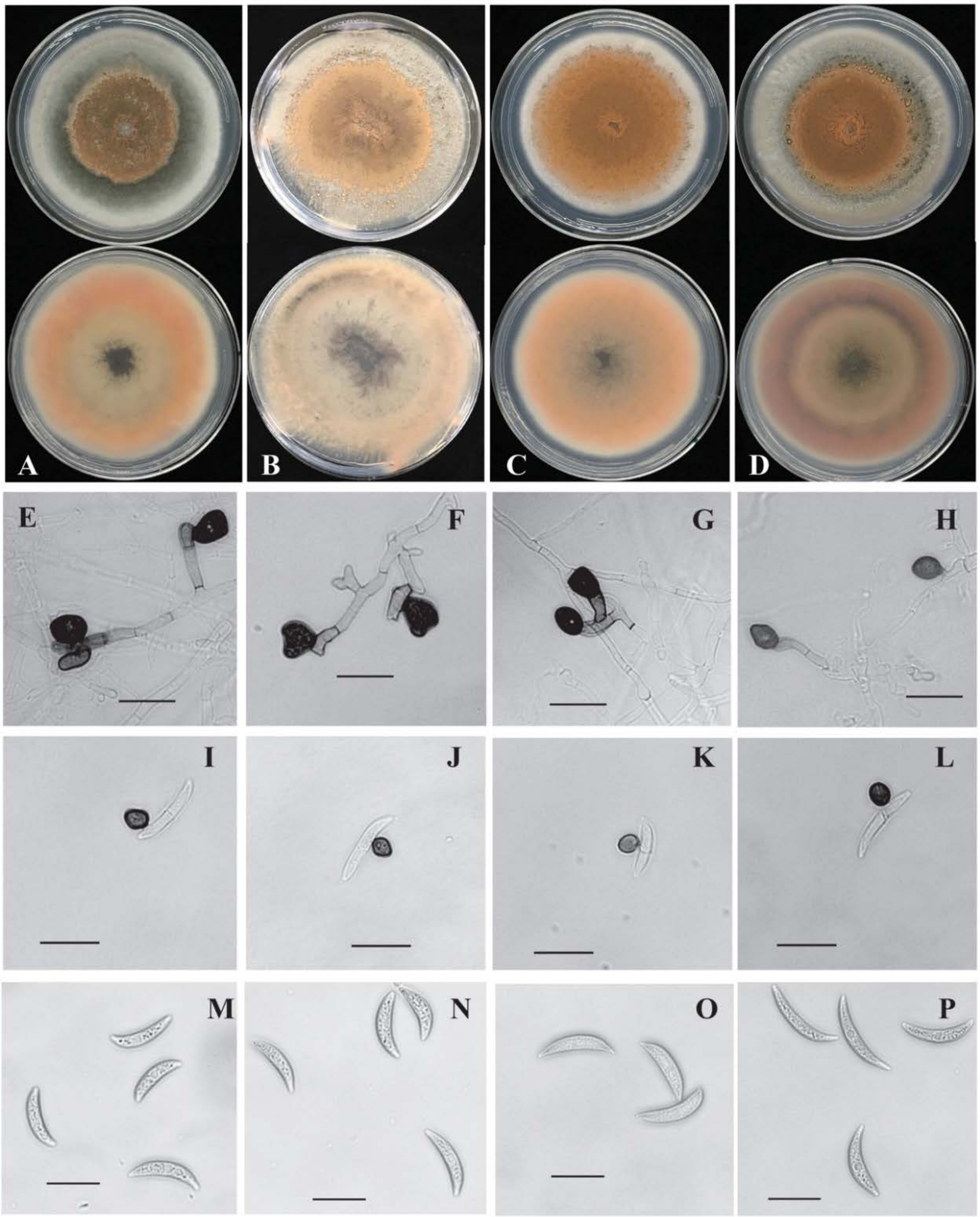


Figure 5. 9. Alignments; (A) *Sod2* and (B) *Mat1/Apn2*. S3.001-_LAB, sequence obtained from our laboratory culture of S3.001. S3001_CBS1 and S3001_CBS, sequence retrieved from JGI, and S3.001_DQ1 and S3.001_FJ3, sequences retrieved from GenBank (accesion number DQ132051.1 and FJ378029.1, respectively).

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment

S3001-_LAB      GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
S3001-CBS1      GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
S3.001_DQ1      GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
                ************************************************************

S3001-_LAB      AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
S3001-CBS1      AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
S3.001_DQ1      AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
                ************************************************************

S3001-_LAB      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
S3001-CBS1      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
S3.001_DQ1      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
                ************************************************************

S3001-_LAB      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
S3001-CBS1      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
S3.001_DQ1      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
                ************************************************************

S3001-_LAB      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
S3001-CBS1      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
S3.001_DQ1      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
                ************************************************************

S3001-_LAB      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
S3001-CBS1      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
S3.001_DQ1      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
                ************************************************************

S3001-_LAB      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
S3001-CBS1      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
S3.001_DQ1      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
                ************************************************************

S3001-_LAB      TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
S3001-CBS1      TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
S3.001_DQ1      TGGCTTTGTTCAGATTTGTAACGTATGCTTGGT-GCT
                ********************************* ***
```

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment

S3001-_LAB      CCACGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGACAAGA
S3.001_CBS      CCACGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGACAAGA
S3.001_FJ3      CCACGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGC----------CAAGA
                *********************************************          *****

S3001-_LAB      AATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCACAGTTGACACCCGGCT
S3.001_CBS      AATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCACAGTTGACACCCGGCT
S3.001_FJ3      AATGGTAGGTGAGAACGGTGGA----------AGCAACATACCACAGTTGACACCCGGCT
                **********************          ****************************

S3001-_LAB      TCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTGGCGCGACCCTCTTCC
S3.001_CBS      TCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTGGCGCGACCCTCTTCC
S3.001_FJ3      TCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTGGCGCGACCCTCTTCC
                ************************************************************

S3001-_LAB      CAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAAAGACTCTTTCCGAGG
S3.001_CBS      CAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAAAGACTCTTTCCGAGG
S3.001_FJ3      CAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAAAGACTCTTTCCGAGG
                ************************************************************

S3001-_LAB      TATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGGGGGACGGAGAAGGTT
S3.001_CBS      TATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGGGGGACGGAGAAGGTT
S3.001_FJ3      TATCGATTCCAG----CGTAGTGGTGGCGGAAGAGCCTGCCAAGGGG-------------
                ************    *******************************

S3001-_LAB      CTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGAGAGGAGGAGGGGTCT
S3.001_CBS      CTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGAGAGGAGGAGGGGTCT
S3.001_FJ3      --------------GCGTCT-------------------------GAGGAGGAGGGGTCT
                              ******                         ***************

S3001-_LAB      TTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGCCTTTGTTCTTTGTGG
S3.001_CBS      TTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGCCTTTGTTCTTTGTGG
S3.001_FJ3      TTGATACCCCCGAGAGAGGT--GCCACCGCCTGCGGCTGCTGTGCCTTTGTTCTT-----
                ********************  *********************************

S3001-_LAB      GCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCTTCGAGGGCGGTTGCT
S3.001_CBS      GCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCTTCGAGGGCGGTTGCT
S3.001_FJ3      --------TGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCTTCGAGGGCGGTTGCT
                        ****************************************************

S3001-_LAB      TGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGATCTGATTCGCTGGCCT
S3.001_CBS      TGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGATCTGATTCGCTGGCCT
S3.001_FJ3      TGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGATCTGATTCGCTGGCCT
                ************************************************************

S3001-_LAB      TGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATGCTCGCTTCAAGGGGG
S3.001_CBS      TGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATGCTCGCTTCAAGGGGG
S3.001_FJ3      TGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGT----------------------G
                *************************************                      *

S3001-_LAB      ATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAGACACTTGGGATGTTG
S3.001_CBS      ATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAGACACTTGGGATGTTG
S3.001_FJ3      ATCTTGTTTGTGAGGTTGGCAACT-TACCGGGCTTGATTGCGTAGACACTTGGGATGTTG
                ************************ ***********************************

S3001-_LAB      GTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCTCTTGGGGGGTGGTCA
S3.001_CBS      GTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCTCTTGGGGGGTGGTCA
S3.001_FJ3      GTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCTCTTG------GGTCA
                ****************** ******************************      *****

S3001-_LAB      GCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
S3.001_CBS      GCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
S3.001_FJ3      GCTCCTCGGAGACGAGACTCTGC------GACTCCAATG
                ***********************      **********
```

Chapter 6 - Summary and Significance of the Study

Populations of *Colletotrichum sublineola* that cause anthracnose disease on grain sorghum have been extensively characterized both genetically and pathogenically, and a high level of diversity has been demonstrated worldwide. This diversity has been used to explain the lack of durability of resistance genes that have been deployed to manage anthracnose in this crop. In contrast, comparatively little was known about the *Colletotrichum* populations that infect sweet or wild sorghum, even though anthracnose is listed as one of most important diseases of sweet sorghum in Kentucky, and is growing in importance across the southeastern U.S. In my dissertation research, I have described the genetic and pathogenic diversity in the population of *Colletotrichum* on sweet and wild sorghum in Kentucky and the southeastern U.S. by using several different approaches. To support my work, I developed many new protocols, including greenhouse and laboratory pathogenicity assays, and novel molecular markers for fingerprinting and for RFLP analysis. I amassed and archived a strain collection containing more than 400 *Colletotrichum* isolates from across the region. This collection will provide a valuable resource for plant breeders who want to efficiently screen sorghum germplasms for new sources of resistance. The greenhouse and laboratory assays that I developed will be useful for pre-screening germplasms with a large number of strains, or with exotic strains that require containment. My work will enable plant breeders to screen sorghum germplasms more quickly and efficiently for new sources of resistance or to characterize new isolates. Indeed, this approach has already been used effectively by a sorghum breeding company that I worked with last year.

Results from my field, greenhouse, and laboratory pathogenicity assays were generally consistent in ranking pathogen aggressiveness and host resistance, and demonstrated that *C. sublineola* has a significant potential to cause anthracnose disease on sweet sorghum in Kentucky when conditions are favorable (i.e. with a combination of a susceptible host and favorable environmental conditions). I discovered that most johnsongrass isolates, despite being much less aggressive than most isolates from *S. bicolor*, were still able to complete their life cycles on sweet sorghum plants. And I also found that some johnsongrass isolates are highly aggressive on sweet sorghum. This implies that johnsongrass could serve as a reservoir for the pathogen during crop rotation schemes, and could also be a source of new races. This would complicate the implementation of an effective disease management program, and efforts to develop new resistant varieties of sweet sorghum.

My studies of genetic diversity using RAPDs and novel RFLP fingerprinting markers identified two distinct sub-populations, one consisting mostly of isolates from cultivated sorghum, and the other including primarily isolates from johnsongrass. The fingerprinting analysis provided strong evidence for relatively rare cross-infections between these hosts. It also revealed that there is a very high level of genetic and pathogenic diversity within both sub-populations. Thus, I anticipate that *Colletotrichum* across the southeastern U.S. has a significant genetic potential to overcome individual resistance genes, and may pose a serious threat to widespread sorghum monoculture in the future. Plant breeders should monitor the levels of genetic diversity in the pathogen population, because resistant varieties carrying major gene resistance against anthracnose may not last for long when deployed widely in regions with high levels of pathogen diversity. The fingerprinting markers that I developed may be particularly helpful for this purpose. A goal for the future could be to develop a multiplexed RAPD primer protocol to facilitate individual strain identifications.

I confirmed that the closely related sibling species *C. sublineola* and *C. graminicola* are host-specific, and I also determined that the non-host reaction in both cases was associated with rapid cell death, a characteristic of ETI. Based on a genomic comparison of these two species, I identified putative pathogenicity genes that differed between them. I used some of these as probes to conduct an RFLP analysis of more than 80 representative *Colletotrichum* strains. Results of the analysis were consistent with the fingerprinting results, confirming that isolates from *S. bicolor* and from *S. halepense* mostly clustered into two distinct groups. However results of the RFLP analysis allowed further separation of the johnsongrass isolates into two groups, one containing isolates associated with *S. halepense* only, and another containing isolates from both *S. halepense* and *S. bicolor*. I identified several genes encoding putative effectors and specialized secondary metabolism proteins that showed allelic variation and presence-absence polymorphisms that differentiated the three groups of isolates. These allelic variants might play a role in pathogenicity and host determination: functional analyses of these genes should be performed to test this hypothesis. The RFLP analysis confirmed that there was a high degree of genetic diversity within these populations. High levels of genetic diversity among putative effectors and secondary metabolites suggests a possible threat to monoculture crops, because it increases the possibilities for rapid selection of new races able to overcome plant resistance genes, and of epidemics (“boom-bust” cycles). Pathogenicity assays revealed strong evidence of the existence of different races within the pathogen populations.

My phylogenetic analysis revealed that the three groups identified based on the molecular markers actually belonged to different species, two of which had not been previously described. I named these two new species *C. halepense* and *C. caselae*. My research findings suggest that *C. sublineola* and *C. caselae* have the ability to infect both *S. bicolor* and *S. halepense*, which has significant implications for sorghum breeding and anthracnose disease management in regions where johnsongrass is common. Identification of these species could be based on the inoculation of a differential host series. However, for more rapid diagnosis a better option would be to generate species-specific primers based on the putative pathogenicity genes. Accurate and rapid diagnosis will be an important tool for improved disease management.

My research findings have provided critical new information for plant breeders that develop and deploy resistant sweet sorghum varieties in areas where johnsongrass is also present. For the growers, my work implies that the pathogen has the potential to quickly overcome plant resistance based on major genes, increasing the risk of epidemics if those genes are widely deployed. The ability of two of the species to infect both *S. bicolor* and *S. halepense* complicates the use of crop rotation where johnsongrass is common, so alternative management strategies might be needed.

Appendix A - Optimization of whole plant inoculation in the greenhouse

The goal of this work was to determine the age at which sweet sorghum plants were most susceptible to anthracnose caused by *C. sublineola* in order to optimize greenhouse pathogenicity assays.

Material and methods: Seeds of sweet sorghum (variety Sugar Drip) were sowed once per week for a period of four weeks. The experiment was conducted using plants of four different ages: (i) plants with eight fully developed leaves (V8); (ii) seven fully developed leaves (V7); (iii) six fully developed leaves (V6); and (iv) five fully developed leaves (V5). Four sorghum seeds were sowed in each 8' inch pot, in a mixture of three parts Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada) and two parts sterile topsoil. The pots were kept at the greenhouse under 14 hours of light and were watered daily to saturation and fertilized with a solution of 150 ppm of Peters 20-10-20 (Scotts-Sierra Horticultural Product Co., Marysville, OH), three times per week, beginning one week after seedling emergence.

Pots were arranged in a randomized complete block design (RCBD) with four blocks, with each experimental unit consisting of three pots, containing a total of 12 plants. The treatments were as follows: (1) V5 plants inoculated with SS1001; (2) V5 plants inoculated with GS1001; (3) V5 plants inoculated with JG9001; (4) V5 plants inoculated with JG18001, (5) V5 plants inoculated with water; (6) V6 plants inoculated with SS1001; (7) V6 plants inoculated with GS1001; (8) V6 plants inoculated with JG9001; (9) V6 plants inoculated with JG18001, (10) V6 plants inoculated with water; (11) V7 plants inoculated with SS1001; (12) V7 plants inoculated with GS1001; (13) V7 plants inoculated with

JG9001; (14) V7 plants inoculated with JG18001, (15) V7 plants inoculated with water; (16) V8 plants inoculated with SS1001; (17) V8 plants inoculated with GS1001; (18) V8 plants inoculated with JG9001; (19) V8 plants inoculated with JG18001, (20) V8 plants inoculated with water. The entire experiment was repeated twice, during the winter of 2013 and the summer of 2014. Inoculum preparation and plant inoculation was performed as described in chapter 2 of this dissertation

The number of plants that were infected (i.e. displayed lesions containing sporulating acervuli) was recorded weekly. Data were subjected to analysis of variance (ANOVA) using the PROC GLM procedure of the SAS software package (SAS Users Guide, SAS Institute, Cary, NC). Mean comparisons were conducted using Fisher's protected least significant difference (LSD) test at $\alpha = 0.05$. Pearson's correlation coefficients for different variables were computed using the PROC CORR command of the SAS program. The analyses were performed using SAS version 9.3.

Results and Conclusions: The results of the two experiments were similar. There were no statistical differences (p value > 0.10) among the treatments within the blocks, or for the isolate X plant stage interaction (p value > 0.10) at any time point in either of the two experiments. Therefore, data were pooled for plant stage and isolate, and reanalyzed for both experiments (Figure A1.1A, B).

The strain SS1001 was statistically more aggressive than all the other treatments across all time points (p value < 0.05) (Figure A1.1A, B). The effect of the plant stage was not significant for most time points (p value > 0.01). However, in the 2013 experiment, the V6 plants were statistically the most susceptible to the anthracnose (p value < 0.05) in the second week of evaluation (Figure A1.2A). In the 2014 experiment, the V5 plants were statistically more susceptible to the anthracnose (p value < 0.05) from the second to the forth week of disease evaluation (Figure A1.2B).

From the two experiments I concluded that the younger vegetative plant stages (V5 and V6) were the most susceptible to *Colletotrichum*. I also concluded that the most aggressive fungal strain was SS1001.

Figure A2.1. Pathogenicity assay of four strains of Colletotrichum on Sugar Drip. Disease evaluation was performed based on the number of infected (susceptible) plants (incidence), during four weeks after appearance of symptoms. (A) First repetition performed in the winter 2013 and (B) second repetition performed in the summer of 2014.

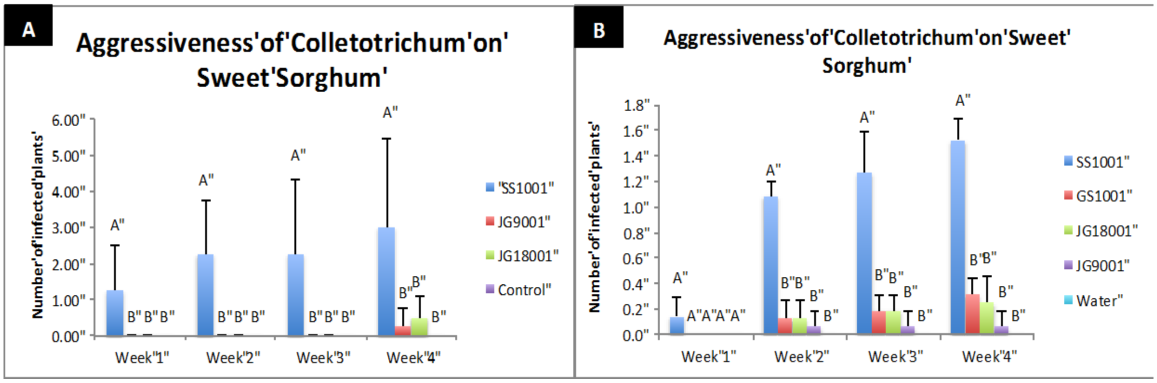
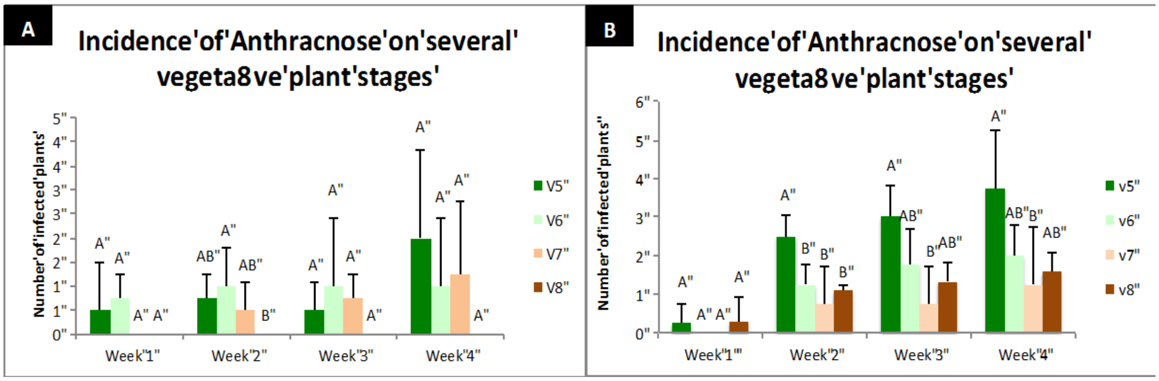


Figure A2.2. Pathogenicity assay of four strains of Colletotrichum on Sugar Drip. Disease evaluation was performed based on the number of infected (susceptible) plants (incidence), during four weeks after appearance of symptoms. (A) First repetition performed in the winter 2013 and (B) second repetition performed in the summer of 2014.



Appendix B - Optimization of sheath assay and rating system for sorghum anthracnose

Introduction: A cytological study of compatible and incompatible interactions of grain sorghum with *C. sublineola* was performed by using a leaf sheath infection assay (195). This assay allowed detailed visualization of the pathogen growth into the plant cells, and of the plant response. The leaf sheaths were spray-inoculated, which made it relatively difficult to find and count the widely dispersed infection sites. Kankanala and her collaborators used detached leaf sheaths to study the interaction between rice and *Magnaporthe oryzae* (97). They applied fungal inoculum directly to the inner surfaces of the detached sheaths which made it more convenient to synchronize and locate infection sites. This method was successfully adapted to study the details of infection of maize by *C. graminicola* (97, 175). My goal for the work described in this appendix was to optimize this detached sheath assay to investigate interactions of sorghum and *C. sublineola*.

Effect of inoculation method

In the protocol described by Torres et al. (2014), a 20 µl drop of a *C. graminicola* spore suspension was applied to the inner surface of a detached maize leaf sheath. Sorghum leaf sheaths are quite small compared with maize sheaths, and it is relatively difficult to apply drops without damaging them. In studies of infection of rice by *M. oryzae*, the entire detached rice leaf sheath was filled with a spore suspension (97). The aim of this experiment was to compare the effect of drops versus the more convenient filling method on the interaction between *C. sublineola* and sweet sorghum.

Material and methods: The fungal strain, CgSl1 isolated from grain sorghum was used for these experiments. Sweet sorghum seedlings (var. Simon) at the V3 growth stage were cut at the soil line and brought to the laboratory. The sheaths were removed from the second leaves and cut into three-inch segments. The segments were inoculated by one of two methods: (i) applying a spore suspension ($5x10^5$ spores/mL) to fill up the whole leaf sheath; and (ii) applying a 20 μL drop of the spore suspension ($5x10^5$ spores/mL) to the inner part of the leaf sheath just above the midrib. The inoculated sheath segments were supported inside a Petri dish with the midrib side downward. The Petri dishes were lined with moistened Whatman paper (Whatman No. 1, Whatman, Hillsboro, OR, USA) to maintain high humidity. Dishes were incubated at 23° C under continuous light. Fungal colonization was evaluated at 72 hours after inoculation (hpi).

To observe fungal colonization, the blade sections of the inoculated leaf sheaths were removed, leaving only the midrib section. This was then turned upside down on a wet slide, and a single-edged razor blade were used to remove excess tissue to leave a thin layer of mostly epidermal cells containing the fungal infection sites. A total of three sheaths, fifty infection sites per sheath, were observed for each treatment. The experiment was repeated twice.

Results and Conclusions: After 72 hours, only half of the propagules in drop-inoculated treatments had given rise to successful infections inside the host cells (Figure A2.1). In contrast, 90% of the propagules in filled treatments had produced infectious hyphae inside the host cells. Thus, for the compatible interaction filling the leaf sheath with the spore suspension resulted in faster development of the pathogen than the drop inoculation. However, in incompatible interaction *C. graminicola* was not able to colonize the cells even in filled inoculations. For the interaction corn x *C. graminicola* we notice that the pathogen was able to colonize and sporulate faster by filling the leaf sheath with the spore suspension than by using a single drop. So it seems that filling leaf sheaths compromises the strength of the resistance response. Some possible reasons for this could be effects on the cuticle of being submerged or some kind of quorum issue related to having multiple infection sites across the entire leaf versus having infection sites concentrated in a single area.

For all the leaf inoculation experiments performed in this dissertation I used the more convenient fill inoculation methodology.

Time course of infection

Torres et al. (2104) while studying the interaction between maize and *C. gramicola* observed that the fungus produced appressoria by 24 hpi, and that more than 90% of the infection sites consisted of hyphae that were colonizing at least two cells by 48 hpi. Secondary necrotrophic hyphae became evident by 60 hpi. Wharton and collaborators reported that *C. sublineola* formed secondary hyphae by 66 hpi. The goal of this experiment was to describe the time course of infection of detached sorghum leaf sheaths that were inoculated by filling with *C. sublineola* spore suspensions. The goal was to determine the best time point for the evaluation of fungal infection for my assay.

Material and methods: Fungal strains CgSl1 (from grain sorghum); M1001 (from maize); and JG18001 (from johnsongrass); and V3 seedlings of sweet sorghum (Simon) and sweet corn (Golden Jubilee) were used for the experiments. The leaf sheaths were prepared, inoculated by filling with spore suspensions ($5x10^5$ spores/mL), and incubated in moist chambers as described above. Fungal colonization was evaluated at 12, 24, 36, 48, 60, 72 and 108 hours post inoculation (hpi). I observed 50 infection sites per leaf sheath, and recorded the developmental stage of the infection for each. A total of 10 leaf sheaths were observed for each treatment, and the experiment was repeated three times.

Results and Conclusions: With the fill inoculation technique, *C. graminicola* colonized sweet corn much more quickly than had been reported by Torres et al., (2014), who had used the drop inoculation technique and a field corn inbred (Table A1.1). *C. sublineola* colonized sweet sorghum more slowly than *C. graminicola* colonized sweet corn. *C. sublineola* formed primary hyphae by 36 dpi, and didn't start to produce secondary hyphae until 72 dpi (Table A1.1). I decided to do my assay evaluations at between 60 and 72 hpi.

Effect of leaf sheath age

The work of Ferreira and Warren (1982) showed that 35-day-old grain sorghum seedlings were more susceptible to *C. sublineola* than 15-day-old seedlings. My goal in these experiments was to discover whether *C. sublineola* aggressiveness varied according to age of the leaf sheath.

Material and methods: Two fungal strains (CgSl1 and JG18001) were used for this experiment. Plant material used was the sweet sorghum variety Simon. Seedlings at the V5 stage of development were cut at the soil line and brought to the laboratory. Four leaf sheaths removed from each seedling, which were designated in order from oldest to youngest as (i) senescent; (ii) old; (iii) mature; and (iv) young, which was the youngest full-expanded leaf. The leaf sheaths were prepared, inoculated by filling with spore suspensions ($5x10^5$ spores/mL), and incubated in moist chambers as described above. Observations were made at 60 hpi. The experiment was performed once with each strain.

Results and Conclusions: The youngest leaf sheaths were the most susceptible to CgSl1 (Figure A2.2). I observed 15 infection sites in which the pathogen was growing in more than two cells (Figure A2.2). The old and mature leaf sheaths were more resistant: the pathogen was found colonizing more than two cells in only 3-5 infection sites on those sheaths. The senescent leaf sheaths were the most resistant to CgSl1. The pathogen was not observed to colonize more than two cells at any infection sites in those sheaths (Figure A2.2). Similar results were found in inoculations with JG18.001, which was generally less aggressive than CgSl1 (Figure A2.3). Based on these experiments I decided to use sheaths from the fourth fully expanded leaves of V5 plants for all subsequent experiments.

Disease evaluation scale

Wharton and collaborators (195, 196) observed at a cellular level how *C. sublineola* attacks grain sorghum and how the plant defends itself. The pathogen produces primary hyphae that grow from one cell to another, invading cells that are alive but killing them rapidly once they have entered them. By 66 hpi, the cells appeared to be disorganized and the pathogen had produced narrower secondary hyphae branching from the primary hypha. These authors described how plant cells responded to invasion by the accumulation of a red pigment, which was distributed throughout the infected cells and which also accumulated in the epidermal plant cell walls. They reported that compatible interactions did not result in the accumulation of these red compounds (195, 196). However, in my own studies I observed production of red vesicles or pigments in incompatible interactions (in which the fungus did not colonize significantly) as well as in compatible ones (in which the fungus colonized several cells). I could observe mixed infection types on the same leaf sheaths (Figure A2.4). The aim of the work described in this section was to evaluate the range of colonization and response types that occurred in the interaction of *C. sublineola* and sorghum, and to develop a uniform diagrammatic scale that I could use to evaluate fungal pathogenicity and plant response for the leaf sheath assay.

Materials and Methods: I used the inoculation to fill up the whole leaf sheath described appendix 1A.

Results and Conclusions: I developed a diagrammatic scale for evaluation of the interaction between *C. sublineola* and sweet sorghum var. sugar drip at a cellular level under the light microscope with no need to stain or fix the plant tissue. This scale includes all the interaction types that I observed in more than a thousand separate interactions. This diagrammatic scale was used to evaluate all leaf sheath assays performed in this dissertation.

Table A2.1. Summary of the time course of maize inoculated with *C. graminicola*, sweet sorghum inoculated with *C. sublineola* (CgSl1 and JG18001). Disease evaluation was performed every 12 hours.

Evaluation (hpi)	**Maize/M1001**	**Sorghum/CgSl1**	**Sorghum/JG18001**
12	App	App	App
24	PH	App	App
36	SH	PH	PH
48	SH	PH	PH
60	Acervuli	PH	PH
72	Acervuli	SH	SH
108	Acervuli	Acervuli	Acervuli

App, appressoria; PH, primary hyphae; and SH, secondary hyphae.

Figure A2.1. Development of CgSl1 when inoculated on sweet sorghum leaf sheaths by drop or fill inoculation at 72 hpi. Number of infection sites with appressoria, first picture; or has grown in one cell (second picture); or has grown in more then 2 cells (third picture). Error bars represent standard deviation.

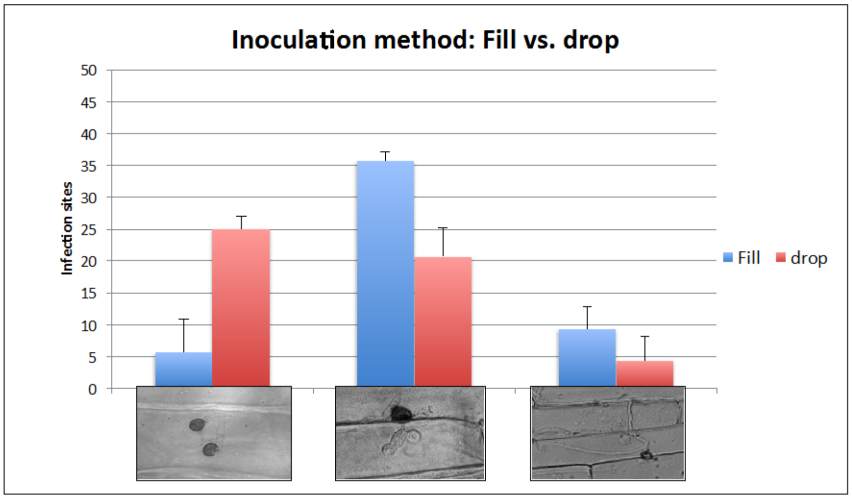


Figure A2.2. Development of CgSl1 on sweet sorghum leaf sheaths at different stages. Senescent leaf sheaths (dark orange bar); old leaf sheaths (light orange bar); mature leaf sheaths (light green bar); young leaf sheaths (dark green bar). Number of infection sites with ungerminated spores, appressoria, hyphae growing in one cell, or more then 2 cells; infection site surrounded by reddish, vesicles, cells entirely red, and red hyphae. Error bars represent standard deviation.

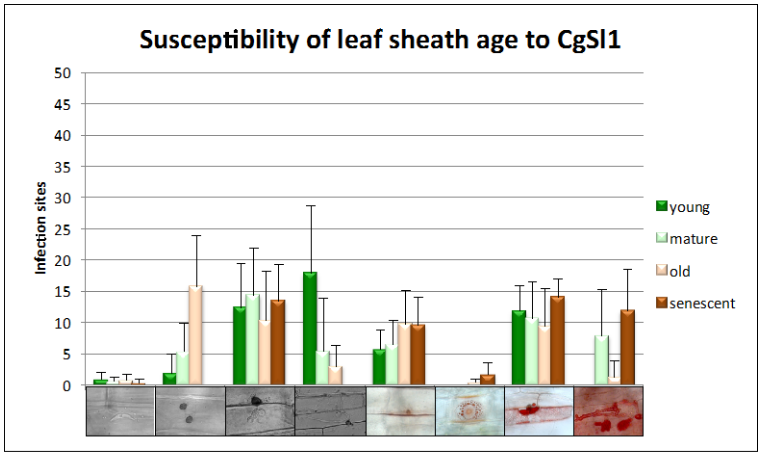


Figure A2.3. Development of JG18001 on sweet sorghum leaf sheaths at different stages. Senescent leaf sheaths (dark orange bar); old leaf sheaths (light orange bar); mature leaf sheaths (light green bar); young leaf sheaths (dark green bar). Number of infection sites with ungerminated spores, appressoria, hyphae growing in one cell, or more than 2 cells; infection site surrounded by reddish, vesicles, and cells entirely red. Error bars represent standard deviation.

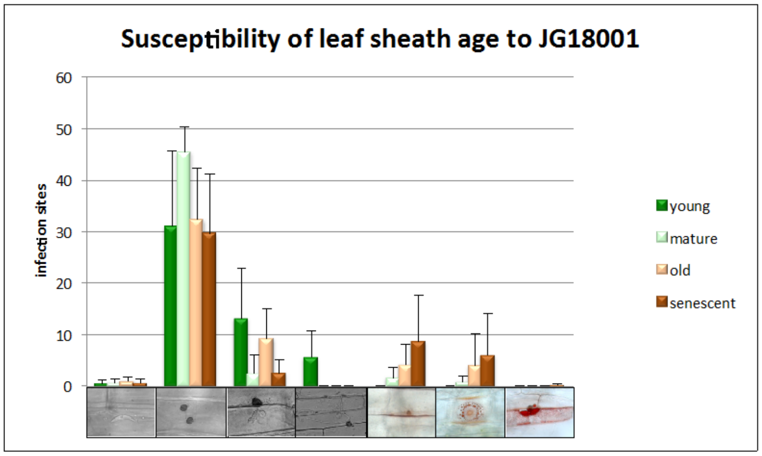


Figure A2.4. Pattern of CgSl1 development on sweet sorghum leaf sheaths at 106 hpi. On the top of the picture there are red cells which a plant defense strategy in a incompatible. In the bottom of the picture we there are acervuli, which are structures produce by the pathogen in a compatible interaction.

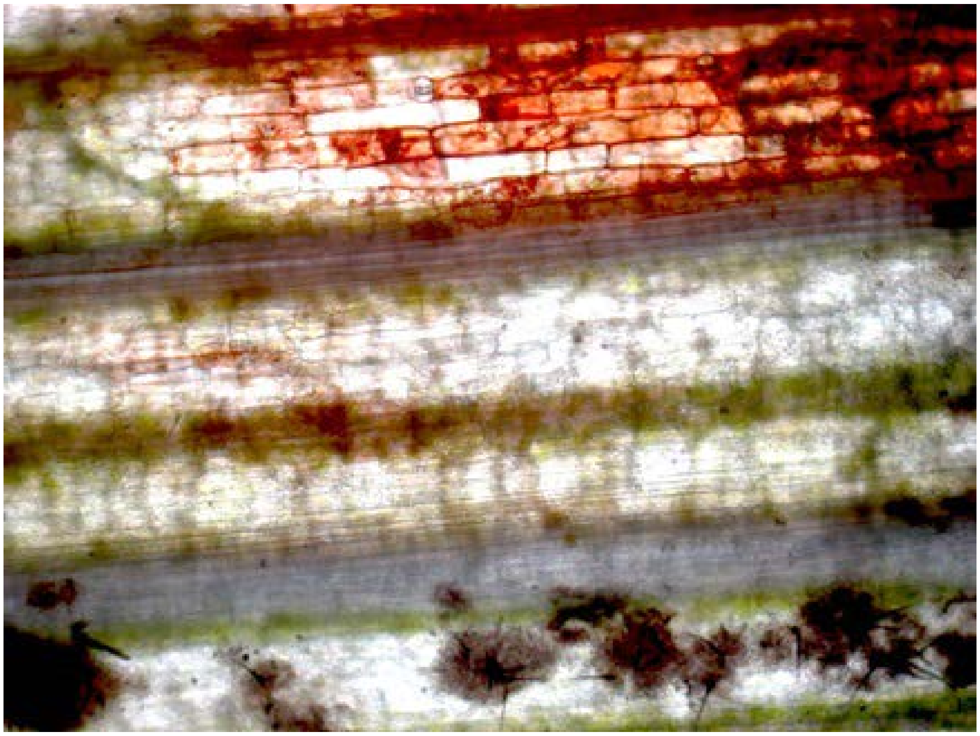
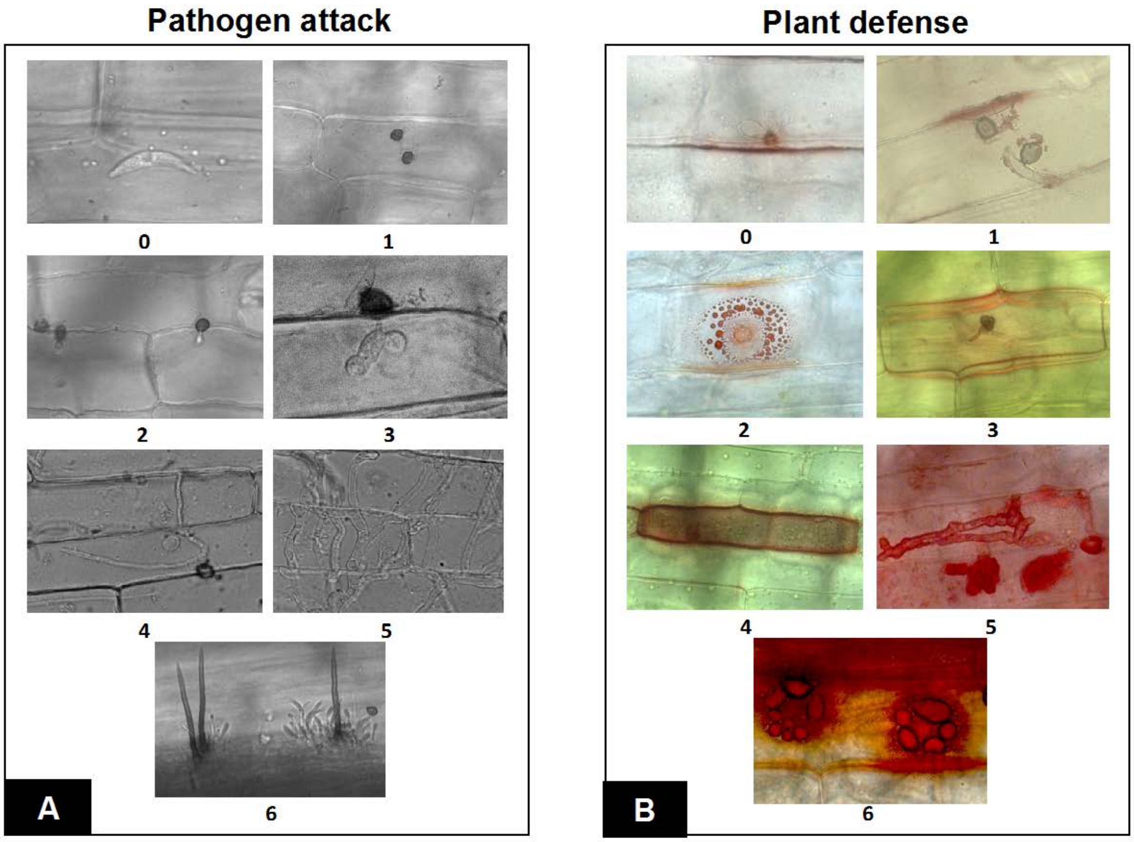


Figure A2.5. Diagrammatic scale for evaluation of the interaction between *C. sublineola* and sweet sorghum. A. Pathogen Attack; (0) spore did not germinate, (1) Appressoria has been formed, (2) Hyphae smaller than appressoria, (3) Hyphae bigger than appressorium but restricted in one cell, (4) Hypha growing in more than two cells, (5) Hypha growing in more than three cells, and (6) Acervuli has been formed. B. Plant Defense; (0) Reddish surround infection site, (1) Reddish surround appressoria, (2) Red vesicles around the infection site, (3) Plant cell is light red, (4) Plant cell is very red, (5) Red compound has covered the pathogen hyphae, and (6) Big red vesicles has accumulated.



Appendix C - Alignments used to create the unrooted Secondary Metabolites Phylogenetic Trees in Chapter 4

PART 1: DMAT protein sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
CH063_0589      ------------------------------------------------------------
GLRG_01200      ------------------MAQLIPPSHNALVVRPHIPITCQEREEVLRSVEYNVFAFPAA
CS1             ------------------MARLIPPSHNALVVRPHIPITAEEREEVLRSIEYTVFAFPAA
CH063_0902      ---------------------------MGEMIMSHEPWDFLNSVLRFRD-----------
GLRG_07878      ---------------------------MGEMIMSYEPWDYLNSILRFRD-----------
CS.DMAT         ---------------------------MGEMIMSYEPWDFLNSILRFRD-----------
CS7             ----------------------------------MPGLTTASRTQAEKS-----LVVWQS
CH063_0738      ---------MEATSTDQFPLWDSPFRGKTKSWMYKGAGATESKSKSTSKPEHDGTLVFPS
GLRG_11574      --------------------------------MHNGDIVADSKPKQYLT----GNLILSS
CS6             ---------------------MKPARDKTKSWMHEGDIVTESKAKQNST----GIPVLPS
CH063_0204      ------------------------------------MVAATNGSDLARQ-----------
GLRG_11853      ------------------------------------MVMSTSGPEPERQ-----------
CS8             ------------------------------------MVIATSGPEPSRQ-----------
CS5             -----------------------MPSMERMEIYRLNDVRDDHSGTVTKN-----PTYNEA
GLRG_11423      -----------------------------MEIYRLTVVNDANGSI-TEK-----SIYKDA
CH063_1280      -----------------------------MEIYRLSAVHDASGGIDTKR-----NIYNDA
CH063_0144      -------------------------MSKTVLIRPRGPVLSAASLDTGP------------
CH063_0491      ------------------MQATASIPPGLPAEHGHAAVGRPGLSAQTQS---------SN
CB2             MAPIIATPTYTATTRSPHPLQLSSRCVRSPMAFGQTEILSPRAMAKQEK-----DLCLPA
CH063_0654      ----MATSTYTTSHRPPFPLQLSPRALRSPMAFGQSEILSPRALAKQEK-----DLCLPA
GLRG_02500      ------------------------------MAFSQADILSPRAMAKQEK-----DLCLPA
CS4             --------------------------------MALMPVRTTS------------------
GLRG_07955      ------------------------MTPETISNDFQTNVDAPRPPQNFKE-------FKSL
CH063_1116      ------------------------------------------------------------
CB.DMAT         ----------------------------------MGEISSGSYNLN--------------
CH063_0401      ------------------------------------------------------------


CH063_0589      ------------------------------MMYEANYPDETSNRFLTF------------
GLRG_01200      LLTCDFLSDSGTSAMTDVQWAAMLRGDESYGRNSGYYCLLEAFRDIFE---------RDV
CS1             LLTCDFLSDSGTSAMTDVQWAAMLRGDESYGRNSGYYCLLEAFRDIFERGEDRKYVFRDV
CH063_0902      -------------NDSQFWWDKTGRMFSKLLKYAG-YSASEQYRELNF------------
GLRG_07878      -------------DDSQFWWDKTGMMFSKLLKYAG-YSTSEQYRELNF------------
CS.DMAT         -------------DDSQFWWDKTGRMFSKLLKYAG-YSTSEQYRELNF------------
CS7             LAKKCLPPAGN--KNTEFWWNLTGYHLAVMIDAAG-YSTQLQYETLLF------------
CH063_0738      VSKWLPPRD----EHSDYWWNMAGPHFATLFRNAG-YSLEEQYEALLF------------
GLRG_11574      VTKWLPSRD----EHSDYWWNMAGPHFSALLQNAG-YSSQAQYETLLF------------
CS6             VSKWLPSRD----EHSDYWWNVAGPHFSALLENAG-YSIQAQYEILLF------------
CH063_0204      -------------EDSRFWWDELAHPLLSLMKSAG-YSAEEQEQYVEF------------
GLRG_11853      -------------EELQFWQDEISHSLLSLMQSAG-YSLKEQEHYLEF------------
CS8             -------------DESQFWRDELSHSLLSLMQSAG-YSLKEQQHYLEF------------
CS5             IDPSTSSSKGSPRPLVSSWLPIIHNSLSSLLRWAGPYPAEVQESHLAF------------
GLRG_11423      VASSP-SSKGS--PLFSSWLPLIHNSLTSLLRWTGTYPAEVQESHLAF------------
CH063_1280      VDSSPSSSKGS--PLYSSWLPLIHNSLSSLLRWTGTYPADVQESHLAF------------
CH063_0144      -------------EDVDFWSSYLVPRFHAYLSEAGSYTPEQQAAHLSC------------
CH063_0491      ALYVGMNPSPPANPDHRFWWDHFAPFLLSWLCASGCYSESDIAAQLTI------------
CB2             QESIYIHK-----PDQDFWAARCSESLDCLMQAAGSYTPAQRAAHLQF------------
CH063_0654      QESIYIHR-----PDQDFWTVRCAEILDGLMQAAGSYTPAQRAAHIRL------------
GLRG_02500      QESIYIHR-----PHQDFWAVRCAESLDGLMQAAGSYTPAQRAAHLQF------------
CS4             ----------------------------------------------PL------------
GLRG_07955      VNNNEVPIDPPLNPEHRWWWDRCAPLLNSLLNSAESYTSEEKADHLRV------------
CH063_1116      -------------MAPAVTLNHDTDRPLSLLTSAGSYTPEEKADHLRV------------
```

```
CB.DMAT         -------------PDQQYWWQFVEPALSSMLVYAGRYTLEEQESHKAW------------
CH063_0401      ------------------------------------------------------------


CH063_0589      ------------------------------------------------------------
GLRG_01200      LAGTADGIFYRQTFPKEVGGGFVNGGPLQLIRPNFFILPQGRCAESLLFSTMSEILRNES
CS1             LAGTADSTLYRQTFLKEVA-----------VAPSL---------ARQLSSAM--------
CH063_0902      ------------------------------------------------------------
GLRG_07878      ------------------------------------------------------------
CS.DMAT         ------------------------------------------------------------
CS7             ------------------------------------------------------------
CH063_0738      ------------------------------------------------------------
GLRG_11574      ------------------------------------------------------------
CS6             ------------------------------------------------------------
CH063_0204      ------------------------------------------------------------
GLRG_11853      ------------------------------------------------------------
CS8             ------------------------------------------------------------
CS5             ------------------------------------------------------------
GLRG_11423      ------------------------------------------------------------
CH063_1280      ------------------------------------------------------------
CH063_0144      ------------------------------------------------------------
CH063_0491      ------------------------------------------------------------
CB2             ------------------------------------------------------------
CH063_0654      ------------------------------------------------------------
GLRG_02500      ------------------------------------------------------------
CS4             ------------------------------------------------------------
GLRG_07955      ------------------------------------------------------------
CH063_1116      ------------------------------------------------------------
CB.DMAT         ------------------------------------------------------------
CH063_0401      ------------------------------------------------------------


CH063_0589      -------------------------------YRDVVCPLLG-GRPKPXSL----------
GLRG_01200      ILNNDTTDPPALISNGFFDTTGANAAATGFQLHTLTQPGLTDPFPFDQVEKSNPFKGNLD
CS1             ---------------GFFDTTGAKAAATGFQLHTFTQPGLTDPFPVEPVGRSSPFKGNLD
CH063_0902      -------------------------------YSLFVAPELG---PSPDCHGNV---RRW-
GLRG_07878      -------------------------------YSLFVAPELG---PSPDSQGNV---RRW-
CS.DMAT         -------------------------------YSLFVAPELG---PSPDSHGNV---RRW-
CS7             -------------------------------HYHWIVPRLG---PAPYPDGN----PKF-
CH063_0738      -------------------------------VYHQVIPRLG---AAPGSSDSAAAAAAGG
GLRG_11574      -------------------------------IYHQITPRLG---PSPLLSNA----DRL-
CS6             -------------------------------IYHHISPKMG---PSPIISNP----DRP-
CH063_0204      -------------------------------VNQFVVPSLG---PRPMTTKTGVRLPHF-
GLRG_11853      -------------------------------TKQHVFSSLG---PPPITTKSGARIPHF-
CS8             -------------------------------TNQHVVSSLG---PPPITTKSGVRIPRF-
CS5             -------------------------------MRDVVVPCLK---PPAAA-------DRL-
GLRG_11423      -------------------------------VRDVVVPRLK---PPAAA-------DRL-
CH063_1280      -------------------------------VRDVVVPRLK---PPAAA-------DRL-
CH063_0144      -------------------------------VRAAAV-ALGPKHPHPHV-----------
CH063_0491      -------------------------------LRDVVIPAFG---PPSVV-----------
CB2             -------------------------------LSEIIVPAMG---PHPST-------APT-
CH063_0654      -------------------------------LTEIIVPAMG---PHPSA-------APT-
GLRG_02500      -------------------------------LSEIIVPAMG---PHPSA-------APA-
CS4             -------------------------------ARTVVVPSSG-----------------F-
GLRG_07955      -------------------------------FRDVVVPSFG--TPTPR--------AKV-
CH063_1116      -------------------------------FRDVVVPSFG--TPTPE--------AKV-
CB.DMAT         -------------------------------YATQVTPIFG---PRPSE-------GNP-
CH063_0401      ------------------------------------------------------------


CH063_0589      ----PTAVGWDGN---------PFEYSF--------------------------------
GLRG_01200      IEKTSTFLSCNGNAARSVMLLLTITNNWAGAQPVSMANIKATAALAQTHDIPLFFDACRF
```

```
CS1             IEKTSAFLSRDGNAARSAMLLLTITNNWAGAQPVSMANIKAASALAQTHDIPLFFDACRF
CH063_0902      ----RSPGTPDST---------PIDFSW--------------------------------
GLRG_07878      ----KSPGTPDST---------PIDFSW--------------------------------
CS.DMAT         ----RSPGTPDST---------PIDFSW--------------------------------
CS7             ----KSLIAYDGS---------PFEYSW--------------------------------
CH063_0738      ----DSGLSLDNT---------HIEYSW--------------------------------
GLRG_11574      ----DSGLSLDGT---------RFEYSW--------------------------------
CS6             ----DSGLSIDGT---------RFEYSW--------------------------------
CH063_0204      ----DSFCSDDFS---------PAELSW--------------------------------
GLRG_11853      ----DSFCSDDFS---------PAELSW--------------------------------
CS8             ----DSFCSDDFS---------PAELSW--------------------------------
CS5             ----SYVANHNHS---------TYEASL--------------------------------
GLRG_11423      ----YYIGTHSHG---------IYEASL--------------------------------
CH063_1280      ----HYIGTHSHG---------IYEASL--------------------------------
CH063_0144      ----KSALTKNGS---------PIELSF--------------------------------
CH063_0491      ----QHALKPTGS---------PFELSW--------------------------------
CB2             ----KSLLTADGS---------PFKPSW--------------------------------
CH063_0654      ----KPLLTADGS---------PFEPSW--------------------------------
GLRG_02500      ----KPLLTADGS---------PFEPAW--------------------------------
CS4             ----QHRRQKSGP-----------SRTM--------------------------------
GLRG_07955      ----RPLLTYDGS---------TFEPSW--------------------------------
CH063_1116      ----KPLLTYDGS---------TFEPSW--------------------------------
CB.DMAT         ----DPIFTHDSS---------PCHISI--------------------------------
CH063_0401      ------------------------------------------------------------


CH063_0589      --------EFKGSTKKAGVRFVLDLSEVRPA--------------------NRDYPL-SL
GLRG_01200      AENAKFIQDFEDGYSNMTVPQVVREMFSHAGGFTISLKKDGLATMGGALCFRDEGMFSRK
CS1             AENAKFIQDFEDGCGNKTIPQIVQEMFSQVDGFTISLKKDGLANMGGALCFRDEGVFSRK
CH063_0902      --------EWGH-DNRAVIRYSFEPIGLHAG--------------------TDLDPL-NR
GLRG_07878      --------EWGC-DNRAVIRYSFEPIGLHAG--------------------TDLDPL-NQ
CS.DMAT         --------EWGH-DNRAVIRYSFEPIGLHAG--------------------TDLDPL-NQ
CS7             --------KWNTTAGEPDIRYSWEAINSSSG--------------------TTADPL-NH
CH063_0738      --------RWNXADTKPEIRMVMEPFSRFAG--------------------TYLDPL-NL
GLRG_11574      --------RWNQSYTKPEVRMVVEPFSRFAG--------------------TYMDPL-NI
CS6             --------RWNQPDTKPEVRMVLEPFSRFAG--------------------THMDPL-NI
CH063_0204      --------NVGP--ARSKIRVGSEPIGPFAG--------------------TSKDPF-NQ
GLRG_11853      --------NIGP--KRSKIRVGSEPIGPYAG--------------------TAKDPF-NQ
CS8             --------NIRP--KQSKIRVGSEPIGPFAG--------------------TTRDPF-NQ
CS5             --------AFSS-YRPPKVRYSVQPLVDP----------------------SPGDPL-GQ
GLRG_11423      --------AFAT-HKPARVRFTVQPLVDPSPA-------------------QGGDLL-GQ
CH063_1280      --------AFGS-HKPARVRFTVQPLVDPSP--------------------QRGDPL-GQ
CH063_0144      --------NLSE-DRLPTARFYIEPLGPETG--------------------TENDPF-GE
CH063_0491      --------NFTP--HGNTIRYTFQPMGSRAG--------------------TDEDPFGSS
CB2             --------TITD-SGSSFIRFSFEPMGRHGG--------------------SASDPF-AQ
CH063_0654      --------NITD-SGSSSIRFSFEPMGRHGG--------------------SESDPF-AQ
GLRG_02500      --------NITD-SGSSLVRFSFEPMGRHGG--------------------STSDPF-AQ
CS4             --------NFTK-GDDGVIWYTFEPLCDTTG--------------------SADDPF-AG
GLRG_07955      --------NFTK-GNDGVVRYTFEPLGDTAG--------------------SEEDPF-AG
CH063_1116      --------NFTE-GDDGVIRYTFEPLGDDAG--------------------SEEDPF-AG
CB.DMAT         --------NWCS-KTKPTVRSGMT---------------------------RPHDVFNSQ
CH063_0401      ------------------------------------------------------------


CH063_0589      ATVEKVLDVLSTK--SPL-YDDH--------------WHRSL-ERW----------FVYS
GLRG_01200      FGADVGIRLKERQIMCYG-NDSCGGMSGRDVMAAAVGLYEVTKEAYLADRIGQVKRFAEG
CS1             FGADVGIRLKERQIMCYG-NDSYGGMSGRDVMAAAVGLYEVTKESYLADRINQVKRFAEG
CH063_0902      HATNDWIFKLQQQKMVPG-LDLE--------------WYNHFTQQI----------LPHG
GLRG_07878      YATNDWIFKLQQQNMVPG-LDLD--------------WYNHFTDQI----------LPRG
CS.DMAT         YATNDWIFKLQQQNMVPG-LDLD--------------WYNHFTNQI----------LPRG
CS7             DPTLDYMEKVPVI--LPE-VDFS--------------WYRHFLAEL----------Y--N
CH063_0738      RPATEMLYSMTPQ--VPS-LDMS--------------LFNHFVAKF----------Y---
```

```
GLRG_11574    RPSMETLYSMKSQ--VPS-LDMT--------------LFNHFIAKL----------Y---
CS6           KPAMETLYSMMPQ--IPS-LDMT--------------LFHHFVAKL----------Y---
CH063_0204    DESQVVMSRLLRQ--GRGPVDGE--------------LWEFFKKHL----------HVEA
GLRG_11853    DESEVVMSRLLRQ--GRGPVDGE--------------LWEFFKKHL----------HVEA
CS8           DEPQVVMSRLLRQ--GRGPIDEE--------------LWEFFKKHL----------HVEA
CS5           KALRNILEGIASK--C-G-ADRK--------------WLDAFIDSS----------FLTT
GLRG_11423    KALRNTLEGMASA--C-G-ADRA--------------WLDAFLDSV----------FLTA
CH063_1280    KALRETLDSLASA--C-G-ADRA--------------WLDAFVDSV----------FLTA
CH063_0144    SWAARGFSCLASQ--MTS-MDTS--------------WYEHLGQAF----------RLKG
CH063_0491    NLDDEVLPLLERH--SAN-VDLR--------------WFCQFRDAW----------MMHR
CB2           KIVPAILPALRMV--SHG-ADTR--------------WFEQFMSVL----------FLTE
CH063_0654    KIVPAILPALRMV--AHG-ADTR--------------WFEQFMSAL----------FLTE
GLRG_02500    KILPAILPALRMV--SHG-ADTR--------------WFEQLMSAL----------FLTE
CS4           EIRRSLIPILSQA--PSD-IDLQ--------------WYKQVMNEW----------FVTP
GLRG_07955    EIGRSLIPILSQV--SSD-VDLQ--------------WYKQVVNEW----------FVTP
CH063_1116    EIGRSMMPVLSRV--SSD-VDLR--------------WYEQIVNAW----------FVTP
CB.DMAT       AFVDELRTAIET---SHE-PDLT--------------LFDALAKSL----------FVHD
CH063_0401    -----MLSQMKEE--IRD-IDLH--------------WADQFIENL----------FPNN
                   *

CH063_0589    HKPSERQRELVAKVGY-----------QTPTI------------------------LGF
GLRG_01200    LIQEGIPVLLPPGGHAIFLDMNGFFAGCARSYGEFASVGFTLELLKDYEIRACEAGPFGW
CS1           LIEEGIPVLLPPGG-------------CARSYGEFASIGFTPELLKDYGIRACEAGPFGW
CH063_0902    TRTKTVDRFVEETTPK------------AGTV------------------------VAL
GLRG_07878    NRTKTVDRFIEETTPK------------AGTV------------------------VAL
CS.DMAT       NRTKTVDRFIEETTPK------------AGTV------------------------VAL
CS7           PDRSVYAKEIEEGEPP-----------ATTLM------------------------HAV
CH063_0738    --DAAHHKYLETNERP----------VMTNVC------------------------LGF
GLRG_11574    --DSEWHRYLDTNERP----------IVTNVC------------------------LGF
CS6           --DSEWHRYLETKERP----------IVTNVC------------------------LGF
CH063_0204    KNAHDIVAKMATNEHM------------TTNT------------------------ISF
GLRG_11853    KHANEIVSKMAPNEHM------------TTNT------------------------ISF
CS8           KHANEIVSKMAPNEHM------------TTNT------------------------ISF
CS5           EEQAALVSKGAVAAAS-----DASHPLPQSAL------------------------VAF
GLRG_11423    EEELTLVGKRGMDAAN-GDAGAAGGPLRQNGF------------------------VAF
CH063_1280    EEEASLVGKAAAAAAGGGGPGAAGGPLRQNGF------------------------VAF
CH063_0144    QVEVDTANSQSRPGMW-----------LPKVF------------------------LGV
CH063_0491    HHADAAAAKEAMKQHA-------NVKIPSQLF------------------------LGY
CB2           SETRAALARIPKGTRA------------PTSF------------------------LAF
CH063_0654    SETKAALAKLPKGTRA------------PTSF------------------------LAF
GLRG_02500    SETRAALARLPEGTRP------------PTSF------------------------LAF
CS4           EGAVAACQNMPPHVKR-----------IPQLF------------------------LAC
GLRG_07955    EEAVAARQNMPPHIKR-----------IPQLF------------------------LAY
CH063_1116    DEAAAARQNMPASVQR-----------IPQVF------------------------LAY
CB.DMAT       PQEVAKVKAVVPLHLH---------TLIPNIA------------------------IAW
CH063_0401    EAEVELALKRAQTELP------PPLDHALTFN------------------------MAL
                                      .

CH063_0589    DINPKITELAPALLPVMVKSYFPPCFVTEDRG----FTRF--RSLALGVR----------
GLRG_01200    EWDKKGAIAQDR-DSIP-NLVRFAVPRYVMSD----RHIEYTATAISALHKRRHTIPGAR
CS1           EWDKKGAIAQDR-DSIP-NLVRFAVHRYVMSD----RHIEYTVAAISALHKRRQNIPGVR
CH063_0902    DIE-------KS-GPVM-KIYIYPGLKAEELG----ITNL--ELVEQSIR----------
GLRG_07878    DIE-------KS-GPVM-KIYIYPGLKAEELG----ITNL--ELVEQSIR----------
CS.DMAT       DIE-------KS-GPVM-KIYIYPGLKAEELG----ITNL--ELVEQSIR----------
CS7           EYN-------RH-ASFGLKSYFLP--RKLFQG----GDPATLEEWDAAIV----------
CH063_0738    EFL-------GH-DILP-KAYFFPR-KLGHVG----ITPM--AVWEDAIA----------
GLRG_11574    EFK-------GS-EILP-KAYFFPR-KLGQSG----LTPM--RVWEDAII----------
CS6           EFQ-------DTNDILP-KAYFFPR-KLGQAG----LTPM--EVWEDAIA----------
CH063_0204    DLE-------GE-LPAP-KVYFYPIPISLLQN----NHAG--EIITDVIA----------
GLRG_11853    DLD-------DE-QPAP-KVYFYPIPISLLAN----NHAG--EIITDVVA----------
CS8           DLD-------GE-QPVP-KVYFYPIPISLLKN----NHAG--EIITDVIA----------
CS5           DLHVNTDKGGKA-ANTM-KTYLFPQYKALITG----QKTV--DTTDSIVR----------
```

```
GLRG_11423      DLEADMDENEKA-VTVM-KTYLFPQLKAIATG----QKTV--NTTDSIVK----------
CH063_1280      DLEANMDESGKA-ATVM-KTYLFPQLKAIATG----RKTV--DTTDSLVR----------
CH063_0144      DFK-------GA-DRML-KCGFCPLLKLSAMGAKWDRLADHNKFVLDVVR----------
CH063_0491      DLV-------GP-KAQL-KAYFLPIFKHFATG----RSTD--ALAAEMIR----------
CB2             DLD-------GD-KTIF-KAYFFPTLKHIATG----ESAE--NITFDAIR----------
CH063_0654      DLD-------GD-KAVF-KAYFFPILKHIATG----ESPE--NITFNAIR----------
GLRG_02500      GLD-------GD-KTVF-RAYFSPILKHVATG----DSTE--SITFNAIR----------
CS4             DVK-------RS-KRIL-KAYLFPALKHSVMG----ATTS--DFVFDMIS----------
GLRG_07955      DMI-------RS-KRML-KAYLFPVLKHFATG----ATTS--DLVFDLIP----------
CH063_1116      DMK-------RA-KRVL-KAYLFPVLKHFATG----AATS--DLVFDLIP----------
CB.DMAT         DLV-------GP-IKKL-KLYHNPQAKKLATG----RTGN--EIVISSIR----------
CH063_0401      DLS-------GA-TRKM-KAYMFPMAKNLATG----RHRDARDAGFDAIR----------
                   . .   .      :

CH063_0589      ---------QLPDIGSHP-NILLGLK---------------------------MIEDFVA
GLRG_01200      ITRGEELRMRVFQSGLEPVPVTSTADDSATSPMPVSTFLSEAKKDVKKLHATLELRREAW
CS1             ITRGKELRMRVFQSGLEPVPVKTSTN----------------------LHAALKLDREVW
CH063_0902      ---------SLPAEQFQSLQAEPLFD---------------------------FLRE---
GLRG_07878      ---------NLPAEQFQSLSAEPLFE---------------------------FLRE---
CS.DMAT         ---------NLPAEQYQSLNAEPLFE---------------------------FLRE---
CS7             ---------KLSPDSKHP-GRDALMH---------------------------FLAT---
CH063_0738      ---------TAVPGS--P-SMXTVFS----------------------------------
GLRG_11574      ---------TAIPQS--P-TVATVFS---------------------------FVQ----
CS6             ---------TAIPQS--T-TMENVFS---------------------------FVK----
CH063_0204      ---------QLP-----V-NLEPSFN---------------------------PVRDFVY
GLRG_11853      ---------QLP-----V-NLEPAFN---------------------------YVREFVY
CS8             ---------QLP-----V-NLEPAFN---------------------------YVRDFVY
CS5             ---------VLAEGN--Q-EMLAAWE---------------------------LIKTFLN
GLRG_11423      ---------RLAGGD--K-DMLAAWE---------------------------LVKTFLI
CH063_1280      ---------RLAQGD--K-QMLASWE---------------------------LLKTFLV
CH063_0144      ---------GLPGCG--A-NMSAALG---------------------------MLERYLV
CH063_0491      ---------SLEPFG--P-ELEDQVR---------------------------QMEQFL-
CB2             ---------SLSPGG--A-DLTSSAN---------------------------AAEEYF-
CH063_0654      ---------SLSPGG--A-DLARSAN---------------------------ATEAYF-
GLRG_02500      ---------SLSPGG--P-DLAPSAH---------------------------ATEEYF-
CS4             ---------KLQSFG---------------------------------------------
GLRG_07955      ---------KLQPFG--D-KLAMPAR---------------------------KLQKYL-
CH063_1116      ---------KLQPFG--D-KLAAPAA---------------------------KLRKYL-
CB.DMAT         ---------ALAKYG--H-DFTEAMD---------------------------ILERYVM
CH063_0401      ---------NLKPHG--D-KLVPAVD---------------------------FLDRY-W


CH063_0589      ----------------------------DNPQYEGCGRG---LSTDFVPAG--Q------
GLRG_01200      ELIEAALDISMQGWGKLTVLDFPSGWRPDVCLNLCPIEYSVAIERNTGKPQ---------
CS1             ELLQAALDVSMQGWGNWTVSECPNGWRSDASLSLCPFEYSVAIERDTGKPQ---------
CH063_0902      ----------------------------GTKKYDFETGI---LAIDCLAPR--D------
GLRG_07878      ----------------------------GTKKYDFETGI---LAIDCLAPK--D------
CS.DMAT         ----------------------------GTKNYDFETGI---LAIDCLAPK--D------
CS7             ----------------------------SPEGQLMKPNV---MGMDNVTPS--K------
CH063_0738      ------------------------------------------------------------
GLRG_11574      ----------------------------NDAPRLGLTLAPLWLGIDVVNPV--E------
CS6             ----------------------------NDAPTLGLTLAPLWLAIDVVNPA--K------
CH063_0204      RYKRQ-----------------------RDNEYILRLEL---IAFDAIRPA--D------
GLRG_11853      RYKKQ-----------------------RNNESILRLEL---IAFDAVRPV--E------
CS8             RYKKQ-----------------------RNNEYILRLEL---IAFDAVRPV--D------
CS5             ----------------------------TNSGKKVMLDF---LAIDCLAPS--KEARSKR
GLRG_11423      ----------------------------TDGGEKINLDF---LAIDCLAPC--KE-----
CH063_1280      ----------------------------NDGGEKMNIDF---LAIDCLAPS--KE-----
CH063_0144      QDREHSAGAQGQGDVDDGRKDASSMCQACAKANRPKPFFNLVSVDCADPSNGK------
CH063_0491      ----------------------------AGCSYPHFIDM---VGIDCCDPA--K------
CB2             ----------------------------KNAPFPIPVEM---ISLDCIDPAAG-------
CH063_0654      ----------------------------RNAPFSIPVEM---IALDCIDPAAG-------
```

```
GLRG_02500      ----------------------------KNASFPIPVEM---IALDCIDPAAG-------
CS4             ------------------------------DKLPCIVES---MAIDCFDPS--K------
GLRG_07955      ----------------------------ATCKEPCLVEM---MAIDCVDPS--K------
CH063_1116      ----------------------------AGCREPCLVEM---MAIDCVDPG--K------
CB.DMAT         ----------------------------KLNAEQLELII---IGLDAADPNLPT------
CH063_0401      ----------------------------DTCPERLILDM---IGMDCIDPS--K------


CH063_0589      ARLKV---------------------------YLRYWG-DSFDEIWDYY-----TLGGRI
GLRG_01200      LRFLVEAQADAATLEAHQKSALQLTADIQSH-YPQTVSLGRLNAVKDLFFPPASQVEGIF
CS1             LRFMVEAQADAATLEAHQKSALHLTADIQSG-YSQIVSLDRFNVVKDLFFPPA-QAEGTF
CH063_0902      ARIKV---------------------------YIRAKH-TTFDYMMDCI-----TLGGRL
GLRG_07878      ARIKV---------------------------YVRAKH-TTFNYMMDCI-----TLGGRL
CS.DMAT         ARIKV---------------------------YVRAKH-TTFDYMMDCI-----TLGGRL
CS7             SRLKM---------------------------YFTSRH-TCFNSVREII-----TMGGLR
CH063_0738      --------------------------------FVRTTP----------------XXSG--
GLRG_11574      ARLKF---------------------------YCVESR-TSFDSVKSIM-----TMGGII
CS6             ARLKF---------------------------YCVEPQ-TCFNSVKSIL-----TMGGII
CH063_0204      SRFKL---------------------------YLRTKE-TCMARVEEVY-----SLGGAL
GLRG_11853      SRFKL---------------------------YLRTKE-TCLARVEEVY-----SLGGAL
CS8             SRFKV---------------------------YLRTKE-TCLARVEEVY-----SLGGAL
CS5             PRLKI---------------------------YFSTNF-KSLAAVREVV-----TLGGLL
GLRG_11423      PRFKV---------------------------YVSTRF-KSLAAVREAF-----TLGGLL
CH063_1280      PRFKV---------------------------YVSSHF-KSLAGVREAF-----TLGGLL
CH063_0144      GRVKL---------------------------YTRTSC-NAFACIRDAV-----TLGGRL
CH063_0491      ARIKV---------------------------YVRVNN-VSRACVKYFL-----TMGGRL
CB2             ACAKL---------------------------YARTRS-NAFSVVRHVM-----TLGGQI
CH063_0654      ARAKL---------------------------YARTRS-NAFNVVRDVM-----TLGGQI
GLRG_02500      ARAKL---------------------------YARTRS-NAFSVVRHVV-----TLGGQI
CS4             ARVQV---------------------------YARRIS-NSKCVLVDVF-----TLGGTQ
GLRG_07955      ARVKV---------------------------YARTTS-NSKSVLADVF-----TLGGTQ
CH063_1116      ARVKV---------------------------YARTAS-NSRSVLADVF-----TLGGAQ
CB.DMAT         TRVKVCHFSKSFHSDGPRTPESFLADTLHLQPYGIVSEANSWETVKNIY-----TLGGQV
CH063_0401      ARIKI---------------------------YAHVSTRNSWDLIRHVS-----TFGGQA
                                                :                        *

CH063_0589      PIADLEDDKEKLRDLIQLSRGSE-------------------------------------
GLRG_01200      AAWHSFVAGHDQRPEWKIYLNPNARGKLQAAAVVREAFERLGMPQAWILLEKVLLRGGRL
CS1             AAWHSFAADNDRRPEWKIYLTPNVRGKSNATGVVREALERLEMPQAWTHLDKVLLRGGRL
CH063_0902      --DMSGEAIEDLKDFWRIFL----------------------------------------
GLRG_07878      --DMSGEAIDDLKDFWRTFL----------------------------------------
CS.DMAT         --DMSGEAIDDLKDFWQAFL----------------------------------------
CS7             --HVSESSLQDLRSLILAILGLPADYPEEDEF----------------------------
CH063_0738      ------------------------------------------------------------
GLRG_11574      -----EIASDMLAMIWDLMKAVCDLPDD--------------------------------
CS6             -----DIAPNMLAMIWDLMKAVCDLPDD--------------------------------
CH063_0204      KGPEIDNGLDLIRSFYMHVLGT--------------------------------------
GLRG_11853      KGPEIDHGLGLIRSFYLHVLGT--------------------------------------
CS8             KGPEIDHGLNLIRSFYLQVLGT--------------------------------------
CS5             --PRSSANMDFLSEAWPLLMDMEDIP----------------------------------
GLRG_11423      ----PSSSADFLSQVWPQLLDMEDIP----------------------------------
CH063_1280      ----PRSSADFLPQVWPQLMDMEDVP----------------------------------
CH063_0144      TDEDTLEGLRRLRSVWHLLLNDP-------------------------------------
CH063_0491      DDERTRRMLADMDRCWHHVIDEE-------------------------------------
CB2             NDEATLESLDALRSIWHLLHNEP-------------------------------------
CH063_0654      HDETTLEGLDILRSIWHLLHNEP-------------------------------------
GLRG_02500      SDEATLEGLRGLRCIWHLLHNEP-------------------------------------
CS4             TDEATLKGVEAAEKVWHLLLDGP-------------------------------------
GLRG_07955      TDEATLKGVEAAEKVWHLLLDEP-------------------------------------
CH063_1116      TDEATLRGVETAEKVWHLLLDEP-------------------------------------
CB.DMAT         VNKERMKGLEILHSIWDLMRCHRGDP----------------------------------
CH063_0401      TDPDRLRGLEILHSLWDVLRNEQSNP----------------------------------
```

```
CH063_0589    ---------------------------------------YP--ADKIRPETDAEKKRRA-
GLRG_01200    TFLALDLCPGPQARVKAYVQYDNVGAAELAEAAALVAPDTASASEIRRFCSTLSGGSEGP
CS1           MYLSLDLSPSSQARVKVYVQYDNVGAAELAEAAALVAPETATAPEIQKFCSTMSGGSEGP
CH063_0902    -------------------------------------------ADAPDVLPEEAPGRAS-
GLRG_07878    -------------------------------------------ADAPDILPADAPGRAS-
CS.DMAT       -------------------------------------------ADAPDILPADAPGRAS-
CS7           ---------------------------------------TPTSAKSGDKWDDFDSLCEG-
CH063_0738    ------------------------------------------------------------
GLRG_11574    ---------------------------------------FPQDKDLPRAPQYNPSTNGID
CS6           ---------------------------------------FPQDKSLPRAPRYNPSTDGID
CH063_0204    -------------------------------------------TAAEEDLPRSTHRTAG-
GLRG_11853    -------------------------------------------TAPEEDLPHSTHRTAG-
CS8           -------------------------------------------TAPEEDLPHSTHRTAG-
CS5           ---------------------------------------QAEIEDLEKPLNDPTSHYQG-
GLRG_11423    ---------------------------------------QAEMEDLEKPLNNPDSHHQG-
CH063_1280    ---------------------------------------PAAMEDMEKPLSNPDSHYKG-
CH063_0144    ---------------------------------------VSQHDDEYSRTVAVESHRQG-
CH063_0491    ---------------------------------------AGYPDEQDKKPKDPATFHKG-
CB2           ---------------------------------------EPYGDDFDKQPRLEGTLHKG-
CH063_0654    ---------------------------------------EPYGDDFDKQPKLLQTLHKG-
GLRG_02500    ---------------------------------------EPYSDDFDKQPKMENTIHKG-
CS4           ------------------------------------------------------------
GLRG_07955    ---------------------------------------QGMAPDQRKDARDMKNLHKG-
CH063_1116    ---------------------------------------EGMAPDQRKECRDARTLHKG-
CB.DMAT       ---------------------------------------LP--DDYHKPKNDASSTRGV-
CH063_0401    ----------------------------------------G--EDYDKPMRHPTSFLGS-


CH063_0589    -IFGTKPTSLYFSLTPEKP--YPIPKLYFYPAFQAPNDEAIAQGIDAWLKRYDWYDGGK-
GLRG_01200    YKGKAPMTCFGFTASDHGNGITAESAVYFPVHDYANDDAEVRLRLEKYLSDTGSGESGI-
CS1           YKGKGPMTCFGFTASGHGHGVTTESTVYFPVHDYATNDAEIRLRLEKYLRDTGSGESGI-
CH063_0902    -------PGFYYTLGCGRA---ISPKVYISPNYFSKNDVDVLGRLRTFFATRRSDSMMD-
GLRG_07878    -------PGFYYTLGNGRE---ISPKVYISPNYFSKNDVDVLARLRQFFSSRRTDSMMD-
CS.DMAT       -------PGFYYTLGSGRQ---ISPKVYISPNYFSKNDVDVLARLRQFFSSRRTDSMMD-
CS7           -------FTYFFDIAPTSG--VPDVKFYLPTRKYGADDLTIARNLVAWLRERGRDAYCN-
CH063_0738    ------------------------------------------------------------
GLRG_11574    AIGLWGTFAYYFDIGLGRG--LPDVKFYIPVCHYGTNDQTIASATANWMKKNSRGQYVD-
CS6           TVGLWGTFAYYFDIGIGRE--LPDIKFYIPVCHYGTNDQAIASATANWMKRNGRGQFVN-
CH063_0204    -------IIFNMELKHNSP--EPLPKVYIPVRHYGGTDLRIAQSLSNFFRACGLTHMAD-
GLRG_11853    -------IIFNMEMKHNSP--APVPKVYIPVRHYGGTDLRVAQSLSNFFRAWGLIHMAD-
CS8           -------IIFNMEMEHNSL--APVPKVYIPVRHYGGTDLRIAQSLSNFFRAWGFTHMAD-
CS5           -------VGFTIAVAPGEA--IPQIKTCVPIWQFARDEARIIACCQRLSQKHGP--TGKF
GLRG_11423    -------VGFAFSVVPGQA--VPQIKMYVPMWQFARDEARIVECYRRVLQTQGT--MGDY
CH063_1280    -------VGFAFSVVPGQA--VPQIKMYVPMWQFAHDEAKIVECYQRVLQTQGT--MGDY
CH063_0144    -------IDINWEISDQLP--APQAKVYVPVSMFHENDLGANRNLSEAFRKLNWLEWAEG
CH063_0491    -------IVMAITLSPALNDGVVRTRPYCSWSNYQSSDWGAVGNFAMILKELGMEKEAE-
CB2           -------VCYGFELKPGAK--WPEVKAYVPLWQYADSDAVIASNLANAFHLRGW-PVAE-
CH063_0654    -------ICYGFELKPGAE--WPEVKAYVPLWQYAGSDAVISSNLAKAFRSRGW-PVAE-
GLRG_02500    -------VCYGFELKPGAE--WPEVMVYVPLWQYADSDAVIASNLADVFRSRGW-PVAE-
CS4           ------------------------------------------------------------
GLRG_07955    -------ICFAFELKQGVE--RIDIKAHLPWGQTARSDFQTIENFAQVLRNVGLDSAEK-
CH063_1116    -------ICFAFELRPGAE--RIDIKAHLPWGQTSRSDARTVDNFSQALRNLGWDESAD-
CB.DMAT       -------LTPSFEVAPGQT--LPNVKLYLSQWQFGKSDREIAECTVEIFRKLGWQREAD-
CH063_0401    -------IMFSFEIVPGRQ--IPEVKIYVPMWQYSPSDGHVAKNLISAFQKLGWHDVAE-


CH063_0589    TLEQMVANVL--------------------------------------------------
GLRG_01200    QTLRAYQGAT---------------------EAAANRPLQDGRGI-HAWVGLKTTRHRGS
CS1           QTLRTYQRGI---------------------DAAANRPLRDGQGI-HAWVGLKMTKNRGS
CH063_0902    NYEQALKDIF------------------------GSKTLESRCGS-HYYVGCALHK-GQL
```

```
GLRG_07878     NYEQALKDIF------------------------GSKTLESRCGS-HYYVGCALHK-GHL
CS.DMAT        NYEQALEDIF------------------------GRKTLESRCGS-HYYVGCALHK-GQL
CS7            RYLDMLDKLA------------------------VHRGLENGKGM-HAYISYQCTEKGEP
CH063_0738     ------------------------------------------------------------
GLRG_11574     AFWDTMHKII------------------------SHRKLDESRGA-QMWLSMMVKQ-GEL
CS6            AFWDSMHKII------------------------SHRKLDESRGA-QMWLSMMVKR-GEL
CH063_0204     TYVDAVQKAF------------------------PTQDFSTTIGR-HSYVGLSYNK-NGP
GLRG_11853     T---------------------------------PNQDFSTTIGR-HSYVGLSYNK-NGP
CS8            TYVDAVQKAFLPVQSLCHSVYEHKLTSLLTVPCSPSQDFSTTIGR-HSYVGLSYDK-NGP
CS5            NFGAAIQDAL--------------------------------------------------
GLRG_11423     DINAAIQGAL--------------------------------------------------
CH063_1280     DIGAALEGA---------------------------------------------------
CH063_0144     RYEKMLGRVF------------------------PEADFATSHV--NTWISYCYYKGRGS
CH063_0491     TYVRG-----------------------------------------HTATASSV------
CB2            QYEDTDSKLI-------------------MDVVSPRSNMEKTTGT-HAYISFAFSKKNGA
CH063_0654     TYEDAMPCCF------------------------PRSNLGKTTGT-HSYISFAFSKQKGA
GLRG_02500     KYEETIRRSF------------------------PRSKLEKTTGT-HSYISFAFSKKNGA
CS4            ------------------------------------------------------------
GLRG_07955     FLKGALETAK-----------------------LPGRDYTKPGGL--SYVSYNYNQ-NGP
CH063_1116     KFYRGASAAA------------------------K-------------------------
CB.DMAT        SYFDFLRDAL--------------------------------------------------
CH063_0401     NYLPNLRRTF------------------------PGADLDSPSSVLHSNLSYSYSPATGA


CH063_0589     -------------------------------------
GLRG_01200     VITFYLASEILVHCL----------------------
CS1            VITFYLASEVLTHCL----------------------
CH063_0902     RVVTYLSPQSFDCEKDAIQDRKP--------------
GLRG_07878     RVVTYLSPQSFDCEKDRIQGVKS--------------
CS.DMAT        RVVTYLSPQSFDCEKDRIQAGGA--------------
CS7            DIKSYISPELYHKARYSAA------------------
CH063_0738     -------------------------------------
GLRG_11574     QVTTYIAPEGFHPKRGQGHQLSHRAVVERVTGSC---
CS6            QVTTYISPEGFHPNRVQRQDISHRAIAERVTSSH---
CH063_0204     YITMYYNTMTYSAGDERNADGKLVGPAAMKQRHLLD-
GLRG_11853     YITMYYNTMTFSAGNERDVNGKLVGPAAMKQRHLLD-
CS8            YITMYYNTMTFSAGNERDITGKLVGPAAMKQRHLLDQ
CS5            -------------------------------------
GLRG_11423     -------------------------------------
CH063_1280     -------------------------------------
CH063_0144     YMTMYHSPPW---------------------------
CH063_0491     -------------------------------------
CB2            YMTMHYSIRPQTSLAF---------------------
CH063_0654     YLTIYYSIRPQDSLMAF--------------------
GLRG_02500     HMTMYYSIRPQDSLVAV--------------------
CS4            -------------------------------------
GLRG_07955     YLSSYFSPKIQDN------------------------
CH063_1116     -------------------------------------
CB.DMAT        -------------------------------------
CH063_0401     YISVYYAVSGNATTATGQAE-----------------
```

PART 2: A-domains of NRPS and Hybrids sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
GLRG_09842      ---------------IAVIGIGCRFPGASSFP-ELADALH-----------EPAD-----
CH030B_PKS      ---------------IAIIGTGCRFPGGSDSPSKLWELLK-----------DPRD-----
CSP24           ----WAQRTPDSVAVKDMNGEILSYSQLMDRAGSVSIALS-N---AGV---EASA-----
GLRG_09715      -VARSAAQFSDSIAIKDDEGNELKYSDMIKTTNQICRGLQ-A---SKI---QPGS-----
CSP6            ---DHAAESHPDHIAVKDGARAYTYEDMMRSVDSIATSLR-A---MGVPRSARP------
CH063_0617      -----EAYPDRVALRDGSR--SYTYREVAAETDRIAAVLL-E---ATA---VSK------
CH063_0392      -IDRIAAQQPNSVAISNGFGSTATYGAMQRCIQTIAQTILGN--TQGR---SAG------
CS.P.NRPS       IEALAEEHPHSLALKSPHGGPSMTYRDMIAKSRAIGAALV-A---VGC---VPGA-----
GLRG_10367      IEALAEEHPHSLALKSPHGGPSMTYGDMIAKSRAIGAALV-A---VGC---APGA-----
GLRG_07434      ----------EKVAVTDSAGASMTYAQLSSRMSTIASELL-R---QGV---QPGS-----
CSP1            -FEVSAEKHKRNTALIHADGTSLTYAQMQQRIDTIAAALE-E---NSI---KPGD-----
CSP10           --------YPGKLAIKDGNGTILTYADMSNRVAAISSGIS-R---LQV---RPGS-----
CSP20           -----ADKYASKAALRDGHGDGLTYAQMVARVEAIAAYLI-K---EGI---GRRK-----
CH063_0306      -IDDMADKHASKAALRDGQGGGLTYAQMAARVNAIAARLE-E---AGI---GSSN-----
GLRG_05027      -----------------------TYAELDQETQALDRLRR-H---YGV---GRDG-----
CH.NRPS1        -----------------------TYAELDQETQALADRLRRH---YGV---GRDS-----
GLRG_08225      -----------------------TYRQVCRLVSRLATFLV-Q---WGV---RPGV-----
CHN24A          -----------------------TYAELHKRSDKLASWLR-Q---RRL---PAGS-----
CSN9D           -----------------------TYAQLNQQSDRLAAWLR-R---REL---APGS-----
G0043A          -----------------------TYTQLDEQSDCLAAWLR-E---RTF---APGT-----
CSN9G           -----------------------TYAQLDRQSDELADWLR-R---RQHRLPGPEAGGRGP
CH.NRPS2        -----------------------TYAELDRRSDDLARWLR-R---RRL---DPET-----
CHN29A          -----------------------TYRELDRESSRIAAYLQRY--HRQM---APES-----
CSN9I           -----------------------TYSQLDQQANRIAAYLG-N---RNM---APET-----
G0046A          -----------------------TYRQLYQQSNQIASYLG-Q---RQM---APET-----
CSN9C           -----------------------TYAELDRQSDQLASWLVHD---RGL---APET-----
CHN23A          -----------------------TYSELDRQSDQLASSLV-R---RGF---APET-----
CHN210A         -----------------------TYGQLDEKSDRLAAWLR-K---KHM---STET-----
G0047A          -----------------------TYAQLSETSDRLAAWLR-L---RKM---EPET-----
CS.NRPS         -----------------------TYAQLNQTSDHLAHWLR-L---QQM---ETET-----
CSN9H           -----------------------TYAQLDEQSDRLAAWLR-RDGQQHHRQRPAES-----
CHN28A          -----------------------TYAELDEQSDRVASWLR-R---RRL---APET-----
CHN22A          -----------------------TYTELDQQSDRLSGWLR-Q---RRL---APET-----
CSN9B           -----------------------TYAELDRESDTLAGWLR-R---RRL---APET-----
GLRG_00469      -----------------------TYAELDRQSDKLAGWLR-Q---RRL---APET-----
G0044A          -----------------------TYSQLDEKSSILAAWLR-N---QHL---AAET-----
CSN9E           -----------------------TYSQLDYESGVLAAWLR-----KQRQQKQEGHSTAAT
CHN25A          -----------------------SYAQLDRQSDILASWLQ-K---QKL---AAET-----
G0045A          -----------------------TYSELDRQSDKFAAWLR-L---RGL---PDET-----
CHN26A          -----------------------SYSQLDQQSDTLAAWLLRR---RDM---ADEA-----
CSN9F           -----------------------TYSELDRRSDTLAAWLR-A---RGM---ADEA-----
CH.NRPS3        -----------------------TFRELNERGNQVASVLA-----PHA---KVGQ-----
CSN1C           -----------------------TFRELNERGNQVANVLA-----PYA---QTGQ-----
GLRG_06542      -----------------------TFRELNERGNQVANVLG-----PYA---QTGH-----
CHN32A          -----------------------SFGELNSVSNRIARFLR-G---QGA---REGD-----
CSN1B           -----------------------SFNTLNSLSNRIGRYLQ-E---QGA---KEGY-----
G0652A          -----------------------SFSTLNSLSNRIARYLQ-E---QGA---KEGD-----
CSN1A           -----------------------SYSQLHHAADRLASRIICSVDKVST---SEDF-----
G0651A          -----------------------SYSQLHHAADKLASRVVCSVEKASA---SEDF-----
CHN31A          -----------------------SYSQLHRAVDSLASRISRS---LGKASTLEHF-----
CHN41A          -----------------------TYAELDDASDRLRDRLR-Q---AGL---APGD-----
CSN7            -----------------------SYARLNRRANHVATMLV-Q---RGL---RPGD-----
CHN62A          -----------------------SYAQVERYSNLFSQHLL-A---AGV---NKGD-----
CSN2A           -----------------------SFKGLEDASNHLGKQLI-Q---KGV---SPAS-----
CH.NRPS9        -----------------------TFSQIDGLSTSLCQHLVHR---LGL---SRGD-----
CHN91A          -----------------------TYRDLGKYSGVICSRLSQNIEMRSAFATGEEDRRHP-
CH.NRPS5        ------------------------------------------------------------
CH.NRPS6        -----------------------TYGELERLSSALAVSLR-H---LGV---RREV-----
```

```
CHN21A         --------------------------------------LS-----LKL---DAGH-----
CSN9A          -----------------------SHGELDAVSSRLAGYIQ-S---LKL---DAGH-----
G0041A         -----------------------SHAEMDGVSSRLAGHIQ-S---LKL---DPGH-----
CSN4B          -----------------------SFSELDIMASRLARILV-E---LCV---VPET-----
CHN102A        -----------------------SFAELDVMASRLARVLV-E---LQV---GPEV-----
CH.NRPS7       -----------------------SYSQLQQCTKQLAHHLS-N---MGV---GPEV-----
CSN6A          -----------------------SYSQLQNHADKLANHLR-S---IGA---GPEA-----
G0681A         -----------------------SYTQLQDHADKLANHLS-S---IGV---GPEV-----
CSN2B          -----------------------TYRELDDLSRNLAHHLK-S---LGI---GPEK-----
CH.NRPS4       -----------------------TYGELDRATGCVAQYLRDS---CGV---VPDG-----
CH.NRPS11      -----------------------------------AVRLQ-S---LGA---GPGE-----
CSN8B          -----------------------SYAELKNLTMRLASRLR-I---LGA---RSGA-----
GLRG_00920     -----------------------SYTELNDLSRRLVSRLQ-T---LGT---RPGA-----
CSN2D          -----------------------TYAELDFASSFLAQKLQ-E---IGV---GPEK-----
CHN44A         -----------------------TYAELTAASVRLASRLRDK---FGD---CRGR-----
CSN2C          -----------------------SYAEMDLWANQISQQIIAT---NLV---KPGD-----
CHN43A         ------------------------------------------------------------
CHN111A        -----------------------TYRELDEASTAVARRIL-A---LGI---GPGA-----
CSN8A          -----------------------SYRELDEAADVVATRIS-S---FGI---GPGA-----
G0091A         -----------------------SYRELDEAADVVVSHIT-S---LGI---GPGA-----
CSN4A          -----------------------SYAELGETSDMVAHELF-Q---HGV---GPEC-----
CH.NRPS10      -----------------------SYAELGEASDKVAHELC-Q---RGV---SPES-----
CHN72A         -----------------------SYATLESMASALGQRLV-D---SGV---SPGD-----
CSN6B          -----------------------------------------E---CGV---SPGE-----
GLRG_06838     -----------------------SYAALESMSSALATHLT-K---SGV---SPGE-----
CHN87A         -----------------------TYGQLEAEASRFSHRLR-R---AGVV--GLGA-----
CH063_1253     -----------------------TYGDVERLSTRLARHLV-H---RGV---KVGS-----
GLRG_08065     -----------------------TYGDVERLSTRLARHLV-H---RGV---KVGS-----
CSN5           -----------------------TYGDVERLSTRLARHLV-R---RGV---KVGS-----
CHN84A         -----------------------TCGELDGLVGRLSRHLVRQ---YGV---GPGM-----
CHN85A         -----------------------SYGELDSLAERLAGDIV-----AALD--TAGS-----
CHN81A         -----------------------SYGALDRLSDQLAIEII-R---RGV---NAGD-----
CH.NRPS8       -----------------------TYRELDQLAAELSYSLVSR---RGI---GPGS-----
CHN86A         -----------------------TFAELDRLTDSLASHLVHH---LGV---EPET-----
CHN83A         -----------------------TYGELDALSLRMSSRLA-Q---LGV---GPEV-----


GLRG_09842     ---------------LSQTIPSDRFSIDGFHHEDGKYHG-----NTNVRKSYFLRDGIRK
CH030B_PKS     ---------------LLKEIPEDRFSVEGFYHPKNSHHG-----TTNVRHSYLLDEDLRR
CSP24          ---P--------VGVLLDPNVDTIASMLGILRIGAAYVPLDVR-NPDERLLDIL-QESG-
GLRG_09715     ---R--------VAVLLTPDVWSICTLLAALQQGLVWVPLDTR-NARQRLAAML-TDCK-
CSP6           ---F--------CCLLIEPSADYIFAYLAVLKAGAVVLSLDTT-NHRERLAAIV-SDCK-
CH063_0617     ---AASTSPPVFCCLLFEPSADYVFALLAVLKVGLVAVSLDTT-NHAERLASIV-ADCK-
CH063_0392     ---L--------IGVFCEPSIDAICSLLAAMRVGFAYVPLDPT-LPSERLAALV-ESAR-
CS.P.NRPS      ---S--------VVVYQDSTPDWICSVLGTFSINGVCVPCDAG-TPVKRLADMV-ASSK-
GLRG_10367     ---S--------VVVYQESTPDWICSVLGTFSINGVCVPCDAG-TVVKRLADMV-ASSK-
GLRG_07434     ---K--------VAVFQHPSVCWIPSLLAILRIGAVYVPLDVN-TPVARLSSIV-RDCQ-
CSP1           ---S--------VGVFQESTTDWVCSMMALWKRGAVFVPLDPG-TKTERLAMVV-KDCK-
CSP10          ---V--------IGVFQEPTANFVCSMLAIWHAGCVYLPLDPA-TPMPRLESQV-YHCS-
CSP20          ---V--------VGVMQASTLDFVCSILAIWKVGATYTPLDPRLNSIDRLAAII-RECQ-
CH063_0306     ---I--------VGVMQASTLDFVCSILAIWKVGATYTPLDPRLNSTDRLAAIV-RECR-
GLRG_05027     ---L--------VGVLMNRSADYVIASLAALRAGGAFLVLELA-YPSGLLRDVI-QDAK-
CH.NRPS1       ---L--------VGVLMSRSADYVVASLAALRAGGAFLVLELA-YPSGLLRDVI-QDAK-
GLRG_08225     ---A--------VPVILDKN-----------KAGGCFVPIDAE--DRANVETFI-RQLA-
CHN24A         ---L--------VGVLSPRSCETIVALFAILKADLAYLPFDTN-VPTARLEDILSEVSG-
CSN9D          ---L--------VGVLVPRSCEAVVTLLAVLKANLAYLPFDNN-VPTARLQDVLAEVAAG
G0043A         ---L--------VGVLAPRSCEAVVTLLAILKANLAYLPFDNN-VPTARLEDILSEVPG-
CSN9G          GTIL--------VGVLAPRSCETVVAFLGIMKANMAYLPLDVN-VPIARLEAILSSVPG-
CH.NRPS2       ---L--------VGVLAPRSCETVVAFLGILKANLAYLPLDVN-VPSARVEMILSFVPG-
CHN29A         ---L--------VGVLAQRSCLTTAAFLGILKANLAYLPLDVN-APAARIGAILSAVPG-
CSN9I          ---I--------VGVLSHRSCQTIATFLGILKANMAYLPLDVN-VPDARVEAILSAVPG-
G0046A         ---M--------VAVLAPRSCETVATFLGILKANLAYIPLDVN-VPAARVEAIL-SSV--
```

```
CSN9C        ---L--------VGVLAPRSCEAIVALLGIMKANLAYLPLDVN-IPLGRLESILTSVEG-
CHN23A       ---M--------IGVLAPRSCEAIVALMGILKAGLAYLPLDVN-VPLGRLESIL-SSIEY
CHN210A      ---I--------VGVLAPRSCQTTVAFMAVLKSSLAYLPLDVN-VPAARLESILSAVPG-
G0047A       ---L--------VAVLAPRSCQTVVAFLAILKAGLAYLPLDIN-VPAARLESILSAVPR-
CS.NRPS      ---L--------VAVLAPRSCQTVVAFLAILKAGLAYLPLDVN-VPAARLESILSAVPG-
CSN9H        ---L--------IGVLAPRSCETIVAFIGILKANLAYLPLDVN-VPAARNETILSTVPG-
CHN28A       ---L--------IGVLAPRSCQTIVAFIGILKANLAYLPLDVN-VPAARNETILSTVPG-
CHN22A       ---L--------VAVLSPRSSLTILTFLAILKAGLAYLPLDVN-MPLARVEAIL-SAVK-
CSN9B        ---L--------VAVLAPRCCRTILTFLAILKANLAYLPLDVN-MPAGRIEAILSAVQG-
GLRG_00469   ---P--------VAVFAPRCCQTIVTFLAILKANLAYLPLDVN-MPAGRIETILSAVQG-
G0044A       ---P--------VALFSPRSCETIVAILGILKANMAYLPMDVK-MPAARLNSIL-STLE-
CSN9E        TTPV--------VAVFSPRSCETIVAFLGILKADMAYLPLDVN-SPVARLTSILSAVEG-
CHN25A       ---P--------VAVFAPRSCETIVAFLGILKANLAYLPLDVN-VPAARLTSILSAVDD-
G0045A       ---I--------VGVLAPRSCETTVALLGILKANLAYLPLDIK-VPLPRLEAILSSLQG-
CHN26A       ---L--------VGVLAPRSCDTIVAFLGILKANMAYLPLDIN-APVARLNAIFSTRPG-
CSN9F        ---L--------VGVLSPRSCDTIVAFLGIMKANMAYLPLDVN-VPAARLEAIL-SAV--
CH.NRPS3     ---T--------IAIHFDKCPEACFSILGILKAGCAFLALDPS-APKARKEFIL-KDSG-
CSN1C        ---T--------IAIHFDKCPEAIFSILGILKAGCAFLALDPS-APKTRKEFIL-KDSG-
GLRG_06542   ---T--------IAIHFDKCAEAIFSILGILKAGCAFLALDPS-APKTRKEFIL-KDSG-
CHN32A       ---I--------VGLIMEKSISLYAAMLGILKAGCAYLPLLPT-TPAERTKVIL-TQAG-
CSN1B        ---I--------VGLVMEKSISLYAAMLGILKAGCAYLPLLPT-TPAERVRVIL-TQAK-
G0652A       ---I--------VGLVMEKSIDLYAAMLGILKAGCAYLPLLPT-TPAERIRVIL-AQAE-
CSN1A        ---V--------VPLLVPQSPHLYIAILAILKAGGAFCPLNLD-APLERVRFIL-QDVS-
G0651A       ---V--------VPLLVPQSPHLYIAILAILKAGGAFCPLNLD-APPERVRFIL-QDVS-
CHN31A       ---V--------VPLLVPQCPHLYIAILAILKAGGAFCPLHLD-APPERVRFIL-QDVS-
CHN41A       ---G--------VLLCFEKSVWMSVAMLAVLKTGAFCSSLDPD-YPEHRVTQIA-EATE-
CSN7         ---C--------VALLLEKSVEAVACILGVLKAGCTYVPLSPD-NSAERNAFIC-RDPS-
CHN62A       ---F--------VPLLLERSPWAPVLMLAVLKLGAAFALLDLS-HPVQRLRTMC-TVLS-
CSN2A        ---F--------VLLSFTHSLWAVVSWLAVLKAGAACVFLDQR-QPAERTKQII-GTTS-
CH.NRPS9     ---H--------VTLLCDKSAWVVVAMLAVLRAGAACAFLSPA-DGLNRLRNMVTEQMR-
CHN91A       ---F--------VAICLEKSVWTIPIMMGVWQAGAAVVLLDPE-LPRYRLRKAV-QQVN-
CH.NRPS5     ------------------------------MKAGNAYVGLSAE-TPLSFLQECS-CIAD-
CH.NRPS6     ---F--------VPLVFDKSKWAIVALLGVLKAGGAYFFLNPT-HPVEYNKTLC-KPFD-
CHN21A       ---A--------VPLCFEKSKWMVAGVLAVLKAGHAFALVDPA-HPAARVAQIC-RQTS-
CSN9A        ---A--------VPLCFEKSKWMVAGVLAVLKAGHAFALVDPA-HPAARVAQIC-RQTS-
G0041A       ---A--------VPLCFEKSKWMVAAVLAVLKAGHAFALVDPA-HPAARVAQIC-RQTS-
CSN4B        ---K--------VAYCFEKSMYTVVSMLAILKAGGTIVPLDPA-HPPERKAFII-QSIA-
CHN102A      ---N--------VAYCFEKSMYTVVSMLAILKAGGTMVPLDPA-HPPERKSFVI-RKIS-
CH.NRPS7     ---L--------VPVIFEKSLWSVVAILAVVRAGGAVVALDPS-LPAARIRMIV-QDVK-
CSN6A        ---L--------VPVIFEKSLWSIVAMLGVIRAGGAVVALDPS-LPTARIRMIV-QDVK-
G0681A       ---F--------VPVIFEKSVWSVVAILGVIRAGGAVVALDPS-LPTTRICMIV-QDVN-
CSN2B        ---I--------VPLCFEKSKWAVVSMLAVLRAGACFVMLDPA-HPNVRVMSIA-EDVE-
CH.NRPS4     ---V--------VPICSEKSIWAVVAMLGVMRAGGCFVMLDPG-HPDSRLLSIV-EQVR-
CH.NRPS11    ---I--------VVLCFEKSLWAMVAMLAVAKSGAAFVHIDIQ-GAPKRTESVI-MQTK-
CSN8B        ---V--------VVLCFEKSLWAIVSMLAVARSGAAFVHIDPA-APFKRIESIV-GQTR-
GLRG_00920   ---V--------VVLCFEKSLWAVVSMLAVARSGAAFVHIDPA-APSKRIESIV-GQTK-
CSN2D        ---T--------TPFMFDKSKWAIVSMLAIWKAGGSFVPLDPK-SPNQRLQHII-QATS-
CHN44A       ---A--------FPILCRKSKWAIVGMLAIWHAGGYLVPMEPT-HPIQRLRLVA-GTVA-
CSN2C        ---F--------VGLCFDKSASAILAMIAVLKAGGAFLPLNPD-HPPARLRALL-KEAN-
CHN43A       ------------------------------------------------------------
CHN111A      ---F--------VALCFEKSIWYSVAMIGVLKSGNSFVPIDVS-NPTSRREEIL-QQLGI
CSN8A        ---Y--------VALCFDKSRWYSVALLGVLKSGNAFVPIDAA-NPISRRTEIL-QQLGI
G0091A       ---Y--------VALCFDKSLWYSVALLGVLKSGNAFVPIDVK-NPTSRRLEIL-QQLGI
CSN4A        ---T--------VALFFEKSIWNTVATVAVLKASGAFVPLDPS-QPNNRISYIL-DQVK-
CH.NRPS10    ---T--------VALFFDKSIWNTVATVAVLRSSGAFVPLDPS-QPDSRISYIL-GQVK-
CHN72A       ---M--------IPMCFEKSVWTVVALLGALKAGAVFVPLDPS-QPISRLREVV-AQVR-
CSN6B        ---M--------VPLCFEKSVWAVVAMLGVLKAGAAFVPLDPS-QPISRLRDVA-AQVK-
GLRG_06838   ---M--------VPLCFEKSVWTVIAMLGVLKAGAAFVPLDPS-QPISRLRGVA-AQVK-
CHN87A       ---M--------VPLCFEKSKWAVVAMLAVLKAGAAFVPLDPSQAPDRRERVLVRTRAR-
CH063_1253   ---T--------VPMCFEKSRWTTVALLAVMKAGAAFALTDPS-QPEARLRTIV-EQTG-
GLRG_08065   ---T--------VPMCFEKSRWTTVALLAVMKAGAAFALTDPS-QPEARLRTIV-EQTG-
CSN5         ---T--------VPMCFEKSRWTTVALLAVMKAGAAFALTDPS-QPEARLRTIV-EQTG-
```

```
CHN84A       ---V--------LPLCFEKSMWTVVAFLAVLEAGAGFVLLDPA-LPEARLRLMV-SQVN-
CHN85A       ---V--------VPVCSEKSAWAVVSLLAVLKSGNTFVLLDPS-QPESRLSAVV-RQTG-
CHN81A       ---L--------VPLCFEKSAWTVVAILAVLKAGGAFVLLDSS-HPVERLRTIV-AQAS-
CH.NRPS8     ---L--------VPLCFEKSKWAVVAILSVLRTGAAFVLMDPS-QPEERLRSIV-RQLD-
CHN86A       ---L--------VPICFDRSKWAVVAILGILKAGGAFVPLDPA-QPPARREQVL-SQTG-
CHN83A       ---L--------VPLCFEKSLWAIVAMLAVLRAGGAFVPIDPD-AARGRRDGIL-KQTD-


GLRG_09842   FDNNF---FNITAV---------------------EASAMDPQQRLL----------LET
CH030B_PKS   FDASF---FGIKAT---------------------EANSIDPQQRML----------MET
CSP24        ---AT---IVLFHS---PTAK--------------RAKTLLGGLTNT----------QHM
GLRG_09715   ---PR---HIIHSDATGVLAE----------ELVAAARPENPELEPLRL--------AEL
CSP6         ---PQ---AIIHCK---AMSN--------------LAQWLGQSCDVA----------SVM
CH063_0617   ---PM---VILHHG---ATKD--------------LAV-RLGEPHEG------IS--TLI
CH063_0392   ---PL---LILAHDETIDRAK--------------EPSSFVGHIPLI----------NVS
CS.P.NRPS    ---AT---VLLVDA---STAE----------QAAGALSTTISDRERVRILRV-----DRI
GLRG_10367   ---AT---ILLVDA---STAE----------QAAGALGTTISDEERVRILRV-----DRI
GLRG_07434   ---PA---SILVHE---FTIG--------------EAEKLDTEPNQIFI--------NVS
CSP1         ---PA---AILIDD---ATKG--------------LHTELGATETTVFI--------DVS
CSP10        ---PH---LILVDN---RFAD--------------AAASFGTPVFNV----------PLW
CSP20        ---PA---CVLIDA---SSRS--------------LFDQLDSAAVLI----------DVE
CH063_0306   ---PT---HVLVDA---STRS--------------LFAQLESTAVPV----------DVT
GLRG_05027   ---PQ---VVITQN---AHVS--------------HIKADVPIIVY-----------DEP
CH.NRPS1     ---PT---VVITQN---AHVN--------------HIKADVPVIVY-----------DEP
GLRG_08225   ---PK---IVVATD---LGWK--------------SLEFIVEGVIVV----------DDS
CHN24A       ---EK---LVLVGA---SVPM---------------PDLGVPDVKLVRIV-------DAM
CSN9D        HGRSD---LPLLVG---ATVP--------------MPDLKMPGVQLVRI--------AEA
G0043A       ---HK---LLLVGS----TVS--------------MPDLKMPDVQLVRIVEA-----TNG
CSN9G        ---RK---LVLVGA---GVPT--------------DEVGLLPDVELVRI--------RAA
CH.NRPS2     ---RK---LVLVGP---DVST--------------DTT--LSDVEFVRIGA------ALN
CHN29A       ---HK---LVLVGA---GVPT---------------LPAESRDVEQV----------QIS
CSN9I        ---HK---LVLLGD---NVPV----------VTLTTTSGSAGEIEQVQISHALGQKAAAA
G0046A       ---PGHK-LVLLGT---GVSP--------------LPTESAEMEQVQI---------SHA
CSN9C        ---EK---LLLLGA---GVPV---------------PATHLPDVEFLTIA-------EAR
CHN23A       ---KK---LLLLGA---GVPV---------------PATQLDDVEFL----------TIA
CHN210A      ---LT---LVLLGA---DVAT---------------PNVRLPAINFVPISE------ALA
G0047A       ---LT---LVLLGD---DVST---------------PNVRLSNLRFVPLGG------AIA
CS.NRPS      ---LT---LVLFGS---EVSA---------------PDVRLPDLQFVSLGD------AIA
CSN9H        ---RK---LVLLGT---DVPA---------------PEPPLVDVELVSIRDILTSVVDHH
CHN28A       ---RK---LVLLGS---DGAA--------------APERPFADVELVPLA-------DVL
CHN22A       ---GS---TVLFLG---EDTP--------------RPEIKVADVEIVSITDALN---DNA
CSN9B        ---HK---LVLLGP---GTQK---------------PPINVTDVEMVPI--------TDA
GLRG_00469   ---HK---LMLLGA---DAPK---------------LPIKVADVEMVS---------ITD
G0044A       ---GR---VIVLLG---SAVT--------------RPALDVPNVSMVL---------ICD
CSN9E        ---QE---VIVLLG---STVT--------------RPAFHLPHVSMVP---------IRD
CHN25A       ---GK---VLLLLG---STVT--------------RPSLDLPHVTMVP---------ISD
G0045A       ---RK---LLFLGA---GVPV---------------PESPLPDVELVR---------VGE
CHN26A       ---RK---LLLLGN---DTPV---------------PDSLAEDVETV----------RLG
CSN9F        ---PGRKLLFLGTS--VDAPV--------------SDSSPLVDVEMVRISEALKEEEARN
CH.NRPS3     ---AL---ALMTNT---DDID------------------FVVEQPLIRV--------DDT
CSN1C        ---A----LALLSN---TNID---------------FTVEQPHIHV-----------DDA
GLRG_06542   ---A----LALLTN---TDID---------------FAVEQPHIHV-----------DDA
CHN32A       ---VK---LCVADV---EVYT--------------NLEDVSSCKFV-----------DTR
CSN1B        ---VK---LCVADA---KVAT--------------TLEGLSDCKFM-----------DIQ
G0652A       ---VQ---LCVADA---KVST--------------TLEGLSNSKFM-----------DIQ
CSN1A        ---AK---VVVVSR---DMVS--------------KIPTDEPDRVVLIV--------DED
G0651A       ---AK---VVIVSR---DMVS--------------KIPADEPDRVVLIV--------DED
CHN31A       ---AK---IVVVSR---DLVS--------------KIPQDEPDRIILIV--------DQD
CHN41A       ---AA---FLVASQ---AQLK------------KLRGYEISQQLGLVQVPSNIGDEHELD
CSN7         ---AR---LVILQR---EQGS---------------FASHIEGIEIL----------PVE
CHN62A       ---AK---RVVCFP---HHQT------------MVENDLLLPEVVF-----------DPE
CSN2A        ---AT---HALADS---EAVA--------------ATLGELGINVMQVPS-------QAA
```

```
CH.NRPS9     ---AN---VIIASP---AHED----KAMALLGDCHRNDRNSSRVVVLDP--------DSD
CHN91A       ---AV---LVLVSRATRYLWD----------ELDDGYAMDGPNITSLLM--------TLP
CH.NRPS5     ---VP---LIITSP---QQKH--------------LAEKIGRPVLVL----------DQD
CH.NRPS6     ---IK---LAVCSP---KHHT--------------LAQNFAPRFVELGSEAGE----DLK
CHN21A       ---AT---FALASR---LHAG--------------TFSGLVSRTIVL----------EDA
CSN9A        ---AT---FALTSR---LHAD--------------TFRGLVPTTVVL----------EDD
G0041A       ---AT---FALASR---LHAD--------------TFRGLVPTTIVL----------EDD
CSN4B        ---AD---VVLASA---SNAH--------------MFEGLDTNVVIV----------DSH
CHN102A      ---AN---VVVASA---ANAA--------------MFLDMDVHVVVV----------DAG
CH.NRPS7     ---AS---VVISSP--GSQYR--------------IPEEILPGNRVIV---------DES
CSN6A        ---AS---IILSSA---SNRH--------------KIPEHITQENRVIV--------GEP
G0681A       ---AS---IVLSSV---SSQD-------------KIPEHILQENRVIV---------NER
CSN2B        ---AE---ILLCSP---RTRP--------------KFEAIKDKVLTI----------ETG
CH.NRPS4     ---PR---VMLSSL---GQRARLERWARDLDLSKSKDDGPPSPISVQVLAITDDWIKRRM
CH.NRPS11    ---SR---LGLTST---AQHA--------------KLASVVETVMIV----------HKT
CSN8B        ---PA---LGLGSA---AGHE--------------KLSHHISPAIVV----------DQS
GLRG_00920   ---PA---LGLGSA---AGYE--------------RLSPYIRPAIIV----------DRS
CSN2D        ---AT---MILAPT---HYCG--------------RCRDLNCTPVLI----------DGD
CHN44A       ---AS---AVLVSS---DCLS-------------TGLETGFPCIVIDEDGDGSGTGIRCV
CSN2C        ---VK---LVFASL---GRIA--------------SLREALNSEVL-----------PIG
CHN43A       ------------------------------------------------------------
CHN111A      SALAG---LIVCSR---DQAP--------------SLRPFSRHLLEL----------DGE
CSN8A        AGDSG---LIICSS---SQFS--------------SLQQLGRHVLQL----------PCR
G0091A       GGNSG---LVICSS---SQAS--------------SLKHLSGHMLQL----------PCT
CSN4A        ---PK---VILCSK---EHEQ--------------RLLKISQNVVAL----------DEA
CH.NRPS10    ---AK---VILCSR---VHEQ--------------RLLGLSDNVVAL----------DEA
CHN72A       ---AR---TVLMSA---FQAK--------------ATDLGTLERIIV----------DRE
CSN6B        ---AK---FVLVSK---FQAK--------------TTDLGVSTRIIV----------DRD
GLRG_06838   ---AK---IVLMSK---FQSK--------------ATDLGIDTRIIV----------DRD
CHN87A       --------VVLVSA---RYSN---------------LDVGAGRRVVVVDGDNND---DDD
CH063_1253   ---AT---IIVTSQ---RQSE--------------LGRRIAPEGVLVVA--------SDE
GLRG_08065   ---AT---VIVTSQ---KQSE--------------LGRRIAPDGALIIA--------SDD
CSN5         ---AN---IIVTSQ---KQSE--------------LGRRIAPDGMLVIA--------SDD
CHN84A       ---FS---VLLASP---ATHE--------------LGRRLSGTVFAVGP--------EMD
CHN85A       ---SR---LVLLCSGPQSAAS---------------DDADAAAAGPFRDFGGRTWHIDSG
CHN81A       ---SG---LVLTSR---ACES--------------LGRCLEKDVLVVCP--------DAS
CH.NRPS8     ---AG---LILTSE---TSYL--------------VSKRLVDESLGVVVVV------EAE
CHN86A       ---AS---FVLTSV---RHSS-------------SSMDFGGRDAVPVGA--------DTE
CHN83A       ---AK---IILVSE---RYHA--------------AEKLELEQDVLVLPVGPSSHSHSRR


GLRG_09842   VYEAFEN--------------ADVSTPQLKG----SDTAVYVGLMCGDYENILLRDIDSA
CH030B_PKS   VYEGLES--------------AGLSIKQLQG----SDTAVYVGVMSADYTEMLARDVEKY
CSP24        RLVSLDE--------------VPWSSTQDIQDVSTLEGLAMMLYTSG-------------
GLRG_09715   MAPTAPS--------------ASVPTPNLSS----RDADAVLLYTSG-------------
CSP6         DISSIPI------------RERQHELRNSSD----PQAPAVLFYTSG-------------
CH063_0617   TLSSSSS------------SSSSSSMGNRSL----PDTPGVLFYTSG-------------
CH063_0392   RQRYLPR----------------DPVPISAR----SQDTMAVHFTSG-------------
CS.P.NRPS    EAATLGH------------ASSTARGRPPPG----PQDPAMMLYTSG-------------
GLRG_10367   EAATLGH------------ASSTARGRLPPG----PQDPAMMLYTSG-------------
GLRG_07434   GLKDTSE--------------EAIPNLAR------PDAPAIILYTSG-------------
CSP1         ALTTKT---------------DAPSSKILAS----PDDLAMVYYTSG-------------
CSP10        IDDQNPT--------------EIMLEVPETK----STDPAAVLYTSG-------------
CSP20        KVPLLSV-----------------KTSTQAK----ADDAAVVLYTSG-------------
CH063_0306   KNSSPDA-----------------KVPTQAR----AGSAAAVLYTSG-------------
GLRG_05027   SKTTING--------------HANGELPPLPASNDLERLAFVSYSSG-------------
CH.NRPS1     SKTTVNG--------------HANGELSPLPASDDLDRLAFVSYSSG-------------
GLRG_08225   LFEVNF---------------PPQAPLPMAA----PDDAACMLLSPR-------------
CHN24A       TASHQGQ------------QNGFVAAPPGPK----ATSLAYVIFTSG-------------
CSN9D        MSHPLKG------------QPRADAAPSRPG----ATSLAYVMFTSG-------------
G0043A       SRHRQRQ------------DGDGEDATPGPQ----ATSLAYVMFTSG-------------
CSN9G        LGGHAGPNGTTDGVTNGVTRNSTARSVTRPG----PDSLAHVIFTSG-------------
```

```
CH.NRPS2      EVAGSNG--------S---AKEPATSSPQPG----ASSLACVIFTSG-------------
CHN29A        DALRLGA--------------AVNPRDLHPRPAATASSLAHVIFTSG-------------
CSN9I         ATVVVSS----------GDARQRGERYKKPT----ATSLAYVMFTSG-------------
G0046A        LSQGETA------------STSSARKYTQVA----ATNLACVIFTSG-------------
CSN9C         DSHRQSD------------ATTVVDRLPKPT----ATSLAYVMFTSG-------------
CHN23A        EARDGDA-------------KTAVKQLPSPT----ATSLAYVMFTSG-------------
CHN210A       AFRPASS--------------EAPARVSLPS----PTNLAYVIFTSG-------------
G0047A        SSRQPST--------------KPPASVSFPS----ATSLAYVIFTSG-------------
CS.NRPS       SSRHSLT------------ASSSSVNPSLPS----ATNLAYVIFTSG-------------
CSN9H         KYADGDD--------------DDDDAAVRPH----AKSLAYVMFTSG-------------
CHN28A        AANIVDR------------DDARVDAPPRPH----AKSLAYVMFTSG-------------
CHN22A        TTTADAR--------------HAGDASAGPT----ATSVAYVMFTSG-------------
CSN9B         LSEAVAD-----------ARHQVDGVSARPQ----ATSLAYVMFTSG-------------
GLRG_00469    ALGEAAE------------IHHQTDGSAGPR----ATSLAYVMFTSG-------------
G0044A        TLRNYPGD-------------KSSQDIAGPT----ATSLAYIIFTSG-------------
CSN9E         ALRSAPD------------RASSQHVLPGPT----ATSLAYIIFTSG-------------
CHN25A        TLRDAPS-------------KSSIQHLPKPT----ATSLAYIIFTSG-------------
G0045A        ALKGKHA--------------TQGAAGSQPS----ARSLAYVVFTSG-------------
CHN26A        EALKTSL--------------AIPPTASKPS----ATSLAYVMYTSG-------------
CSN9F         TATNTND-------GATTNGAITNGTVSQPS----ATSLAYVMFTSG-------------
CH.NRPS3      LLDGASR---------------VFNSPESLS----IQDNSYCLYTSG-------------
CSN1C         LLDSASP---------------IFTSPASLS----VQDNSYCLYTSG-------------
GLRG_06542    LLNSASP---------------IFTSPKSLS----VQDNSYCLYTSG-------------
CHN32A        SLDMSDL--------------ADDDLRLTAE----PSRPAYVIYTSG-------------
CSN1B         SIDMSGF--------------ADANLCLPPD----PSRPAYVIYTSG-------------
G0652A        SIDMSGF--------------ANANLRLPPN----PSRPAYVIYTSG-------------
CSN1A         EVEKSTS--------------STYDSCHDTE----PHSLAYVMYTSG-------------
G0651A        EVEKSAS--------------STHGSFRDPE----PRSLAYVMYTSG-------------
CHN31A        DVEEPTS--------------PHSGPVRDSK----PHDLAYVMYTSG-------------
CHN41A        TILQNESRRTKQK--------QYQQQQTRPS----PDSTAFVVFTSG-------------
CSN7          DISLSEG--------------EAPSVALT------PKHIAYIIYTSG-------------
CHN62A        VVTDVSK---------STRDAASPAGLAEID----MDAPACVVFSSG-------------
CSN2A         VSHGFPD--------------DGGQTWPRTR----PEDAAFVVFTSG-------------
CH.NRPS9      VFIRDPC------------INAAKGLLDRSQ----PDGVAFVLFTSG-------------
CHN91A        LEEEDDS------------YSQEETKLRESR----PENACYCVFTSG-------------
CH.NRPS5      LLGTLGS------------PANIADFTSPAI----PSDLAYLVFTSG-------------
CH.NRPS6      VLSPADE------------DPPEEEGQETPS----PSQAMYATFTSC-------------
CHN21A        LLESLPA------------ATAAAPFTPVAK----PHDLAYIIFTSG-------------
CSN9A         LLQSLSG------------PRAGGFFTPVAK----PHDLAYIIFTSG-------------
G0041A        LLRSLSG--------------SRAGFTPVAK----PQDLAYIIFTSG-------------
CSN4B         LFTRLEE------------LKVAPFSSTDVQ----PHNAATVLFTSG-------------
CHN102A       LFARLEE------------LQVAPFSSTDVQ----PHNAATVLFTSG-------------
CH.NRPS7      FCRRLPA------------RTAATSLPPTKPPWTTPDNALYVVYTSG-------------
CSN6A         LFTTLST------------RATSWAVRPMVT----SHNALYVVYTSG-------------
G0681A        LFPTLST------------RAASWPARLVTS-----NNALYVVYTSG-------------
CSN2B         FAHTLPL------------TDPTQPVCPEVS----PDNAMYVVFTSG-------------
CH.NRPS4      AMTEISG--------------LAICICPEVR----PRHAMYVQFTSG-------------
CH.NRPS11     SVEELPS------------PESRDQPVASAD----PSSVLYVIFTSG-------------
CSN8B         LVDGITS--------------VGNNSDATVE----PGSVLYICFTSG-------------
GLRG_00920    LVEGIPS--------------VSNIYDVAVD----PGSVLYICFTSG-------------
CSN2D         SVPALSK--------------IPVTKQSTVG----KHNLAYILFTSG-------------
CHN44A        DMERCQGNPEEEGETPDHDPNPNPTSSSTPT----PSDLAYILFTSG-------------
CSN2C         DFRPSEE--------------TAAPLDVAVT----PANAAYLLFTSG-------------
CHN43A        -------------------------------------DAAYLLFTSG-------------
CHN111A       TLAHIAS-------------ADCSSSLPSAS----LTDPAYIIFTSG-------------
CSN8A         KLDA-----------------EIRPPSLPVS----VDDPAYVIFTSG-------------
G0091A        KLDA-----------------KLRRPSLTVS----VDHPAYVIFTSG-------------
CSN4A         WIQSVPS--------------SSLPHPSFPR----PWNLAYIMFTSG-------------
CH.NRPS10     WIQSALN--------------TPLPHPLSPQ----PWNLAYIMFTSG-------------
CHN72A        SLERPGS------------SDDPEMLATRST----PDSPAYVIFTSG-------------
CSN6B         SLEGLTS------------SSSPSRLTRHSE----PDCPAYVIFTSG-------------
GLRG_06838    SLEAFTS------------WSSFETLAPHSE----PDRPAYVILTSG-------------
```

```
CHN87A          DDEPLPE--------------HDEPGMAPQA----PSSTAYVLFTSG-------------
CH063_1253      TLNQDTL--------------ELAPELPYIP----PSSPLYIQFTSG-------------
GLRG_08065      TLTQDTL--------------ELAPELPYIP----PSSPLYIQFTSG-------------
CSN5            TLTQDTL--------------ELAPELPYIP----PSSPLYIQFTSG-------------
CHN84A          TSVTVATAAAVVSSPRPTTLNTPAAAAAAAAAGFDPSFPMYVIFTSG-------------
CHN85A          YFDNNKP------AGLRSPPPPPPPRRSSAPPAVDPDSNMYIVFTSG-------------
CHN81A          VWHSDPQ--------------KQKHDLPQTA----PESTMYVVFTSG-------------
CH.NRPS8        LESLSEP--------------NKRDSLPRVD----PSSTMYVVFTSG-------------
CHN86A          SMRARPE----------------SFVQPRVP----VDAAAYILFTSG-------------
CHN83A          LASHGEEDR------------PSSRTTNHAS----PSSAAYMIFTSG-------------


GLRG_09842      PRYQAT-GV-GRSIMANRISYI--------WDLHGPSM--TIDT-A----CSSSLVALHQ
CH030B_PKS      PTYFAT-GT-ARSILSNRLSYA--------FDWRGPSM--TIDT-A----CSSSLIAMHQ
CSP24           ----ST-GK-PKGIPLTNANIR--------TPILGASE--KISL------GQEVVLQQSG
GLRG_09715      ----ST-GV-PKGVRVTHLGLL--------NQIYVCKD--IIG-------ERQVVLQQTS
CSP6            ----TT-GV-PKGAILKHRSYA--------NLIEASSR--RLQVRP----QKEVVLQQTG
CH063_0617      ----TT-GV-PKGALLTQGNYK--------HIIAAASR--RLGIRPA---TQEVVLQQTS
CH063_0392      ----TT-GT-PKGVVLCHDTVV--------NVVEACAD--IYRT------ASNVVLQQTA
CS.P.NRPS       ----ST-GT-PKGIRLKHEGFTNWAEFTVPRLLTDTCT--GSGG-G----GDLVVLQQSS
GLRG_10367      ----ST-GT-PKGIRLKHEGFTNWAEFTVPRLLTDTCT--GSGG-G----GDLVVLQQSS
GLRG_07434      ----ST-GT-PKGVILRHSSLK--------HEFDHCAA--TYGM-N----EHDVVLQQSA
CSP1            ----ST-GN-PKGIAVKHEGVR--------DVFESACW--LYGV-N----EKVVTLLQSS
CSP10           ----ST-GV-PKGIVLSHQNLI--------HELEFSSK--TYDF------EFENVLCQSA
CSP20           ----TT-GT-PKGILLSHRSVR--------NNIEIATH-HLWGFKE----GSDIMLQQAA
CH063_0306      ----ST-GT-PKGISLSHASLR--------NNIELATR--QFGFKE----GSDVMLQQAA
GLRG_05027      ----TT-GQ-PKGIANPHRAAV--------LSYDLRFG--LSDL-Q----PGDRVA--CN
CH.NRPS1        ----TT-GQ-PKGIANPHRAAV--------LSYDLRFG--LSDL-Q----PGDRVA--CN
GLRG_08225      ----SSRGKDAKGIIFHHAALS--------SAFLTQGR--ALNL-N----RDSRVMQLSS
CHN24A          ----ST-GR-PKGVMVEHRAIV--------RLVKNTNK--HLQI-G----AGSSMPHVSN
CSN9D           ----ST-GR-PKGVQIEHRSIV--------RLVKNTHP--HLQP-G----VGVSMAHLSN
G0043A          ----ST-GR-PKGVMIEHRSIV--------RLVKNTSP--HLQA-G----VGVGLAHVSN
CSN9G           ----ST-GK-PKGVMIPHRAIV--------RLVMETNT--VPRK-Q----ALAVVAHMSN
CH.NRPS2        ----ST-GK-PKGVMIEHRGIV--------RLVKETNT--HPPT-ETNPSATATVSHVSN
CHN29A          ----ST-GK-PKGVMVEHRSII--------RLVKSTNF--MTQS-Q----TAQPVAHISN
CSN9I           ----ST-GR-PKGVMVEHRGIM--------RLVKSTNL--MTQSQA----ATKPFAHMSN
G0046A          ----ST-GR-PKGVLVEHRGIV--------RLVKSSNV--THPL------KAMTVFHISN
CSN9C           ----ST-GR-PKGVMVEHRGIV--------RLVKNTNI--QSAA-Q----SAVPVAHIAN
CHN23A          ----ST-GR-PKGVMVEHRGIV--------RLVKNTNI--QTQA-Q----SAVPVGHIAN
CHN210A         ----ST-GK-PKGVMVEHRGIV--------RLVKKSNI--VREL-P----DATTISHVSN
G0047A          ----ST-GR-PKGVMVEHRGIV--------RLVKQSNI--VREL-P----EETTIAHVSN
CS.NRPS         ----ST-GR-PKGVMVKHRGIV--------RLVKQSNI--VSEL-P----EATTIAHVSN
CSN9H           ----ST-GK-PKGVMIEHRGIV--------RLVRQSRL--TYKL-P----KTPRVAHLSN
CHN28A          ----ST-GL-PKGVMIEQRGIN--------RLVRQSSK--MYKM-P----QHPRVAHLSN
CHN22A          ----ST-GQ-PKGVMIEHRGIV--------RLVRQSNV--STKL-P----ESPRVAHMSN
CSN9B           ----ST-GQ-PKGVMVEHRGII--------RLVKQSGS-VNSDL-P----PAARVAHLSN
GLRG_00469      ----ST-GQ-PKGVMIEHRGII--------RLVRESNA--HSKL-P----AAARVAHMSN
G0044A          ----ST-GK-PKGVMIEHRGIV--------RLVKNSNV--VAKL-P----PTARLAHLSN
CSN9E           ----ST-GK-PKGVMIEHRGIT--------RAVKNTNV--NAKL-P----PAARVAHLSN
CHN25A          ----ST-GR-PKGVMVEHRGII--------RLVKNSNV--VAKL-P----PAARVAHLSN
G0045A          ----ST-GK-PKGVMIEHRGIV--------RLVRNTNC--VADL-P----PAARVAHMSN
CHN26A          ----ST-GT-PKGVMVEHRGIV--------RLAKNTNA--LARL-P----PSARMAHMSN
CSN9F           ----ST-GK-PKGVMVEHRGIV--------RLVKNSNV--VAKL-P----PAARVAHLSN
CH.NRPS3        ----TT-GT-PKGCEITHENAV--------QAMKAFQRLFAGHW-D----ETSRWMQFAS
CSN1C           ----TT-GT-PKGCEITHENAV--------QAMKAFQRLFDGHW-N----ENSRWMQFAS
GLRG_06542      ----TT-GT-PKGCEITHENAV--------QAMKAFQRLFDGHW-N----EDSRWMQFAS
CHN32A          ----TT-GT-PKGVVVTQQNIV--------SNLDVLSR--IYPHHG----NRSRFLQACS
CSN1B           ----TT-GT-PKGVVVTQGNIV--------SNLDVLSK--LYPH-S----NRSRLLQACS
G0652A          ----TT-GT-PKGVVVTQGNVM--------SNLNVLSK--LYPH-S----NRSRLLQACS
CSN1A           ----ST-GT-PKGVGVSHDAAA--------QSLLAHNR----HV-P----PFKRFLQFAS
G0651A          ----ST-GT-PKGVGVSHDAAA--------QSLLAHDR----HI-P----PFKRFLQFAS
CHN31A          ----ST-GT-PKGVGVSHDAAA--------QSLLAHNR----HI-P----SFRRFLQFAS
```

```
CHN41A        ----ST-GV-PKGIVVSHANVC--------TAVAALGD--AFGV-D----GATRTVQFAA
CSN7          ----ST-GQ-PKGIKVPHQAVT--------AAVKSMCE--VEGRTK----GEQRVLQLSN
CHN62A        ----ST-GL-PKAIQLPHRALS--------TSAYHLST--TGHL-S----QCSRVFHFAS
CSN2A         ----ST-GI-PKGVVLSHSCIY--------TRAKDMT---LLDV-K----SDSRVLQFGS
CH.NRPS9      ----ST-GM-PKGIRLTHSSLC--------LSILNYTR--RLGI-S----HRSRLFQFSA
CHN91A        ----ST-AE-PKGVIVQHRAIS--------TSAVYHGL--EAAL-D----HTSRVLQFSS
CH.NRPS5      ----ST-GV-PKAVMTEHRAYV--------TDALAQQQ--SALL-D----AASRVLHFAS
CH.NRPS6      ----TT-GT-PKSITTSHDAFY--------HMAIATASPCALHI-S----PSTRMLQFAS
CHN21A        ----ST-GE-PKGAMLEHRAFA--------SCALEFSP--GLDI-D----RDTRSLQFSS
CSN9A         ----ST-GE-PKGAMMEHRAFA--------SCALQFSP--GLGI-D----KDTRSLQFSS
G0041A        ----ST-GE-PKGAMLEHRAFA--------SCALQFSP--SLGI-D----QNTRSLQFSS
CSN4B         ----ST-GE-PKGVVQEHKTLC--------SAAYAHSG--AMQM-S----ESSRILQFSA
CHN102A       ----ST-GE-PKGVVQEHKTLC--------SAAHAHAG--AMDM-S----GSSRVLQFSA
CH.NRPS7      ----ST-GK-PKGVVVPHAAFC--------SSAKGFSK--AIRLDS----PSARVLQYSS
CSN6A         ----ST-GK-PKGVVVTHSAFC--------SSSRGFSK--AIHLDS----PKARVLQYSS
G0681A        ----ST-GK-PKGVVVTHTSFF--------SSSRGFSK--AIHLDS----PKARVLQYSS
CSN2B         ----TT-GV-PKGSITSHRAYC--------TGFKEHAW--AIEV-G----PETRSLQFWA
CH.NRPS4      ----TT-GA-PKGVVISHRNFA--------TGFRRHCA--ATEV-W----PHTRALQFSA
CH.NRPS11     ----TT-GV-PKGVVIQHKSFC--------SAVASNRS--WLQI-K----AESRVLQFTN
CSN8B         ----TT-GT-PKGVVIPHESFC--------SAVASNSR--ELQI-T----ADSRVLQFTH
GLRG_00920    ----TT-GT-PKGVVIPHESFC--------SAVASNSP--ELQI-T----AKSRVLQFTH
CSN2D         ----TT-GT-PKGVMIEHGSLS--------SSITALGK--GMGL-D----TETRFLQYSA
CHN44A        ----ST-GT-PKGVMVEHRSLS--------SSLVALGS--SHDM-T----PATRTMQFAS
CSN2C         ----ST-GK-PKGVVMEHRAWA--------SAITPLSK--AIGI-E----PDTRTLQFSS
CHN43A        ----ST-GE-PKGVVVQHRALC--------SGMAAQAA--AIRC-T----SETRMLQCAN
CHN111A       ----ST-GK-PKGVVVQHGAYS--------YAAQAHSP--GIHI-N----EDSRVLQFAS
CSN8A         ----ST-GT-PKGVIVNHGAYS--------YAAQAHSG--GIRI-N----QDSRVLQFAS
G0091A        ----ST-GT-PKGVAVNHRAYT--------YALQAHSG--GIRI-D----QDSRVLQFAS
CSN4A         ----ST-GE-PKGVMIDHSACA--------AAVSCHGN--ATGF-G----RKTRALQYAR
CH.NRPS10     ----ST-GE-PKGVMIDHSACA--------AAVISHGK--ATGF-N----QKTRALQYAR
CHN72A        ----ST-GT-PKGIVVSHAAFA--------SSAAAHGS--VFGM-S----RDNRVLQFSA
CSN6B         ----ST-GT-PKGVVISHSAFA--------SSGASHGS--AFGM-S----CNNRVLQFSA
GLRG_06838    ----ST-GT-PKGVVVSHSAFA--------SSAASHGS--AFGM-S----CTNRVLQFSA
CHN87A        ----ST-GQ-PKGVVVSHRAIS--------SSCYHHGL--RMGH-G----PETRTLQFAA
CH063_1253    ----ST-GK-PKGVMVSHENFT--------SGAVPRGW--AVGY-R----AHSRCFDFAS
GLRG_08065    ----ST-GK-PKGVMISHENFT--------SGAIPRGW--AVGY-R----AHSRCFDFAS
CSN5          ----ST-GK-PKGVMISHENFT--------SGAIPRGW--AVGY-R----SHSRCFDFAS
CHN84A        ----ST-GV-PKGAIIPHRSVA--------SALHHQLP--RLGY-G----PGSRHCDFSA
CHN85A        ----ST-GT-PKGVMVPHRSFA--------AALAHQTR--LFGF-G----PGSRHYDFMS
CHN81A        ----ST-GT-PKGVVISHASFY--------SGLHHQLD--RLGF-K----PGCRHYDFMA
CH.NRPS8      ----ST-GN-PKGVVIPHSSFR--------AALHHQLP--RMGF-T----SATRHYDFLS
CHN86A        ----ST-GH-PKGVLIPHRAAS--------TSCTYYGD--RVGF-D----ETSRILFFSS
CHN83A        ----ST-GE-PKGILVEHRAIS--------TSCYHHGT--RIGY-T----RDSRVLQFAS
                  :: .   ..    .

GLRG_09842
GVQALRLGETQLAVVAGSNLLLGPEWYVTESNLNMLSPTGTSRMWDADANGYARGEGVAA
CH030B_PKS    AVQALRSGESKVAIAAGTNLLLGPEQYIAESKLQMLSPSGRSRMWDADADGYARGEGVAA
CSP24         QGFDAAVFQIFIALANGGTLVM---------------A--DNRIDPAELAALMSRERVTC
GLRG_09715    HGFDIALGQIFDALTRGGKLIV---------------VGQHGRGDPKHLSKLMLTESVTY
CSP6          VGFDVCVFEIFTALANAGTLVM---------------ASKDQRRDPLAIANLIKDEKVTF
CH063_0617    VGFDLCPFEIFMALSNAGTLVV---------------ASKEQRQEPSAVVGLIKSEGVTF
CH063_0392    LNFDMALWQVLVTLCSGGKLIV---------------VPQDKRLDMTTVTRLIRQEEVTL
CS.P.NRPS     IGFDMAYLQAFLALAYGGALCI---------------VPRAQRVDAGAITRIMAAEGVTV
GLRG_10367    IGFDMAYLQAFLALAYGGTLCI---------------VPRAQRVDAGAITRIMAAEGVTV
GLRG_07434    WSFDLSVTQIFLALGVGAQLHT---------------VSHTMRADSHAMAELIQRQGVTA
CSP1          LGFDITLIQMFISWPVGGTVCM---------------ISRDMRGDSVALMEFMTKHNVTH
CSP10         FGFDMSLTQIFTALAFGGCVHM---------------VTLDERGDALSITKRIVDDGITF
CSP20         FSFDMSLAQTLTALANGGTLVV---------------APSQLRGDAAGLTRLLAAEGVTF
CH063_0306    FSFDMSLSQTLTALANGGTLVV---------------VPAQLRGDAAGLTRLVAAENVTF
GLRG_05027    VFF---VWEMLRPLLRGATVFA---------------IPDSASYDPSALVSLLDSRQITD
CH.NRPS1      VFF---VWEMLRPLLRGAAVFA---------------VPDSASYDPSALVNLLDSRQITD
```

```
GLRG_08225   LNVDTALVELLSTLVHGGCVCI---------------PSSVET--TSDIVGAACRMDVNW
CHN24A       LAFDAATWEIYAPLLNGGTLVC---------------LDQMTVLSPPLLLEAFRVNKIAV
CSN9D        LAFDATTWEMYAPLLNGGTVVC---------------LDQMTVLNPSHLLNALRTNNITM
G0043A       LAFDATTWEIYGPLLNGGTVVC---------------LDQMTVLSPSLLLDAFRANDIKM
CSN9G        LAFDAATWEMYAALLNGGTLVC---------------VNHLTVLNPADLGRVLSSNRVSA
CH.NRPS2     LAFDAATWEIYAPLLNGGTLVC---------------IDRMTVLNSSELARVLSSQRVTT
CHN29A       LAFDAATWELFTALLNGGTVVC---------------IDHMTVLDQSALGQALRANGVTA
CSN9I        LAFDAATWEMYAPLLNGGTVVC---------------IDHMTVLNPTALADTLAAHGVRA
G0046A       LAFDAATWEIYTPLLNGGTVVC---------------IDHMSVLNPTALGQAFVANGATA
CSN9C        LAFDAATWEIYTALLNGGSLVC---------------VDHTTVLDVAMLGRVFEKEGVKA
CHN23A       LAFDAATWEIYTALLNGGSLIC---------------IDHTTVLDVAALGRVFEQEKVKA
CHN210A      VAFDASAWEMYTALLNGATLVC---------------IDYFTTLDSRALEAVFAREHIRV
G0047A       IAFDASAWEMYTALLNGATLVC---------------IDYYTTLDSRALGAVFARESIRV
CS.NRPS      IAFDASAWEMYTALLNGATLVC---------------IDYFTTLDSRALEAVFARENVRV
CSN9H        TAFDNSTWEIYIALLNGGTVVC---------------IDYYTTIDSKALESVLVQERVQA
CHN28A       TAFDNSTWEIYMALLNGGTVVC---------------IDYYTTIDSKALESVFTREGVQA
CHN22A       TAFDASTWEIWSALLHGGTLVC---------------IDYFTALDSKALAASFTEHRIQA
CSN9B        VAFDASTWEVWSALVNGGSVVC---------------IDYFSTLDSKTLEAAFAEHKIQA
GLRG_00469   IAFDASTWEIWVALLNGGTAVC---------------IDYFTTLDSEALGAAFAQHGIHV
G0044A       LAFDASTWEIYTALLNGGTLVC---------------VDYFTTLDPAALNTLFAKEQIQV
CSN9E        VAFDASTWEIYSALLNGGTLVC---------------VDYFTTLDPTALNTVFLEERIQA
CHN25A       VAFDASTWEIYSALLNGGTLVC---------------IDYFTTLDPAALNTLFLRERIQA
G0045A       LAFDISVWEIFGSLLNGGTMIC---------------VDYFTLLDPKALAALFIRERVEA
CHN26A       IAFDAATWEIYCALLNGGTLVC---------------VDYFTAIDVKTLQALFASRRVEA
CSN9F        VAFDAATWEIYAALLNGGVVVC---------------VDYFTLLDTKALQTLFEHHRVNA
CH.NRPS3     FHFDVSVLEQYWSWSVGIPLVG---------------APRDLI--LDDLAGTIRRLNITH
CSN1C        FHFDVSVLEQYWSWSVGIPLVA---------------APRDLI--LDDLAGAIRRLRITH
GLRG_06542   FHFDVSVLEQYWSWSVGIPLVA---------------APRDLI--LDDLAGAIRRLNITH
CHN32A       QAFDVSVFEIFFTWKVGACLCS---------------GTNDTL--FEDLERAIRFLGCTH
CSN1B        QAFDVSVFEIFFTWIVGACLCS---------------GTNDTL--FEDLERAIRFLKCTH
G0652A       QAFDVSVFEIFFTWKVGACLCS---------------GTNDTL--FEDLERAIRFLECTH
CSN1A        PTFDVSVFEIFFPLFRASTLVS---------------CDRGRM--LNDLPAVIRTLQVDA
G0651A       PTFDVSVFEIFFPLFRGSTLVA---------------CDRGRM--LNDLPAVIRTLQVDA
CHN31A       PTFDVSVFEIFFPLFRGSTLVT---------------CDRGRM--LNDLPAVIRMLQVDA
CHN41A       HSWDVSVQDYLGSLLRGATVCV---------------PSDNER--WNDLEGYMVRTAVNW
CSN7         YVFDASIIDIFNTLSTGGTLCM---------------APTAEL--LSDMAGYIRSMDVRQ
CHN62A       FAFDLSIGELLFSLSAGATVCV---------------PTEEER--RANPSKAAGDLGVTW
CSN2A        YAFDMSIAEMVIGLLTGACLCI---------------PSESDR--LDRLQDYVQEAHATW
CH.NRPS9     YTFDVGIGDMLASLATGAVLCV---------------PCEQDR--LGDLNGAIDGYRATH
CHN91A       YSFDVAIDEMITTLVHGGCVCV---------------VGEEDRKSPADLVDAMREMEVNV
CH.NRPS5     YNFDATNFDILSTLIAGGTICV---------------PSEFDR--INRLAGAINDLGANF
CH.NRPS6     YTFDVANRDMLIPLMFGGCVCV---------------PSERER--VEDLGGFIERTGVTL
CHN21A       YAFGSSLAETMCTLMQGGCVCI---------------PSDHDR--MNDVAGFIRRARVNW
CSN9A        YAFGSSLAETLCTLMQGGCVCI---------------PSDHDR--MNDIAGFIQRARVNW
G0041A       YAFGSSLAETMCTLMHGGCVCI---------------PSDHDR--MNDIAGFIKRARVNW
CSN4B        HVFDVSIIEIVNSLVLGACICI---------------PSDEDR--MNNLASFVSKSRADW
CHN102A      HVFDVSVIEVVNSFILGACICI---------------PSDEER--MNDLAGFITRSRADW
CH.NRPS7     FSFDISMLEILSTLMVGACLCI---------------PSEEER--MNNLARCISSMGVNW
CSN6A        FSFDISMLEILSTLMAGACICI---------------PSEEER--MNNLAGCITGMGVNW
G0681A       FSFDISMLEILSTLMAGACICI---------------PSEEER--MNNLAGCITGMDVNW
CSN2B        YSFDASVGDIMTTLLVGGCICI---------------PSEGDL--NAEIFNFVARSRATW
CH.NRPS4     YSFDSSVGEILTTLAVGGCLCS---------------PSDEDR--SMAIVAFIARSGANW
CH.NRPS11    YCFDASLEEIFTVLVAGGCICI---------------PSETDR--LSDIPGFVARHQVNW
CSN8B        FCFDASLEETFTVLAVGGCICI---------------PSENER--ISDVAGFAARKSVNW
GLRG_00920   FCFDASLEETFTVLVAGGCICI---------------PSEEER--MSDLAGFVTRKSVNW
CSN2D        FTFDGMLLDTFAALVHGGCVCV---------------PSEHQK--MNDLAGFIQNFQINT
CHN44A       YAFDAMLVEIWSTLAFGGCVCV---------------PSDQQR--MNDITGFITTARVNH
CSN2C        YTFDVMLLDIFATLISGGCVCT---------------PSESER--MNDVGGYIARENVTA
CHN43A       FAFDASILEIFAPLSVGGCVCL---------------PDDAER--MADLAGFIRRHDVNA
CHN111A      YGFDTSMEDHLTTFAVGACLCV---------------PSEEDRLNLQNLASFTSKSGANW
CSN8A        YGFDTSMEDHLTTFAVGACLCV---------------PTDEGRHTLQGLASFVHSSGANW
G0091A       YGFDTSMEDHLTTLAVGACLCV---------------PSDEDRHTLPGLASFARSSGANW
```

```
CSN4A         HTFDASIAEILTTLAFGGCVVV---------------PSEYER--LNNIAAVIQEKKVNW
CH.NRPS10     HTFDASIAEILTTLAFGGCVVV---------------PSADER--LNNITAVIQEKRVNW
CHN72A        YSFDASLFEILTTLIHGGTVCV---------------AMEAER--LESIGGFMRRMNVDL
CSN6B         YSFDASLFEIFTTLIHGGTVCV---------------PTEKER--IEAIEEFIERHHVNF
GLRG_06838    YSFDASLFEILTTLIHGGTIFV---------------PTEKER--LEAIGEFIEGHRVNF
CHN87A        YTFDASCYDIFTTLVWGGVVAT---------------PSEDQR--LGRLSEAMDEMKVNT
CH063_1253    YAFDVAIDCMLCTLAAGGCLCI---------------PSDEDR--MNDLSGAIRNSKCNM
GLRG_08065    YAFDVAIDCMLCTLAAGGCLCI---------------PSDEDR--MNDLSGAIRNSKCNM
CSN5          YAFDVAIDCMLCTLAAGGCLCI---------------PSDEDR--MNNLSGAIRNSKCNM
CHN84A        YSFDISIHNAIGMLATGGTLLI---------------PSDRDR--KSNVRGALVSMQATS
CHN85A        YSFDASIWNTVAPLVAGGCLCI---------------PSEHER--RNNLAGSIAAMKATS
CHN81A        YSFDIAIHNTIATLAVGGCLCI---------------PSESER--KNNLRGSMLAMGATS
CH.NRPS8      YAFDASIWNTVATLAAGGCLCI---------------PSEHDR--KNNLTRSVESLRATS
CHN86A        FVFDALILEVITPFVRGSCICI---------------PSEDAR--LDDLTGCINELDVNI
CHN83A        YVFDVSLQEILTTLVFGGCVCV---------------PAEAER--LSNVAHCISSMSVNM
                   .

GLRG_09842    VVL-----KRLSDAVT-DGD---------FIE--CVVRETGVCQDGRTNGITTPSSESQA
CH030B_PKS    IVMKPL--SQALADGD-------------SIE--CVIRETGLNQDGKTPGITMPSSSAQA
CSP24         SVFIVSEMQVMLRFGY-EEL-RRC--P--SWR--FAVVA----------GETFT-VNLLD
GLRG_09715    TCIVPSEFHMMHQHGL-EYL-KQC--R--QWR--TAHVA----------GEKVT-HQLRR
CSP6          MAGTPVEFKHIFQHAS-GILKTC---T--SFR--TVAVG----------GEIFT-AQLAA
CH063_0617    MSGTPAEFKHIFRHDA-LGTLRTC--G--GFR--TVVVG----------GEVFT-PQLAG
CH063_0392    TIATPSEYSAWLSHGA-EDL-RRN--R--QWQ--IAVIG----------GEQYS-AQVDN
CS.P.NRPS     TCAVPSEYKNWLTYGD-RAL-LAAGLS--SWR--TAVSG----------GEPGS-DAVVE
GLRG_10367    TCAVPSEYKNWLTYGD-RAL-LAAGVS--SWR--TAVSG----------GEPGS-DAVVE
GLRG_07434    TYATPTEYKSWLRRES-QDG-LRS--S--AWR--LALVA----------GEPVT-EPLLQ
CSP1          TVGTPSEYNSWLRFGD-QAK-LRS--S--AWR--VAICG----------AEAFP-PTLLE
CSP10         TGATPSEYLSWLVYGV-PTLGRSE--S--SWR--RAISG----------GEPIT-PALLN
CSP20         TVATPTEYASWIGAGS-QHL--KS--S--KWR--VAASG----------GEKMT-QSLLL
CH063_0306    TQATPTEYASWIGAGA-QHL--KS--S--KWR--VAASG----------GEKMT-QSLLS
GLRG_05027    TLMTPTLLATVLSRHP-DLSKRL---P--KLR--SLWLN----------GEVVT-TDLCR
CH.NRPS1      TLMTPTLLATVLSRHP-DLSKRL---P--KLR--SLWLN----------GEVVX-TDLCR
GLRG_08225    TYMTPVL-ARKMT--P-AAI------P--SLK--TVCFR----------TRRLD-EDTCA
CHN24A        VFFTTALMRQLLEEMP-SII------G--SLD--VLFTG----------GETMRPQDAFK
CSN9D         AFFTTALLRQYLEEMP-SIV------A--SLE--VLLAG----------GEAIRPQDAFK
G0043A        AFFTTALMRQFLEEMP-SII------G--SLD--VLLVG----------GETMRSKDAFK
CSN9G         ALLTTALLKQILQEAP-AVI------S--GLD--LLFSG----------GEWMRPQDAQK
CH.NRPS2      AFFTTALLKQVLEEAP-SII------S--DLD--YVFAG----------SEWMRPQDAQK
CHN29A        ALFTTALLNQFLNEMP-SVI------A--GLD--LLISG----------GEAMRPKDASR
CSN9I         AFFTAALLRQFLEDAP-FLI------S--RLD--LLFVG----------GEAMRPRDAFK
G0046A        AFLTTALLRQVLEEAP-SVI------S--GLD--LLLSG----------GETMRPKDAFR
CSN9C         AMFTPALLKQCLAESP-STI------E--GLE--LLFSS----------GDRLDVRDALS
CHN23A        VMFTPALLKQCLAESP-STI------K--GLE--LLFAS----------GDRLDVQDALT
CHN210A       AMLSPALLKQCIANIP-DTL------A--ALE--ILYVG----------GDRLNPSDARD
G0047A        AMLSPALLKQCIINTP-STL------S--DLE--ILYVG----------GDRLNPLDATD
CS.NRPS       AMLSPALLKQCIANIP-ATL------A--ALK--VLYVG----------GDRLNPSDAAD
CSN9H         AMLLPSLLKQCLANAP-AAV------A--ALD--IILAA----------GDRFDGRDAVA
CHN28A        TMMLPSLLKQCLAHAP-GAV------A--ALD--IILAA----------GDRFDGRDAVE
CHN22A        ALLPPALLKQCLSNAP-DML------R--GLA--LLFVG----------GDRFDGRDAIE
CSN9B         AMLPPALLKQCLANAP-DML------R--GLR--LLFVA----------GDRFDSRDAIE
GLRG_00469    LLLPPALLKQCLANTP-DTL------R--GLD--MLFVG----------GERFDGRDAIE
G0044A        TILAPALLKQCIKPMS-AAI------G--ALD--VLLVG----------GDRFDGHDAAA
CSN9E         AMLPPALLKQCLKPMA-SAL------S--ALD--LLFAA----------GDRFDGHDAAA
CHN25A        TMLPPALLKQCLKPME-ATL------G--ALS--LLFAA----------GDRFDGPDAAA
G0045A        MILPSPLLKQYLATVP-EML------A--SLK--LVLSG----------GDRFDAHDAIR
CHN26A        AVFPPALLKQCLALAP-ATL------A--SMQ--FLLTA----------GDRCDPRDAAR
CSN9F         AMFPPALLKQCLAHAP-AML------G--PME--LLFAA----------GDRFDARDAVK
CH.NRPS3      LDLTPSL-ARLLD--P-SEV------P--SLSRGVFITG----------GEALK-QEILD
CSN1C         LDLTPSL-ARLLD--P-SEV------P--SLSRGVFITG----------GEALK-QEILD
GLRG_06542    LDLTPSL-ARLLD--P-SEV------P--SLSKGVFITG----------GEALK-QEILD
CHN32A        LSMTPTV-ASLVD--P-RNV------P--TVE--FLVTA----------GEPMT-AAVAE
```

```
CSN1B           LSMTPTV-ASLVD--P-RNV------P--TVE--FLVTA----------GEPMT-AVVAE
G0652A          LSMTPTV-ASLVD--P-RNV------P--TVE--FLVTA----------GEPMT-AVVAE
CSN1A           CELTPTVAGSLLR--SREKV------P--CLK--LLLTI----------GEMLT-ESVIR
G0651A          CELTPTVAGSLLR--SRKNV------P--CLN--LLLTI----------GEMLT-ESVIR
CHN31A          CELTPTVAGSLLR--SRKNA------P--CLK--LLLTI----------GEMLT-EPVIR
CHN41A          AHLTPTV-ARTLR--P-ERT------S-KSLR--TLLLV----------GEPMG-DADIA
CSN7            AAITPTV-ASILQ--P-DHV------P--SLE--TLCMV----------GEPLS-RCVRD
CHN62A          ALLTPSV-IGLLD--P-EEV------P--TLQ--VLASA----------GEQLT-TQIVE
CSN2A           AILTPTV-ARLLK--P-SML------R--HMR--TLILG----------GELVG-ESDIR
CH.NRPS9        VILTPSV-AAFLR--P-ERLEN----G--GLR--TMVLI----------GEVAT-SELYK
CHN91A          ALLTPAV-LQTLD--P-YSV------P-ACLK--TLMVG----------GSPLN-SSLLT
CH.NRPS5        LGVTATL-AQTLD--P-DDV------P--LLK--VVILC----------GEANS-TELVH
CH.NRPS6        ASLTPSM-AGTLR--P-ASV------P--SLK--GLVLG----------GEQMT-ESHVQ
CHN21A          STLTPSF-VATIP--P-ESL------S--GLK--TLLSG----------GESFS-AYQRD
CSN9A           STLTPSF-VATVP--P-ESL------S--GLR--TLLSG----------GEAFS-AYQRD
G0041A          STLTPSF-VSTIP--P-ESL------S--GLR--TLLSG----------GESFS-AYQRD
CSN4B           AFFTPSF-ARTLD--P-RDL------A--SLK--TVCLG----------GEAMT-LDNIQ
CHN102A         AFFTPSF-ARTLD--P-RDL------P--TMK--TVCMG----------GEAMT-LDAIQ
CH.NRPS7        AMLTPSV-ASLMS--P-EEA------P--GLE--VLCFV----------GEALP-QAVAD
CSN6A           AMLTPSV-ASLMS--P-EEV------P--SLE--VLCFV----------GEALP-QNVAD
G0681A          AMLTPSV-ASLMS--P-EEV------P--SLE--VLCSV----------GETLP-QDVAE
CSN2B           AGWTPSF-ASLID--P-DTV------P--TLT--VLLMA----------GEPLP-ASQVD
CH.NRPS4        AAWTPSF-ASLID--P-DDV------P--SLD--AMCLA----------GEPLS-ASLVD
CH.NRPS11       AAFTPSY-LRTLD--P-DEL------E--SLE--FITVH----------AEPMS-QDLVA
CSN8B           AAFTPSL-LRILN--P-DEL------MF-SLR--FITVH----------AEPMC-GDLVS
GLRG_00920      AAFTPSL-LRTLN--P-DEL------A-SSVC--FITVH----------AEPMC-ADLVS
CSN2D           TCFTPSL-YRIID--P-DSV------P--SLK--VVMIG----------GEPVL-QSDVD
CHN44A          LFCTPSL-SRTIP--P-DAV------MPWPLQ--KVSLG----------GEAMS-QGDVD
CSN2C           LFSTPSV-IRLID--P-AQV------P--SLK--TILSL----------GEPLA-QSDVE
CHN43A          CTFTPSL-LRLLD--P-ADV------P--GLR--TIVSG----------GEPLT-RTDIE
CHN111A         AHLTPSF-AELLT--P-TLM------P--TMK--TMVLG----------GEAMT-ARNIR
CSN8A           AHLTPSL-AAVLS--P-SLV------P--TIR--YMVLG----------GEPMT-ADNIR
G0091A          AHLTPSF-AAMLS--P-SLL------P--TIR--YMVLG----------GEPMT-ADNIR
CSN4A           AFFTPSV-IRLLE--P-FEV------P--GLT--TIVLG----------GEAIG-QDNIE
CH.NRPS10       AFFTPSL-IRLLE--P-FEV------P--GLA--TIVLG----------GEPIG-QDNIE
CHN72A          ALLTPSV-ARIVD--P-AEV------P--SLR--TLILG----------GEAPD-GVLVK
CSN6B           ALLTPSV-ARIID--P-AEV------P--SMQ--TLVLG----------GEEPD-QLVVK
GLRG_06838      ALLTPSV-ARIIE--P-TEV------P--SIK--TLILG----------GEEPD-RLVVK
CHN87A          AALTASV-SRLIE--P-SRV------P--SFH--TLTLA----------GEAPL-MTDFT
CH063_1253      AHMTPSV-ARVLD--P-DVI------P--SLE--VLGLG----------GEAVS-AGDAS
GLRG_08065      AHMTPSV-ARVLD--P-DVI------P--SLE--VLGLG----------GEAVS-AGDAS
CSN5            AHMTPSV-ARVLD--S-DVI------P--SLE--VLGLG----------GEAVS-AGDAS
CHN84A          TVLTPSVARLVLR--E-GIP------P-SSLN--AIGLV----------GEALT-RDDVQ
CHN85A          VSLTPSV-ARLVS----GAL---------DVD--TLVLL----------GEGVSVEDTRL
CHN81A          TSLTPSV-ARLLG--D-DLP------A--TLE--SLVLI----------GEAVTLADTKR
CH.NRPS8        VSLTPSVLGLLLL--P-DFP------P--SLQ--TVVLL----------GEAVT-AEHAR
CHN86A          IFFTPTV-ARLIE--P-SEV------P--SLR--TILMG----------GEPPS-QTDLG
CHN83A          AFLTPTV-ARVVD--P-DHV------P--SLK--TIMLG----------GEKAS-RQDFS


GLRG_09842      SL------------------IRH---VYA-KAGLDVVKD--GCQYF-------------E
CH030B_PKS      AL------------------IRS---TYA-RAGLDLSKPSDRPQFF-------------E
CSP24           QF-----RALSR----SDLTIIN---AYG-PTEASICSS--IWQVS-Y--------ADLE
GLRG_09715      AF-----RDLAL----PELRLWN---AYG-PTEASVTCG--RGPMP----------YQTE
CSP6            QF-----KQAMP----PGARVFD---AYG-PTEACIFAT--IGEVN-----------YLE
CH063_0617      QF-----KRSMS----SNVKVFN---VYG-PTETCIFCS--VEEVPYS--------RGDD
CH063_0392      GL-----RTLQL----PHLRLMN---FYG-PSEVSFVSH--YMEVF-----------PGM
CS.P.NRPS       LLAVRGPRQSKGQ---QLPRVMH---IYG-PTEITFFAT--FGALG-----------PDD
GLRG_10367      LLAVRGPRQSKG----QLPRVMH---IYG-PTEITFFAT--FGVLG-----------PDD
GLRG_07434      LF-----REVDR----DGLRLFN---VYG-PTETTCGST--KTELS-Y--------RIPG
CSP1            KL-----RELNK----PDLQLFH---MYG-TTETTVYAT--QQELNWK--------QESF
CSP10           AF-----ASVAN----PQLRLFN---AYG-PTETTCSVT--RIEVPYR--------DLVQ
```

```
CSP20          SF-----LSLGK----PDLDLVN---AYG-PTETSFASN--TRKMPYQ--------DLGS
CH063_0306     AF-----RSLGK----PDLGLFN---GYG-PTETSFTSN--TREMP---------YHDTE
GLRG_05027     RA------MKAL----PETKLLN---VYS-ASETHEVAAGDIRTFV-------------D
CH.NRPS1       RA------IKAL----PETRLLN---VYS-ASETHEVAAGDIRTFV-------------D
GLRG_08225     PW---------M----AKTKVLL---AYGAPDTCPLGVS--VLEVT--------------
CHN24A         AR--------GL----VRKALCH---VYG-PTENTVFST--MYTMR-----------SDE
CSN9D          AR--------GL----VRKAFCN---IYG-PTENTTYST--LYEMR-----------PDE
G0043A         AR--------GL----VRKAFCH---VYG-PTENTTFST--LYEMG-----------PDD
CSN9G          AA--------RL----VRHSFFH---AYG-PSENATYST--VYQVR----------AEED
CH.NRPS2       AA--------RL----VRRSFRH---MYG-PTENTTYST--VFQMPTAAAASEKDDEQRE
CHN29A         AC--------QL----VKQALYN---AYG-PTENTTIST--AYRID-----------ADE
CSN9I          AR--------DL----VRQGLLN---GYG-PTENTTFST--VYAVG-----------RDE
G0046A         AR--------EL----VRQRFYH---VYG-PTENTTYST--LYHMG-----------EDE
CSN9C          AQ--------AL----AKGGIIN---ALG-HTENTVYST--AF-----------------
CHN23A         AQ--------AL----AKGGIIN---ALG-HTENTVYST--AFHMR-----------PAE
CHN210A        AQ--------AL----VKTGVYN---AYG-PTENSVAST--IYKVQ-----------PGD
G0047A         AK--------AL----VKTGVYN---AYG-PTENSVAST--IYKVK-----------EGD
CS.NRPS        AQ----ALVSTD----GGGVVYN---AYG-PTENTVAST--IYKIN-----------KGD
CSN9H          AQ--------EL----VRVNVYN---AYG-PTENTVTST--VYAVR-----------GDE
CHN28A         AQ--------GL----ARINVYN---AYG-PTENTVTST--VYAVR-----------GDE
CHN22A         TK-------ALV----RNGGVYN---AYG-PTENTVMAT--VYEVT----------DEDT
CSN9B          TK-------ALL----ARGDVYN---AYG-PTENTILST--VYQVA-----------END
GLRG_00469     AK--------AL----VRGSIYN---AYG-PTENTVIST--FYEVT----------DGDA
G0044A         VQ--------AL----VRGSVYN---AYG-PSENTVIST--IYELS-----------EMD
CSN9E          AR-------ALV----APGSVYN---AYG-PTENTILSS--IYEVS-----------DAD
CHN25A         AS-------RAL----VSGSVYN---AYG-PTENTILST--IYEVS-----------GQD
G0045A         TR------SLIS----ASGSVYN---AYG-PTENSVLST--MHKIS-----------KKA
CHN26A         IK--------SI----VPGTFIN---AYG-PTENTVLST--FHEVT------------DE
CSN9F          VA--------SI----ISGSVYN---AYG-PTENTILST--IHEVS-----------SQD
CH.NRPS3       VW---------G----PVGVIYN---AYG-PTEATIGVT--MYQRV-------------P
CSN1C          VW---------G----PVGVIYN---AYG-PTEATIGVT--MYQRV-------------P
GLRG_06542     VW---------G----PVGVIYN---AYG-PTEATIGVT--MYQRV-------------P
CHN32A         KW---------------TGCLFQ---GYG-PAETTNICT--VKKMR--------------
CSN1B          KWTGYLFQGVSSWHRNDNYRVLTDLAGYG-PAETTNICT--VKKMQ--------------
G0652A         KW---------------TGYLFQ---GYG-PAETTNICT--VKKMK--------------
CSN1A          EF-----GGTKE----SEGILWG---MYG-PTEAAIHCT--LRPQY-------------P
G0651A         EF-----GGTED----SEGILWG---MYG-PTEAAIHCT--VRPEY-------------P
CHN31A         EF-----GGTGE----SGSILWG---MYG-PTEAAIH-----------------------
CHN41A         GW--------VE----AGTRAFN---VYG-CTEATWVQV--SRPKT------------GA
CSN7           KW---------L----PFCRLLN---SYG-PTECSVVVC--TKEVQ-------------S
CHN62A         RW---------A----GRVRLFN---MYA-PAECTVISH--ISRIL------------PG
CSN2A          PW--------ID----AGIRVYN---GYG-PAEVTFVAT--IAQAS--------------
CH.NRPS9       TW---------H----GRVRLFN---AYG-PAECSVLTT--VHEVE-------------S
CHN91A         QW---------E----SRVRLLV---AYG-PSECSITAT--MSSDV--------------
CH.NRPS5       KW-----TRPGA----GPRDVIN---GYG-PSEASCAFS--YNVYT-------------R
CH.NRPS6       SW-----TGDVG----NKVGLFN---AYG-VSESTGIAA--LARDV-----------QPG
CHN21A         AF---------A----GRIRVIN---AYG-QSESSTMAG--AARVY-------------A
CSN9A          AF---------A----SRLRVIN---AYG-QSESSTMCG--AARVY-------------E
G0041A         SF---------S----SRIQVIN---SYG-QSESSTICG--GSRVY-------------P
CSN4B          RW---------S----DQVQLIN---GYG-PCEGS-VCT--TANFS-------------N
CHN102A        TW---------S----SRVQLIN---SYG-PCEGS-VCT--TANLS-------------S
CH.NRPS7       TW---------A----GHVQAIN---AYG-PAECSAITI--VSEPR-------------T
CSN6A          TW---------A----DRVKTIN---AYG-PAECSAVTI--VSQPR-------------T
G0681A         TW---------A----GHVKTIN---AYG-PAECSAVTV--VSKPR-------------T
CSN2B          AW---------V----DRLKLLN---IYG-PSECSVACV--VNKNV-------------T
CH.NRPS4       RW---------A----DRVRMVN---IYG-PSECSMAST--VNNPV-------------T
CH.NRPS11      RW---------A----GKIRMRP---SYG-PTECSVTST--VGAPF-------------T
CSN8B          RW---------A----GKVHMRS---SYG-PTECSVTST--VGARM-------------T
GLRG_00920     RW---------A----SKVHMRP---SYG-PTECSVTST--VGARM-------------T
CSN2D          RW---------A----PKVRLIA---GYG-PTEACIATL--AGELT--------------
CHN44A         RW--------CS----RGAWLSN---GYG-PTEACIAST--VQVMT--------------
CSN2C          PW-------LGQ----PGICCIN---GYG-PTETCILST--CRKLT--------------
```

```
CHN43A        AW-------LGRLPHGEERHMYN---VYG-VTEACVVSA--AMPMA--------------
CHN111A       NW-------AAS----PDTELIQ---VYG-PSECCVTST--ISPAL-------------S
CSN8A         TW------SCSS----PGPQLIQ---VYG-PSECCVTST--ITRPV-------------L
G0091A        TW------SCPS----PGPQLIQ---VYG-PSECCVTST--ITQPV-------------L
CSN4A         KW---------A----PGRKLVV---GYG-PTENTVFST--MHPLL--------------
CH.NRPS10     KW---------A----PGRNLVA---AYG-PTENTVFST--MHPLP--------------
CHN72A        KW--------RD----AGTRLFN---AYG-PSECSVIAA--CHCYS--------------
CSN6B         KW--------LD----AGIRLFN---AYG-PSECSIIAA--CYSCS--------------
GLRG_06838    KW--------LD----AGTRLFN---AYG-PSECSVIAA--CYSCS--------------
CHN87A        RW---------S----HLPRVMN---CFG-PTETAIVST--IYDLD-------------F
CH063_1253    TW---------S----KTTSVII---AYG-PSECTVGCT--INSNV--------------
GLRG_08065    TW---------S----KTTSVII---AYG-PSECTVGCT--INSNV--------------
CSN5          TW---------S----KTTSVII---AYG-PSECTVGCT--INSNV--------------
CHN84A        GW--------LD----AGVRVTN---LYG-PSECTSYST--VARSA-------VVVDDDA
CHN85A        LW---------G----RARVIIN---AYG-PAECTPNSV--ISYAS-------------E
CHN81A        IW--------------NKTRLIN---SYG-PAECTPYST--TNDQA-------------A
CH.NRPS8      LF--------RG----KKVAIVN---AYG-PAECTPNSV--IGTSS------------SS
CHN86A        RW---GWGRGPG----PDRTVFN---GYG-PTECVVFCA--LQKID-------------P
CHN83A        RW--------SG----RVGVVLN---GYG-PAECAVCCC--INDVD-----------GWE
                          .

GLRG_09842    AHGT-GTPAG----DPVEA-EAIHSVFTDSSRDGVRT-------------DPLYVGSVKT
CH030B_PKS    AHGT-GTPAG----DPIEA-AAIREAFFGADS------------------HFVPRGPDDT
CSP24         TGAF-SIPIG----KPIAN-YGTY--VVDEES------------------NPVPVGWPGE
GLRG_09715    EDIM-SDSDG---IRALPN-YNLT--IVDEKL------------------NPVPAGFPGE
CSP6          DGLG-LVHAG----FALDN-VSIY--IVDDRM------------------NLVPDGCAGE
CH063_0617    DDLEEPIPVG----RTLDN-VSVY--VVDEAM------------------NLLPDGCAGE
CH063_0392    VSGDGAVPVG----SPLHN-YAAY--VLDSSQ------------------RPVPIGMTGE
CS.P.NRPS     AKQL-PAAAG----RPLHN-YSVY--VLDESL------------------RPVPPGVVGE
GLRG_10367    AKQL-PPAAG----RPLQN-YSVY--VLDESL------------------RPVPPGVVGE
GLRG_07434    FYQG-TIPVG----VASAN-EWFY--ILDASQ------------------NLQPVGQTGE
CSP1          YDDG-VMPAG----FSLPN-KSMY--IVDEDM------------------HLLPVGLPGE
CSP10         ASSKLPLTAG----HVAPN-CSVC--IVDSDL------------------RPLPIGMPGE
CSP20         SSSS-SSSIADLSLMTWPN-YSIS--IVDRDL------------------QPVPAGVSGE
CH063_0306    SNSG-GRNIAALPLTTWPN-YSIS--IVDRDL------------------RPVPAGVSGE
GLRG_05027    FETH-VCPVG----PPMDP-EHIY--IMDESG------------------NRVGTNVSGE
CH.NRPS1      FETR-VCPVG----PPMDP-EHIY--IMDENG------------------NRVGNNVSGE
GLRG_08225    KPSD-LSRVA----PPFLG--RFW--IVNPED-HG---------------KLMPIGALGE
CHN24A        GCVN-GLPIG----RAIHN-SGAF--VMDSRQ------------------QIVPLGVIGE
CSN9D         RCVN-GLPIG----RAIKH-SGAF--VMDSRQ------------------QIVPLGVIGE
G0043A        RCVN-GLPIG----RSIKN-SGAF--VMDSRQ------------------QLVPLGVIGE
CSN9G         FSNG-VVPIG----RAVSN-SGAY--VMDRRQ------------------RLVPIGVMGE
CH.NRPS2      HCAN-GVPIG----RAVSN-SGAY--ITDTKQ------------------RLVPFGVMGE
CHN29A        SFVN-GVPIG----GAISN-SGSY--IMDAHQ------------------QLTPIGTMGE
CSN9I         RCVN-GVPLG----RSVTN-SAAY--IMDARQ------------------QLTPVGTIGE
G0046A        RCIN-GVPIG----RAISN-SGAY--VMDALQ------------------QLTPIGTIGE
CSN9C         -----HMRPG--------------------------------------------------
CHN23A        LCIN-GVPVG----RAISN-SGAL--VLDPQQ------------------RPVPLGVMGE
CHN210A       AFEH-GVPIG----QAVSS-SGAY--IMDSRQ------------------QLVSIGVMGE
G0047A        LFEH-GVPIG----RAVSN-SGVY--IMDSRQ------------------QLVSIGVMGE
CS.NRPS       VFEH-GVPIG----QAVSN-SGAY--IMDSRQ------------------QLVSIGVMGE
CSN9H         DFVN-GVPIG----RAIDN-TGAY--IVDARQ------------------QVVGAGVMGE
CHN28A        GFAN-GVPIG----RAINN-TGAY--IVDARQ------------------QLVSMGVMGE
CHN22A        CFAS-GVPVG----RAVSN-SGAY--IMDPIQ------------------QLVPVGVVGE
CSN9B         ASLANGVPIG----RVVSD-SGAF--IMDSDQ------------------QLVPLGVMGE
GLRG_00469    SLAN-GTPIG----RAISN-SGAF--IMDPDQ------------------QFVPVGVMGE
G0044A        SFVN-GVPIG----KSVSH-SGAY--ILDTNQ------------------QPVPAGVMGE
CSN9E         PFVN-GVPIG----KAVSN-SGAY--ILDANQ------------------QLVPVGVMGE
CHN25A        PFVN-GVPIG----RAVSN-SGAY--ILDPSQ------------------QPVPVGVMGE
G0045A        QYVN-GVPIG----HALSN-SGAY--IMDVRQ------------------NPVPIGVMGE
CHN26A        PYVN-GVPIG----RTATN-SGAY--IMDARQ------------------QPVPIGVVGE
CSN9F         PYVN-GVPIG----GAVSN-SGAY--VMDPRQ------------------NPVPIGVMGE
```

```
CH.NRPS3        QNGR-ASNIG----RQFDN-VGSF--VLQPGT--D---------------IPVMKGAVGE
CSN1C           QNGR-ASNIG----KQFDN-VGSF--VLQPGT--D---------------IPVMKGAVGE
GLRG_06542      QNGR-ASNIG----KQFDN-VGSF--VLLPGTD-----------------IPVMKGAVGE
CHN32A          PGDF-IDHLG----FSFEN-TSTF--VLGSST-TE----------------VVPIGCVGE
CSN1B           PGDF-IDHLG----FSFEN-TSTF--VLGSAT-TD----------------VVPIGCVGE
G0652A          PGDF-IDHLG----FSFEN-TSTF--VLDSAT-TD----------------IVPIGCVGE
CSN1A           RDCA-SSNIG----FPLDS-VSCF--VASIPQEGD-----------AYEFRILPLGEPGE
G0651A          RDSS-PSIIG----FPLDS-VSCF--VASIPQEGN-----------AYEFRILPRGEPGE
CHN31A          ------------------------------------------------------------
CHN41A          QVGR-GHVSG----YGI-N-TRVW--IVRRGD-QA----------------LSPVGCRGE
CSN7            SHGP-AGCIG----RPYPT-AVAF--ILDPDGN-----------------SLCPYGAVGE
CHN62A          ADTP-PACIG----RSP-G-AVSW--VASQND-HR---------------KLVPIGAIGE
CSN2A           -TDQ-ASSIG----RGL-N-TTTW--IVDPN--RG---------------QLAPIGAVGE
CH.NRPS9        PDDN-PAVIG----AAVAPHCRVW--LVDAAD-PS---------------RLVPVGSSGE
CHN91A          GNSP-PGTIG----RPV-G-CRAW--VLGPVQTNDRDEATSKQDPNPRMLRLVPLGCVGE
CH.NRPS5        QSPR-ANNVG----RALEGACWGW--VVNPDN-HS---------------QLLPVGAKGE
CH.NRPS6        GGSS-PSNIG----RGN-G-STLW--VLSMTD-PS---------------RLAPVGAMGE
CHN21A          DTTD-LQNIG----RALG--ARYW--VTDPDDPG----------------RLAPIGAVGE
CSN9A           DTTD-LQNIG----RAL-G-SRYW--ITNPDDPG----------------RLAPIGCVGE
G0041A          DTAD-LQNIG----RAL-G-ARYW--ITNPDDPG----------------RLAPIGCVGE
CSN4B           NPSK-PDSIG----RGA-N-SLIW--IVNSNN-HD---------------ILVPIGSIGE
CHN102A         GHFG-TDAIG----RGA-N-SLVW--IVEPDN-HD---------------RLVPIGSTGE
CH.NRPS7        KGVK-SISLG----RPA-N-CAVW--IVGEDG------------------RLAPFNTVGE
CSN6A           KGVK-SISIG----QPAN--CAVW--IVTENG------------------KLAPFNSIGE
G0681A          KGVK-SISIG----RSA-N-CAVW--IVTENG------------------KVAPFNSIGE
CSN2B           RNTN-ASNIG----RGY-R-CVTW--VVDEND-HE---------------RLRPIGSVGE
CH.NRPS4        RDSL-PSNIG----RAHRS-----------------------------------------
CH.NRPS11       VDTD-ATNIG----WPV-G-CRGW--VVHPENHD----------------ILMPIGAVGE
CSN8B           ADSD-PANIG----YPV-G-CHGW--IVHPENHK----------------ILAPVGAVGE
GLRG_00920      AHSD-PANIG----HPV-G-CHAW--IVHPEN-HR---------------VLAPVGAVGE
CSN2D           PSTS-PNTIG----RPVV--CRAW--VVNPLK-PS---------------ELTPIGGIGE
CHN44A          QATP-IGTVG----FPLDS-CRLW--VVNPLKQQQQQQDDQEYHPGQY--ELAPIGAIGE
CSN2C           PVDS-SSNIG----TPIA--AATW--LVSPFTR-----------------TLAPIGAVGE
CHN43A          LSTS-PRTVG----YPV-G-ASLW--LVSPVSG-----------------MLAPPGAVGE
CHN111A         LYSD-PTNIG----SAVPG-CRTW--IVRPDD-PN---------------ALQAIGVVGE
CSN8A           QDGD-PTDIG----TAVPN-CKTW--IVKPGN-PD---------------NLQVIGTVGE
G0091A          PDGD-PTDIG----TALPN-CKTW--IVKPSD-PD---------------SLQVVGAVGE
CSN4A           LSAK-AGVIG----RGV-G-TRCW--IADPEN-PS---------------RLSPIGAPGE
CH.NRPS10       PSAK-AGVIG----RGV-G-TLCW--IADPED-PN---------------RLSPVGAPGE
CHN72A          AHTD-PRTIG----LPV-G-CSSW--IVNADDNEA---------------SLVADGDIGE
CSN6B           HEMD-PRTIG----LPV-G-CSSL--IVDPDD-ES---------------SLVPEGEVGE
GLRG_06838      HETD-PRTIG----ISV-G-CSSL--IVDPDN-EL---------------SLAPEGEVGE
CHN87A          SRPD-ASCIG--------------------------------------------------
CH063_1253      -SST-STNIG----KGCGG--LTW--IVDPDDHE----------------RLMPVGAVGE
GLRG_08065      -SSS-STNIG----KGCGG--LTW--IVDPDD-HE---------------RLMPVGAVGE
CSN5            -SST-STNIG----KGCGG--LTW--IVDPDD-HE---------------RLMPVGAVGE
CHN84A          PSAT-ALDTG-------AG-CATW--IVDPNDHN----------------VLMPLGATGE
CHN85A          TPEA-AARIG----KGA-G-AVTW--VANPDD-HE---------------RLVPIGAVGE
CHN81A          GPET-AVRIG----TGA-G-ALTW--VVHADD-HN---------------KLVPLGLVGE
CH.NRPS8        SPEA-ASLIG----TGS-G-AVTW--VADIRD-PS---------------RLAPIGAVGE
CHN86A          TSIR-ETCIG----KST-G-SASW--IVAPDDSD----------------SLTPVGSVGE
CHN83A          DLRR-PHCIG----TAV-G-SATW--VVAPGD-AS---------------QLVPVGAIGE


GLRG_09842      VIGHTEGTAGLA-GLL------KTTAALQ--RKTIF-----P-NRHFKRL----------
CH030B_PKS      LY--VGSI-KTVIGHTEGTAGLA-AVVKA--TEALKAATLPP-NRLFNRL----------
CSP24           IA--ISGP-GVASGYA------N-LPLL--------------------------------
GLRG_09715      IC--ISSPYSISRGYN------N-RPEET--ERQFI-----D-SKLIRGG----------
CSP6            VC--IGGA-GVGHGYL------D-RPEPT--TKAFV--ED-P-CATVEDQ----------
CH063_0617      VC--IGGA-GVSRGYFGLSSGFG-SGTRN--SSAFV-----P-DPFASPE----------
CH063_0392      VY--IAGA-GICKGYL------G-NPSLT--GQKFV-----SDPFVRSDL----------
CS.P.NRPS       IF--VGGA-GVAAGYVSGETETG-TETESATADKFLPDHLAP-AAFLSQG----------
```

```
GLRG_10367    IF--VGGA-GVAAGYVLGETETG-TETESATADKFLPDHLAP-AAFLSQG----------
GLRG_07434    IA--IGGV-GVAMGYL------N-MPDRT--QASFL-----P-DPFATEE----------
CSP1          IV--LGGV-GVAQGYN------N-NELMT--KASFV-----P-NTFAPNG----------
CSP10         VL--IGGA-KVAIGYL------G-NDTLT--KERFFDHDHVP-EEFVAEG----------
CSP20         VC--IGGA-GVGLGYF------N-NEELT--SGAFVLDKSAP-PEFAARG----------
CH063_0306    VC--IGGA-GVGLGYL------N-NDKLT--SEAFVSDKTAP-PEFAARG----------
GLRG_05027    LY--VGGD-LLARGYL------N-LPETT--TKAFQ-----PDPFDKKAG----------
CH.NRPS1      LY--VGGD-LLARGYL------N-LPETT--AKAFH-----PDPFVKKEG----------
GLRG_08225    LI--IESP-TLAHHFV------P-GQSLS--HLPQD-----S-DFSLEDG----------
CHN24A        LV--VTGD-GLARGYT--------DPSLD--LNRFV-----Q-VTVEGKS---------
CSN9D         LV--VTGD-GLARGYT--------DPSLD--AARFV-------RVTIGDG---------
G0043A        LV--VTGD-GLARGYT--------DSSLD--QARFV-----Q-VTIDGQS---------
CSN9G         LV--VTGD-GLALGYT--------DAKLD--AHRFV-----T-VTIAGRD---------
CH.NRPS2      LV--VTGD-GLARGYT--------DPALD--RGRFI-----H-ISVAGRK---------
CHN29A        LV--VTGD-GLARGYT--------DSTLN--ADRFV-----E-VDIAGRS---------
CSN9I         LV--VTGD-GLARGYT--------DPALD--KDRFV-----Q-VEIAGRS---------
G0046A        LV--VTGD-GLARGYT--------DSSLE--KDRFV-----Q-VHVGGQS---------
CSN9C         ------------------------------------------------------------
CHN23A        LV--LTGD-GVARGYT--------DAALN--KGRFF-----DMTTADGET---------
CHN210A       LV--VTGD-GLARGYT--------DPSLD--VNRFL-----H-VTIEGQL---------
G0047A        LV--VTGD-GLARGYT--------DHALD--KNRFM-----Y-VTIEGRL---------
CS.NRPS       LV--VTGD-GLARGYT--------DQALD--KNRFM-----H-VTIEGRL---------
CSN9H         LV--VTGD-GLARGYT--------DPALD--VDRFV-----P-ISIGSNN---------
CHN28A        LV--VTGD-GLARGYT--------DPSLD--TDRFVHIRI-P-GYDEP-----------
CHN22A        LV--VTGD-GLARGYT--------DPALD--RDRFI-----H-ISVAGRE---------
CSN9B         LV--VTGH-GLARGYT--------DPTLN--RDRFV-----E-IDMAGKL---------
GLRG_00469    LV--VTGD-GLARGYT--------DQKLN--QDRFI-----Q-IDIAGEL---------
G0044A        LV--VTGD-GLARGYT--------DASLN--EGRFA-----E-VTIHDKV---------
CSN9E         LV--VTGD-GLARGYT--------DAALN--KDRFV-----Q-VAVHGQV---------
CHN25A        LV--VTGD-GLARGYT--------DAALN--KDRFA-----Q-VAVNGQV---------
G0045A        LV--VSGD-GLARGYT--------DPALD--KGRFV-------EVTLGDG---------
CHN26A        LV--VLGD-GLARGYT--------DPALE--VGRFV-----E-ISVDGQP---------
CSN9F         VV--VVGA-GLARGYT--------DRTLD--EGRFVEIIS-P-EGGGGGG---------
CH.NRPS3      LC--VSGK-LVGKGYL------N-RPELT--RERF------P-LLER-LG----------
CSN1C         LC--VSGK-LVGKGYL------N-RPELT--RERF------P-SLER-LG----------
GLRG_06542    LC--VSGK-LVGKGYL------N-RPELT--QERF--------PFLEGFC----------
CHN32A        LC--FGGD-QVAAGYL------G-MPDVT--NAKFL--DH-P-EF---------------
CSN1B         LC--FGGD-QVAAGYL------G-MPELT--NAKFI--DH-P-RF---------------
G0652A        LC--FGGD-QVAAGYL------G-MPELT--NAKFV--DH-P-HF---------------
CSN1A         LV--VGGH-QLAKGYL------N-RPEQT--QSAFL--DT-P-Y----------------
G0651A        LV--VGGH-QLANCYL------N-RLEQT--QSAFL--DT-P-Y----------------
CHN31A        ------------------------------------------------------------
CHN41A        IW--LEGP-MVTQGYI------NVDPLLE--QRSFP--NR-P-DWAE-RL----------
CSN7          LC--IGGP-QLTDGYV------G-RDDLT--SAAFV--------WHDDLQ----------
CHN62A        LL--VEGP-VVSSGYL------G-DAEKT--DAAFLSPSDPP-SWLT-QI---------R
CSN2A         LV--VESL-TLALGYL------N-DPSRT--AESFI--IN-P-SWAQ-LD----------
CH.NRPS9      LL--IEGQ-HVASGYL------N-DPEKT--ARAFISADEAP-AAIRPLL----------
CHN91A        LA--IEGP-ILAEGYI------S-NPRLT--AERFI--FA-S-HVAG-DF----------
CH.NRPS5      LL--IQGP-TLSRGYL------N-EPEKT--AKVFI--ES-P-AWLP-AK----------
CH.NRPS6      LV--IEGP-SVARGYL------N-DPERT--ERSFQ--TE-P-EWKR-AFLQRLVENKDE
CHN21A        LL--IESA-GIARGYV------V-PPPPE--QSPFL--DAAP-AWYPAEW----------
CSN9A         LV--IESP-GIARGYV------V-PPPPD--QSPFL--DEAP-AWYPAEW----------
G0041A        LM--IESP-GIARGYV------V-PPPPD--QSPFL--DDAP-AWYPAEW----------
CSN4B         IL--IEGP-NVARGYL------C-DQEKT--DKAFI--TN-P-AWIT-HF----------
CHN102A       IL--IEGP-NVARGYL------G-DPEKT--AKAFI--TN-P-AWIA-HF----------
CH.NRPS7      IV--IEGP-PVARGYL------G-DREKT--GAVFL--DD-P-GFLRGVT----------
CSN6A         II--IEGP-PLARGYL------G-DEEQT--DAVFL---GDP-SFLRGVV----------
G0681A        IV--VEGP-PLARGYL------G-DEGQT--NAVFL---GDP-SFLRGIV----------
CSN2B         LL--IEGP-ILARGYL------K-RPEKT--AEVFI--DA-P-SWL--KN----------
CH.NRPS4      ----VTS-----------------------------------------------------
CH.NRPS11     LL--LDGP-IVGKGYL------D-DEAKT--AAAFI---DPP-AWAVEVE----------
CSN8B         LL--LDGP-IVGKGYW------R-DEVKT--EAAFV---EPP-EWWARGV----------
```

```
GLRG_00920   LV--LDGP-IVGNGYW------G-DEAKT--GEAFV--ES-S-DWWGRVV----------
CSN2D        LC--VEGP-CVARGYL------G-NEEAT--QAAFL--ES-P-SWLPQSS----------
CHN44A       LY--IEGV-NVARGYL------H-DDAKT--AAAFL--ER-P-PWLDGRR------IEDS
CSN2C        IC--IEGP-TLARGYL------G-DPERT--AMAFE--RNPP-CFPNDVG----------
CHN43A       LF--LEGP-ALARGYH------E-DRPRT--QTSFV--DD-P-AWLV-------------
CHN111A      LL--MEGP-ILAKEYL------N-SPEQT--NNSFT--QG-L-HWAP-EK----------
CSN8A        LL--VEGP-ILAKGYL------G-DAERT--RMAFV--EG-L-RWAP-DK----------
G0091A       LL--VEGP-ILAKGYL------G-DAERT--SMAFV--EG-L-RWAP-DK----------
CSN4A        LL--LESP-QLARGYL------N-FPENEGPNRAFI--RN-S-FGLT-KH----------
CH.NRPS10    LL--LESP-QLARGYL------N-FAENEGPNQAFI--RN-P-PWLAKCG----------
CHN72A       ML--LGGA-TLADGYL------N-DPART--AAAFV--DS-S-SWPL-ES----------
CSN6B        LL--IGGP-ILADGYL------N-NPERT--HSAFI--NT-P-SGSV-DT----------
GLRG_06838   LL--IGGP-ILADGYL------N-DPERT--HSAFT--NT-P-LCSV-GT----------
CHN87A       ------------------------------------------------------------
CH063_1253   LL--IEGP-VVGLGYL------N-DKAKT--DEVFI---EDP-AWLLAGG----------
GLRG_08065   LL--IEGP-VV-LGYL------N-DKTKT--DEVFI--ED-P-AWLLAGG----------
CSN5         LL--IEGP-VVGLGYL------N-DKAKT--DEVFI--ED-P-AWLLAGG----------
CHN84A       LI--LEGP-IIGTGYL------D-EPEKT--AATYI--RD-P-AFLALGA----------
CHN85A       LL--LEGP-IVGAGYL------D-APDKT--AEAFV--ND-P-SWLRRGA------PGGG
CHN81A       LL--LEGP-ILSQGYL------D-DPEKT--AKAFI--ED-P-TWLLQGS----------
CH.NRPS8     LL--LEGP-IVGNGYL------H-DAEKT--KASFI--ES-P-PWLLGGG----SSSSSS
CHN86A       LL--VEGP-ILARSYL--------DPQNT--AKAFI--EN-P-PWLLQGG---GGGDADG
CHN83A       LL--VEGP-ILARGYF------K-DPQLT--AAAFI--EA-P-AWLT-RGGGGG-GGCSP


GLRG_09842   ------NPRIEPFYKNLKV------P--TVEIPWPKPPE--NQPLRASVNSFGFGGTNAH
CH030B_PKS   ------SPKVEPFYKNLEVVTSA--RP-WPKIPANGMRR-----VSVNSFGFGGANCHAI
CSP24        ------------------------------------------------------------
GLRG_09715   -----ALPPGIPVYRTGDLGRLS--AE-DGTLTVLGRIG-QDSQVKIR------------
CSP6         ------RNDWKTMYRTGDSGYLR--P--DGSLVILGRIN-GDSQVKIR------------
CH063_0617   ----DVSRGWTGMYRTGDSGYLK--P--DGSLVVLGRIG-GSTQVKIR------------
CH063_0392   ----AKSRGWTRMYRTGDKARIN--Q--DGTLSILGRID-GDAQLKIR------------
CS.P.NRPS    ---------WTTMHRTGDNGHWR--A--DGQLVIQGRRS-GDTQHKIR------------
GLRG_10367   ---------WTTMHRTGDNGRWR--A--DGQLVIQGRRS-GDTQHKIR------------
GLRG_07434   ----DSRCGWATMYRTGDVGFLQ--H--DGSLVLKGRIGD--------------------
CSP1         --SPFHRKGWTTMYRTRDRGRLL--P--DGSIVVEGRIG-GDTEIKLR------------
CSP10        ---------WTSVHRTGDVGRLL--L--DGQLILEGRVA-GDSQIKLR------------
CSP20        ---------WTKKHRTGDRGRLS--P--DGGLILEGRIA-GDTQIKLR------------
CH063_0306   ---------WKTKHRTGDRGRLS--P--DGGLVLEGRIA-GDTQIKLR------------
GLRG_05027   ----------ARMYRTGDLARIL--P--SGLLEITGRAG---GMIKTRGYTVQPGAVENA
CH.NRPS1     ----------ARMYRTGDLARVL--P--SGLLEITGRAG---GMIKTRGYTVQPGAVENA
GLRG_08225   -------KPTTRYFKTGHLVRYM--E--DGLLDFVSSGR---DAVESNGRVVPVTEIEQR
CHN24A       ----------LRAYRTGDRARYR--PA-DGEIEFIGRVD---NQIKIRGYRIETAEVETA
CSN9D        -----RRTATVRAYRTGDRVRYR--PS-DGEIEYIGRTD---NQVKIRGYRIETAEVEQA
G0043A       ----------VRAYRTGDRARYR--PA-DGELEFIGRTD---NQVKIRGYRIETAEVEQA
CSN9G        ----------VRAYRTGDLARFR--PV-DGQIEFHGRMD---QQVKVRGHRIEPAEIERA
CH.NRPS2     ----------VRAYRTGDRARFR--P--DGQIEFHGRMD---QQVKVRGHRIEPAEIEQA
CHN29A       ----------VRAYRTGDRARIR--PR-DGQIEFFGRMD---QQVKIRGHRIEPVEIEQA
CSN9I        ----------VKAYRTGDRARYR--PG-DGQMEYCGRMD---QQIKVRGHRIEPAEIEHA
G0046A       ----------IKAYRTGDRARYR--PG-DGQIEFFGRMD---QQTKIRGHRIEPAEIEQT
CSN9C        ------------------------------------------------------------
CHN23A       ----------TRAYRTGDRARYR--PV-DAQMEFFGRMD---QQLKIRGFRVEPAEVEQV
CHN210A      ----------VRAYRTGDRARYR--PS-DGQIEFFGRID---QQTKIRGHRIEPAEVELA
G0047A       ----------VRVYRTGDRARYR--PS-DGQIEFFGRID---QQTKIRGHRIELAEVELA
CS.NRPS      ----------TRVYRTGDRARYR--PS-DGQIEFFGRID---QQVKIRGHRIEPAEVELA
CSN9H        --NNNNDQQLVRAYRTGDRARIR--PG-DGQIEFFGRID---HQIKVRGHRIEPAEIEQA
CHN28A       ----------IRAYRTGDRARFR--PG-DGQIEFFGRID---QQIKVRGHRIEPAEIEQA
CHN22A       ----------VRAYRTGDRVRYR--PT-DGHIEFAGRID---QQVKIRGHRIEPAEIEQA
CSN9B        ----------VRAYRTGDRVRYR--PV-DGQIEFSGRMD---QQVKVRGHRVEPAEIEQA
GLRG_00469   ----------VRAYRTGDRVRYR--PT-DGQIEFSGRID---QQVKIRGHRIEPAEIEQA
G0044A       ----------LRAYRTGDRVRHR--PT-DGQIEFFGRMD---QQVKIRGHRVEPAEVEQA
CSN9E        ----------VRAYRTGDRVRYR--PT-DGQIEFFGRMD---QQVKIRGHRVEPAEVEQA
```

```
CHN25A        ----------VRAYRTGDRVRYR--PT-DGQIEFFGRMD---QQVKIRGHRVEPAEVEQA
G0045A        -------QPPIRAYRTGDRARYR--P--DGEMEFFGRID---DQVKVRGHRIEPAEIEHA
CHN26A        ---------PARAYRTGDRVRYR--P--DGQIEFFGRMD---QQVKIRGNRVEPAEVEQA
CSN9F         --GGGGSQSPVRAYRTGDRARYR--P--DGQMEFFGRMD---QQVKIRGHRVEPAEVEQA
CH.NRPS3      ----------ERVYRTGDLVRLL--H--DGCFDFLGRAD---DQVKLRGQRLEIGEINHS
CSN1C         ----------ERVYRTGDLVRLL--H--DGCFDFLGRAD---DQVKLRGQRLEIGEINHS
GLRG_06542    ----------ERVYRTGDLVRLL--H--DGCFDFLGRAD---DQVKLRGQRLEIGEINHS
CHN32A        ----------GRIYRSGDMGRML--P--DGSLMVLGRMD---DQLKLRGQRIDTGEIGSI
CSN1B         ----------GRIYRSGDMGRML--P--DGSLMVLGRMD---DQLKLRGQRIDTGEISSI
G0652A        ----------GRIYRSGDMGRML--P--DGSLMVLGRMD---DQLKLRGQRIDTGEISSV
CSN1A         ----------GRVYRTGDKAIMN--P--DGTLECLGRIS--DGQVKLRGQRIELGEIEQA
G0651A        ----------GRVYRTGDKAIMK--S--DGTLECLGRIS--DGQVKLRGQRIELGEIEQA
CHN31A        ------------------------------------------------------------
CHN41A        -LGRTHDDGQHRFYRTGDLGRLC--A--DGALEVMGRTD---TQAKLGGQRLELSEVEHY
CSN7          ----------MRIYRTGDLARSL--P--GGDIECLGRKD---NQVKLHGFRVELGEIEAV
CHN62A        GEAQGGKV--GRVYKTGDLVRQR--E--DGTLLFLGRKD---DQIKLHGQRLEVEEVEHS
CSN2A         ---MPETPLSRKFYKTGDLARYL--P--DGSLECLGRLD---TQVKLGGQRVELSDIEHH
CH.NRPS9      --------PRAGLYLTGDLARLV--P--HHGLVFLGRKD---LQIKLRGQRVEMGEIEFH
CHN91A        -CRVTGCRPETRLYRTGDLVRQN--P--DGSLLFLGRID---SQVKVRGVRVELEAIEHC
CH.NRPS5      -----TAPQLRRLYKTGDLVRQL--P--DQSYEVYGRID---TQVKLNGQRIELGEIEHK
CH.NRPS6      DGKSDERFPWTRAFRTGDLVRYSPRL--DGTLELMGRRD---HQVKVNGQRLELAAVEGH
CHN21A        --RQTQRANGYRFYRTGDLVCYR--P--DGTVAYLGRRD---SQVKIRGQRVELGDVESH
CSN9A         --RRHQKANGYRFYRTGDLVCYR--P--DGTVAYLGRRD---SQVKIRGQRVELGDVESH
G0041A        --RRHQKANGYRFYRTGDLVCYR--P--DGTVAYLGRRD---LQVKIRGQRVELGDVESH
CSN4B         --ETSSAA--RKMYKTGDLGMLN--H--DGTISYLGRKD---AQIKLRGQRIEPGEVEHQ
CHN102A       --EAGSVA--RRMYKTGDLGMLN--S--DGTVSYLGRKD---TQIKLRGQRIEPGEVEHH
CH.NRPS7      -------RDSGRCYRTGDLGRMR--V--DGSVDFLGRKD---GQVKINGQRVELGEVEHH
CSN6A         -------VDPMRCYRTGDLGRMN--V--DGSIDFLGRRD---SQVKINGQRVELGEIEHQ
G0681A        -------ADPMRCYRTGDLGRMN--V--DGSIDFLGRRD---SQVKINGQRVELGEIEHQ
CSN2B         --GPHPRT--NRLYKTGDLVRYN--S--DGTINFIGRKD---TQLKINGQRVEIGEIEHS
CH.NRPS4      ---------------------------DGSINYVGRKD---TQLKIHGQRVEVGDIEQH
CH.NRPS11     -SAFPEEGLPRKLYKTGDLVSYA--E--DGSLLIHRRKD--HSQVKIRGQRVELGEIQFH
CSN8B         -SSAVSPRSQRKFYKTGDLVRYA--E--DGSLLIQGRKD--DSQVKIRGQRVELREIQHH
GLRG_00920    -DSPEPCRSRRRVYKTGDLVRYA--E--DGSLLIQGRKD--DSQVKIRGQRVELREIQHH
CSN2D         --------PGTRVYRTGDFVYYN--A--DGTLSFFGRKD---SQVKIRGQRVELSEIEEA
CHN44A        DGTTPDDEKHHRVYRTGDLAYYN--P--DGSLRLVGRKD---AQVKLRGQRVELGEIEEA
CSN2C         ----------RRIYHTGDLARYN--T--DGSLTFIGRTD---EQVKLYGHRIELGEIEEN
CHN43A        ---------------------------DG-------------------------------
CHN111A       -----------RLYKTGDLVKYD--S--SGHLHFVHRRD---GQVKLRGQRIELGEIERQ
CSN8A         -----------RLYRTGDLVRYD--S--AGHLHFVGRRD---TQVKLRGQRIELGEIERQ
G0091A        -----------RLYRTGDLVRYD--S--AGHLHFVGRRD---TQVKLRGQRIELGEIERQ
CSN4A         -----GRQ--SRLYRTGDICRFD--E--NGVIHFEGRKD---TQLKISGQRLEASEVEYH
CH.NRPS10     --------RRSRLYRTGDMCRFD--Q--NGVIHFEGRKD---TQLKISGQRLEASEVEYH
CHN72A        ---RPRDA--GRLYKTGDLVRKD--S--DGNMVYVGRKD---MQVKVNGQRIEIGDIESH
CSN6B         ---QPTVN--RRFYKTGDLVKRI--E--DGNMVYVGRKD---RQVKITGQRIELGDIESN
GLRG_06838    ---RAGVD--RRFYKTGDLVKKT--E--DGNMVYVGRKD---MQVKINGQRIELGDIESN
CHN87A        ------------------------------------------------------------
CH063_1253    ---GPARGRHGRLYKTGDLVRYD--PDGTGSIAFVGRKD---QQVKLRGQRIELAEVEHH
GLRG_08065    ---GPARGRHGRLYKTGDLVRYD--PDGTGSIAFVGRKD---QQVKLRGQRIELAEVEHH
CSN5          ---GPARGRHGRLYKTGDLVRYD--PDGTGSIAFVGRKD---QQVKLRGQRIELAEVEHH
CHN84A        --PGHHRGRRGRLYKTGDLVHYS--AG-DGRLHFVSRMD---TQVKIRGQRVELGEVEHG
CHN85A        GGGGGKAGRSGTLYKTGDLVRYD--R--EGSLEYIGRKD---TQVKVRGQRVELDEVAHV
CHN81A        -AQTKQPGRRGRLYKTGDLVRYN--D--DGSLTFVGRKD---TQSKIRGQRIELAEVEYR
CH.NRPS8      AAAAAAPGRRGRLYKTGDLVRYE---A-DGSLGYLGRKD---TQVKIRGQRVELGEVEFH
CHN86A        GGGRRRPGRRGRLYKTGDLVRYD--Q--DGNLCYVARKD---NQVKIRGQRMELGEIEAH
CHN83A        SSRRSGRT--GRLYKTGDLVCYD--G--SGKLTFVGRKD---TQVKIRGQRMELEEVELH


GLRG_09842    VILES-----------
CH030B_PKS    LES-------------
CSP24         ----------------
GLRG_09715    ----------------
```

```
CSP6            ----------------
CH063_0617      ----------------
CH063_0392      ----------------
CS.P.NRPS       ----------------
GLRG_10367      ----------------
GLRG_07434      ----------------
CSP1            ----------------
CSP10           ----------------
CSP20           ----------------
CH063_0306      ----------------
GLRG_05027      IVKHLAVRDCAV----
CH.NRPS1        IVKHLAVRDCAV----
GLRG_08225      LRRC------------
CHN24A          IAGLSGVRDSAV----
CSN9D           LAGLAGVRDSAV----
G0043A          LAGLTEVRDSAI----
CSN9G           MLGYAGIGDAVI----
CH.NRPS2        MLSYAGVGDSIV----
CHN29A          ML--------------
CSN9I           MLGIDSVLDAAV----
G0046A          ILTISTVHDAAV----
CSN9C           ----------------
CHN23A          LLRHPAVRDAAI----
CHN210A         MLRHPVVRDAAV----
G0047A          MLRHPVVRDAAI----
CS.NRPS         MLRHAIVRDAAI----
CSN9H           MLCQPDVLDAAV----
CHN28A          MLGHPAVLDAAV----
CHN22A          VLAH------------
CSN9B           VLGHGAVLDAAA----
GLRG_00469      ILSQGGVLDAAA----
G0044A          LLNHADVLDAAV----
CSN9E           LVQHKEVLDAAV----
CHN25A          LLTHADVLDAAV----
G0045A          MLAQKAVLDAAV----
CHN26A          ILRHTAASDAAV----
CSN9F           ILRH------------
CH.NRPS3        IRTRVPEISDVA----
CSN1C           IRTRVSEVSDVA----
GLRG_06542      IRTRVSEVSDVA----
CHN32A          LTTSGLATSSAV----
CSN1B           L---------------
G0652A          L---------------
CSN1A           AMR-------------
G0651A          AMRTRGCHSAVA----
CHN31A          ----------------
CHN41A          LREASAFQAASV----
CSN7            IRKS------------
CHN62A          ITSLCSSVRKVAV---
CSN2A           IRSNTSILESAV----
CH.NRPS9        VRR-------------
CHN91A          ILELTDVRGVVV----
CH.NRPS5        ISHQKDIS--------
CH.NRPS6        VVAC------------
CHN21A          LRKVLPGHLTPVAEAV
CSN9A           LRKALPKHLVPVAEAV
G0041A          LRKALPKHLVPVAEVV
CSN4B           LAKL------------
CHN102A         LAKL------------
CH.NRPS7        LARC------------
CSN6A           LKQ-------------
G0681A          LKQSLQVA--------
```

```
CSN2B           LRNS------------
CH.NRPS4        LR--------------
CH.NRPS11       LDNLSGIIQ-------
CSN8B           LDNLPDVVRC------
GLRG_00920      MDNLPD----------
CSN2D           IR--------------
CHN44A          IRQ-------------
CSN2C           IREI------------
CHN43A          ----------------
CHN111A         MALDPRVQHC------
CSN8A           LALNPSVQHC------
G0091A          LALNPSVQHC------
CSN4A           LRAAFNTTEIVV----
CH.NRPS10       LRAAFGVTDIVV----
CHN72A          LSGC------------
CSN6B           LAGCLGVNRGVV----
GLRG_06838      LAGCLDVNLAVV----
CHN87A          ----------------
CH063_1253      LR--------------
GLRG_08065      LR--------------
CSN5            LRGKMPPGVKIVA---
CHN84A          LR--------------
CHN85A          VRDGLPGCEQAVV---
CHN81A          LRESVPRLQQVAV---
CH.NRPS8        VRECVPSARQVAA---
CHN86A          LRDCVVVSDA------
CHN83A          LR--------------
```

PART 3: TS protein sequences

CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment

```
CH063_0494      --------------------------------------MGVIFY-LLHPSQLRAIIQWKV
GLRG_07002      --------------------------------------MGVIFY-LLHPNQLRSIIQWKV
CSUB_TS_8       --------------------------------------MGIIFY-LLHPNQLRSIIQWKV
CH063_0492      ---------------------------------------------MNRTRIAGHGLRRLI
GLRG_10556      ---------------------------------------------MNRTRIAGHGLRRLV
CSUB_TS_12      ---------------------------------------------MNRTRIASHGLRRLV
CH063_01_0      -----------------------------------------------------MAAKTFH
CH063_0916      --------------------------------------------------MISLQSLASF
CSUB_TS_17      -----------------------------------------------MTAVLPFKDDEIH
CSUB_TS_15      ----------------------------------------MIES-PTAAADARAERRSKY
CH063_0378      ------------------------------------------------------------
GLRG_02475      MGYDYALVHVKYTIPLAALLTVFSYPVFTRLDVVRTLFIVTIAFVATIPWDSYLIRTNVW
CSUB_TS_2       MGYDYALVHVKYTIPLAALLTFFSYPVFTRLDVVRTLFIVTIAFVATIPWDSYLIRTGIW
CH063_0666      ------------------------------------------------------------
GLRG_03139      ------------------------------------------------------------
CSUB_TS_4       ------------------------------------------------------------
CH063_01_3      --------------------------------------MQLRSGAAAVLRSTSSITSRSL
GLRG_09287      ---------------------------------------MQLRQGAAVLRSTTSITSRSL
CSUB_TS_11      --------------------------------------MQLRPGAAVLRSTTTSITSRSL
GLRG_10302      ---------------MASTIENYQDCGPMRYKSSVPVPASLYANTAYPSRFRPRISKHVD
CH063_0097      ---------------MASTINNYQDCGPMRYKSSIPVPASLYENTAYPSKFRPRISKHVD
CH063_0347      ----------------------MLSEEELYPYSVPVDRETVVRS-GALTTLPVRIYKHDD
CSUB_TS_16      -------MEVRDGEAKSCGNDHSHSVRIKNLISTAVLECIKVDRAGAMCMLEAYGKKWLR
```

```
CH063_0282    ------------------------------------------------------------
CH063_1125    ---------------------------------MLPATSTSSASSKTTNKSPASAPTPAP
GLRG_03419    ---------------------------------------MLPTTPTLPTSKKAADKSRDL
CSUB_TS_5     ---------------------------------------MLPTTSTLSTSKKAADKSWNP
CH063_0190    --------------------------MDDIWKYSVDIDREVALSTGCFTSLAIRIHQRND
CH063_1346    ------------------------------------------------------------
GLRG_00455    -----------------------------MPYFSLPIPESLIQQSGSHTRFATVRHCNYE
CH063_0519    ----------------------------MASFSRPVPGAIVASS-GSRSQFPTYIHQNFD
CSUB_TS_1     ----------------------------MPSFSRPVPDGIVISS-GLRSRFPPYVHQNYE
GLRG_07956    -----------------------------------MASSVAPVQACLHPALSLQPGRDNP
CSUB_TS_9     -----------------------------------MSSSVVPAPVSLHPALSQNSEKSSS
CH063_0939    ----------------------------------------------------------MA
GLRG_03977    ------------------------------------MCVDQQLFHSGVRGRADTYKKNAK
CSUB_TS_6     ------------------------------------MCVGRQLFHSGARSRANTRKQKSK
GLRG_02961    -----------------------------MISPSQLTPGAHDAAVAAGSPMFRMLGSLWG
CH063_1495    ------------------------------------------------------------
CSUB_TS_3     -----------------------------MISASQLTPGAHDAAVAAASPLFRMLSSLWG
GLRG_11695    ----------------------------------------MKDGLVHPSTSPRREPTLAI
CH063_0640    ------------------------------------------------------------
CSUB_TS_13    ------------------------------------------------------------
CSUB_TS_14    ------------------------------------------------------------
GLRG_08962    ------------------------------------------------------------
CSUB_TS_10    ------------------------------------------------------------
CH063_1011    ------------------------------------------------------------
GLRG_06676    ------------------------------------------------------------
CSUB_TS_7     ------------------------------------------------------------


CH063_0494    WHDPVHRRDPSKEDPTLTACFEHLNKTSRSFSA---VIQELNPELLVPVCLFYLVLRGL-
GLRG_07002    WHEPVHRRDPSKEDPTLRACFEHLNKTSRSFSA---VIQELNPELLVPVCLFYLVLRGL-
CSUB_TS_8     WHEPVHRRDPSKEAPTLRACFEHLNKTSRSFSA---VIQELNPELLVPVCLFYLVLRGL-
CH063_0492    QAPPRGTRGIVTEADVVKARKYCQSQLQHSDYD---AHLISRLVPVRTADAYLALRSLN-
GLRG_10556    KTPPRGTRGIVTDADVVKARKYCQTQLQHSDYD---AHLISRLVPARTTDAYLALRNLN-
CSUB_TS_12    KTPPRGTRGIVTDADVVKARKYCQTQLQHSDYD---AHLISRLVPARTADAYLALRNLN-
CH063_01_0    NFPLLPDELRDEIWDKAVRRQGYRGVHYFSVFD---ATSYCTRPVPEEFTNEYLLARVS-
CH063_0916    ILPLSTASPKTLAFAESEPKQYNPANGIAGALF---KAKIHPREPEISAEVDSFFLDHWP
CSUB_TS_17    NGLSPIGQPLKNGNGITKQKLLNGCRTIPLSSW---APVCHPLVDPVTAEVDAWFLDNWA
CSUB_TS_15    HEADVVVVGAGVFGTAIAYALAQQGRSVILLER---WMKEPNRIVGELLQPGGIVALRQ-
CH063_0378    -MSSVSLRVVEGTEGAMTEGIVSSVAREAXFFG---SLEIEEASFFLVTNMLIVFGLAA-
GLRG_02475    TYPPDAVLGPTLYDIPAEELFFFIIQTYITAQL---YIILNKPVLHAQYLNSPATLPQW-
CSUB_TS_2     TYPPDAILGPTLYDIPTEELFFFIIQTYITAQL---YIILNKPVLHAQYLNSPATLPQW-
CH063_0666    MAKKTTLAEFESVFPKLEEVLLEHAQKYKLPKE---ELAWYKKASIQPHQPAEVHRHRR-
GLRG_03139    MAKKTTLAEFESVFPKLEEVLLEHAQKYKLPKE---ELAWYKKASIQPHRPEKLHYSHH-
CSUB_TS_4     MAKKTTLAEFESVFPKLEEVLLEHAQKYKLPKE---ELAWQAYSHTNPTSPATTTTTNT-
CH063_01_3    LSSTASRCAQCRRIAPLEVRLPLTSALYHSSSR---RSSAWGAAVSVASNMAANAVNRV-
GLRG_09287    LSSSSSRCAQCRRTATLDLRLPLTSALYHSSSK---RSSAWGAAVSVASNMAANAVNRV-
CSUB_TS_11    LSSSSSRCPQCRRSAPLDLRLPLTSALYHSSSK---RSSAWGAAVSVASNMAANAVNRV-
GLRG_10302    VADKACWEACDDFENATGLKLKADSVGCINPIGGNVNALWFPEAIPERLHIISYLSELL-
CH063_0097    VADKACWEACDDFENATGLKLKADSVGCINPIGGNVNALWFPEAIPERLHIISYLSELL-
CH063_0347    LADAGAICLTGDWGRIMRDGQDKKSNGSPCVVG---NWG--SFIWPESIPDRMGLLCYL-
CSUB_TS_16    VMETYKSNEIDNIDDYFEARANNGGMGAYYAML---EFSLGIIITDSEYDLMSEAIRHV-
CH063_0282    ---------------------------------------MSPAEIIRPAAKVQLMDSKA-
CH063_1125    TNNSSLPPSPPNINPSRKSSIRQSSAVPVSFSA---HTSAPLPATSSKAPLRTVIEADW-
GLRG_03419    TNNPSFPSPPPNFIPSRKSSIGYTSTVP----G---SLAAAPTPTTSRVPFRPVLETDW-
CSUB_TS_5     TNNPSFPPPPPNFIPSRKSSI-GHTSAVPVSLS---THAAAPTPSSSRAPLRLVPETDW-
CH063_0190    LADEATKLSIQDWGSHVGDGWERKSGSSWSPVG-----NWGAFIFPESLPERLGVITYL-
CH063_1346    ------------------------------------------------------------
GLRG_00455    ICTKAAEQVQAEFNQAMDTEIITNTVAPVPGIG---HLHAVAFTIPNCLPERLSVLTRF-
CH063_0519    KCVEAAKETEHEYNQAMSTNIKSNTLAEIPGLG---LVHPMALTIANCLPDRLAAITRF-
CSUB_TS_1     MCVEAAKGMEDEYNQTMSTNIKSRTLANVPGLG---LVHPMALTIANCLPERLAAITRF-
GLRG_07956    GRNSQPQNSQGQTRAQILRQLHGQDVRIPDVGK---MMADWPRVQNMHLEDVNSRILEI-
CSUB_TS_9     RGNSSAKNSQGLTREQILWQLHGQNVQIPDLGK---LMADWPNDQSMHLEAVNSKILRI-
```

```
CH063_0939     NTTADATNPKTIETQAMITALKGTAIRVPNLYG---ILKGWPVKANINYKRLIPVIEST-
GLRG_03977     DTNASKMSPKTQEARVISEALKGQTIHVPNLYS---ILKGWPVKANVNYERLVPVVEDA-
CSUB_TS_6      ERVAEKRSSKSQETRVIIEALKGQAIHVPNLYS---IFKGWPVKANINYERLVPVVEDA-
GLRG_02961     MFSPKRTRSEDDLATAPERLASLVKPICTDMLK---GLR-YPGVPQIKEESVEALLQYM-
CH063_1495     -------------------------------ML---EELSYPGVPKLKRESVEALLKYM-
CSUB_TS_3      LFAPKRIRLEDDLASAPDRLASLVKPICTDMLK---GMK-YPGAPQLKKESVEALLKYM-
GLRG_11695     DVELNDMAKSPVVTPTVGCLRQRLSYLISEYLK---CINYTRPPKV----DQEALRDAL-
CH063_0640     ------MSESPAVTPTVEGLRHHLSCLIPDFLK---CINYTQPPKA----DQDALREAL-
CSUB_TS_13     -------MTGSVFLQNAESPRQRLSILIPKFLK---SINYIPPSNV----NKQVLRQAV-
CSUB_TS_14     -------MPGPPVSQAAEGVRQRLSILIPKFLQ---SINYTPPLAVNKQELREAV-----
GLRG_08962     ------------------------------------------MQKIVNACDFSYFCSVA-
CSUB_TS_10     ---MSTTNISTQVFPFDDSRLHLSGGH-----G---AVADWCSVFDSGYLRSDLETSRQ-
CH063_1011     ------------------------------------------------------------
GLRG_06676     ------MAVTQTESAQLASQLDGRVMVIPDLRR---MLSHWPSGRNIHAPDIEVLINDL-
CSUB_TS_7      ------MAVMQTESAQLASQLDGRVMIIPDLGR---MLSHWPSGKNIHAPDIEVLINGL-


CH063_0494     DTIEDDMTIDI-----KEKEPLLRQFDQYMEQD--GWTFTKNGPNEKDRDLLV-------
GLRG_07002     DTIEDDMTLSI-----EEKEPLLREFDQYMVKD--GWTFTKNGPNEKDRDLLV-------
CSUB_TS_8      DTIEDDMTIDI-----KEKEPLLREFDQYMEKD--GWTFTKNGPNEKDRELLV-------
CH063_0492     LELVRLPELVS-----NPAIGQMRMQFWRESID--KTFA---------------------
GLRG_10556     LELVRLPEQVS-----NPVIGQMRMQFWRESID--KTFA---------------------
CSUB_TS_12     LELVRLPEQVS-----NPVIGQMRMQFWRDSID--KTFA---------------------
CH063_01_0     EKQLHVFGAPT-----SSGNPSSPSWTIGNQST---------------------------
CH063_0916     FVNEEARKRFV-----AAGFSKVTCFYYPHALD---------------------------
CSUB_TS_17     FPNEKARKKFV-----AAGFSKVTCLYFPFAKD---------------------------
CSUB_TS_15     LGLADALEGID-----AVPCYGYKVSYHGEGVD--VPYPSFDENGHVIHPSAN-------
CH063_0378     FDKAVAVCDAF-----PDKFDK--------------------------------------
GLRG_02475     IKSGKLVGQLA-----LSGSVLLGTWLIAKKGE--GTYLGLILVWACTFALFT-------
CSUB_TS_2      IKSGKLFGQLA-----LFGSVLLGTWLIAKEGE--GTYLGLILVWACTFALFT-------
CH063_0666     RDDLEPPGRPS-----CWRSLEINTLGGKCNRG---------------------------
GLRG_03139     HHHQHCRHHDN-----DRCSLEINTLGGKCNRG---------------------------
CSUB_TS_4      TTTDTPTTRLG-----CR-SLEINTLGGKCNRG---------------------------
CH063_01_3     IPKGDMHIDPL-----RTVAKEMKFLTGNIRKLLGSGHPSLDRVAKYYTQAEG-------
GLRG_09287     LPKGDMHIDPL-----RTVAKEMKFLTGNIRKLLGSGHPSLDRAAKYYTQAEG-------
CSUB_TS_11     IPKGDMHIDPL-----RTVAKEMKFLTGNIRKLLGSGHPSLDRAAKYYTQAEG-------
GLRG_10302     FRHDDLTDDAVTPEQFDEVHGPLARFLGSESKQSDHTTKHNAMNTMQARVAIE-------
CH063_0097     FRHDDLTDDAITPEQFDEVHGPLARFLGSESKQSDHTTKHNAMNTMQARVAIE-------
CH063_0347     LDVGCFHDDAC-----EEMTIA-AAHAEHLDLD--AAMDVEDNRSLSNDSRSA---KTKE
CSUB_TS_16     ERCMLLTNDYW-----SWPREREQAKHQETGKV--FNIVWFLMHTNKSWSEDD-------
CH063_0282     HGNNH-------------------------------------------------------
CH063_1125     LARANYTSSNS-----GNHPYSHKHSHSHSSHQ--SRRPSFANQTTTSFTLTM-------
GLRG_03419     LSRPRCASSSR-----THSRSQSSHRPRQPSLG--NHATSFTVTMQ--------------
CSUB_TS_5      LSRPNCVFSSR-----THSRSHSSHRPRRPSLGNQATSFTITMQPH--------------
CH063_0190     ANMGNIHDDLC-----DDLTYEEALKEHNNLSQ---AMEISAIAPHQGPKALDRSMKMKK
CH063_1346     ------------------------------------------------------------
GLRG_00455     AEFTILNDDFY-----DIAKNE-DIHKTNDDIQ-----SVLNGAIESSRSIESNMTKSKQ
CH063_0519     ADFTILNDDYY-----DSAKKE-EIQEVNDGIQ--SAIRGFSSPSALPTPSSSSAFKAKQ
CSUB_TS_1      ADFTILNDDYY-----DFAKKE-EIQALNDGIQ--SAIQDVSAAGIKSEASSTSTFKSKQ
GLRG_07956     IETHAINDTVR-----MRLTKAMLSSQ---------------------------------
CSUB_TS_9      IETHAVNDTVR-----MRLTKAMLSSQ---------------------------------
CH063_0939     FDRLVKPSQLR-----EKYRKA--------------------------------------
GLRG_03977     FDRQSPNLREK-----YRRAD---------------------------------------
CSUB_TS_6      FDRQSPNLREK-----YRRAN---------------------------------------
GLRG_02961     YKRAVEIGYPL-----DTPISA-KAFRLGHSLGLVNYLPRPSHYLTASHRRIV-------
CH063_1495     YKRAVDIGYPL-----DTPISA-KAFRLGYSLG---------------------------
CSUB_TS_3      YERAVDIGYPL-----DTPISA-KAFRLGYSLGLVRCHHLANSHCV--------------
GLRG_11695     LERGRQAGVHI-----DPDDGSNMRFEAGLAVA---------------------------
CH063_0640     LERGRQAGVHV-----EPEDGSNMRFEAGLAVA--AVCKPDGSMREIDK-----------
CSUB_TS_13     LERGRQAGVDV-----VQKDGSKMRFEAGIAVA---------------------------
CSUB_TS_14     LERGRQSGVSV-----DSDNGSVMRFETGISVA---------------------------
GLRG_08962     AP----------------------------------------------------------
```

```
CSUB_TS_10      V-----------------------------------------------------------
CH063_1011      ------------------------------------------------------------
GLRG_06676      LDKDDAPAQDI-----ANIREA--------------------------------------
CSUB_TS_7       LNKNGASPQDI-----ANITEA--------------------------------------


CH063_0494      ------------------------------------------------------------
GLRG_07002      ------------------------------------------------------------
CSUB_TS_8       ------------------------------------------------------------
CH063_0492      ------------------------------------------------------------
GLRG_10556      ------------------------------------------------------------
CSUB_TS_12      ------------------------------------------------------------
CH063_01_0      ------------------------------------------------------------
CH063_0916      ------------------------------------------------------------
CSUB_TS_17      ------------------------------------------------------------
CSUB_TS_15      ------------------------------------------------------------
CH063_0378      ------------------------------------------------------------
GLRG_02475      ------------------------------------------------------------
CSUB_TS_2       ------------------------------------------------------------
CH063_0666      ------------------------------------------------------------
GLRG_03139      ------------------------------------------------------------
CSUB_TS_4       ------------------------------------------------------------
CH063_01_3      ------------------------------------------------------------
GLRG_09287      ------------------------------------------------------------
CSUB_TS_11      ------------------------------------------------------------
GLRG_10302      ---------ALEQDERLGKLVIEKWKGIVSVRGQDAFMEHKTLDSYMHVRHYDAGAYSVW
CH063_0097      ---------ALEQDERLGKLVIEKWKGIVSVRGQDAFMEHKTLDSYMHVRHYDAGAYSVW
CH063_0347      LVSMAILECVKVDRVGALRMLEAYRKKWLAIMETYNTEEIDNLDDYFFSRANNGGMGAYY
CSUB_TS_16      ------------------------------------------------------------
CH063_0282      ------------------------------------------------------------
CH063_1125      ------------------------------------------------------------
GLRG_03419      ------------------------------------------------------------
CSUB_TS_5       ------------------------------------------------------------
CH063_0190      YISKCLLEAMNIDRPRALRMINTYRSKWLDVMERRDVNEIETLDEYLVFRNLNGGMEAFW
CH063_1346      ------------------------------------------------------------
GLRG_00455      FQSSLILDMVNIDADLAMDIMTTYSKGL--DLATFAPDTLKTLDDYLPIRMVNSGLDVFQ
CH063_0519      LQAGFMLELFSLDQEFALHIMSSYSQGL--DIATFAPDDLKTLDDYLPVRSINSGLDVTE
CSUB_TS_1       LQAALILDLMILDQEFAMDIMSSYSQGL--EIATFAPDDLKTLDDYLPVRSINSGLDVTA
GLRG_07956      ------------------------------------------------------------
CSUB_TS_9       ------------------------------------------------------------
CH063_0939      ------------------------------------------------------------
GLRG_03977      ------------------------------------------------------------
CSUB_TS_6       ------------------------------------------------------------
GLRG_02961      ------------------------------------------------------------
CH063_1495      ------------------------------------------------------------
CSUB_TS_3       ------------------------------------------------------------
GLRG_11695      ------------------------------------------------------------
CH063_0640      ------------------------------------------------------------
CSUB_TS_13      ------------------------------------------------------------
CSUB_TS_14      ------------------------------------------------------------
GLRG_08962      ------------------------------------------------------------
CSUB_TS_10      ------------------------------------------------------------
CH063_1011      ------------------------------------------------------------
GLRG_06676      ------------------------------------------------------------
CSUB_TS_7       ------------------------------------------------------------


CH063_0494      ------------------------------------------------------------
GLRG_07002      ------------------------------------------------------------
CSUB_TS_8       ------------------------------------------------------------
CH063_0492      ------------------------------------------------------------
GLRG_10556      ------------------------------------------------------------
```

```
CSUB_TS_12     ------------------------------------------------------------
CH063_01_0     ------------------------------------------------------------
CH063_0916     ------------------------------------------------------------
CSUB_TS_17     ------------------------------------------------------------
CSUB_TS_15     ------------------------------------------------------------
CH063_0378     ------------------------------------------------------------
GLRG_02475     ------------------------------------------------------------
CSUB_TS_2      ------------------------------------------------------------
CH063_0666     ------------------------------------------------------------
GLRG_03139     ------------------------------------------------------------
CSUB_TS_4      ------------------------------------------------------------
CH063_01_3     ------------------------------------------------------------
GLRG_09287     ------------------------------------------------------------
CSUB_TS_11     ------------------------------------------------------------
GLRG_10302     SQILFCCDISLTDEELTSLEPLTWLAFTQMILWHDYCSWDKEAATYLEREEGGSNMSAVQ
CH063_0097     SQILFCCDMSLTDEELTGLEPLTWLAFTQMILWHDYCSWDKEAATYLEREEGGSNMSAVQ
CH063_0347     AMLEFSLGIVVTDEEYEMMAAPIKHVERCMLLTNDYWSWPREREQAKTQEAGKVFNTVWF
CSUB_TS_16     ------------------------------------------------------------
CH063_0282     ------------------------------------------------------------
CH063_1125     ------------------------------------------------------------
GLRG_03419     ------------------------------------------------------------
CSUB_TS_5      ------------------------------------------------------------
CH063_0190     SMVEFGMAMDVSDSDKTRIRPLFAAAESALVLTNDYWSWDREWRQAQQAEESRIVNAVHL
CH063_1346     ------------------------------------------------------------
GLRG_00455     TMSCFGMGIKLSADDKEKLSEFVNTAMFSTTLINDLHSWPKEVKHHIEHPGSEYPFNAVA
CH063_0519     DMACFGTGVRISRAEKEKLRKATDMAKYAITIVNDLYSWPKEIKCHLETPGSNPPFNAVA
CSUB_TS_1      DMACFGTGVRVSKSEKEKLRGIVDLTKFAITIVNDLYSWPKEIKCHLETPDSKPPFNAVA
GLRG_07956     ------------------------------------------------------------
CSUB_TS_9      ------------------------------------------------------------
CH063_0939     ------------------------------------------------------------
GLRG_03977     ------------------------------------------------------------
CSUB_TS_6      ------------------------------------------------------------
GLRG_02961     ------------------------------------------------------------
CH063_1495     ------------------------------------------------------------
CSUB_TS_3      ------------------------------------------------------------
GLRG_11695     ------------------------------------------------------------
CH063_0640     ------------------------------------------------------------
CSUB_TS_13     ------------------------------------------------------------
CSUB_TS_14     ------------------------------------------------------------
GLRG_08962     ------------------------------------------------------------
CSUB_TS_10     ------------------------------------------------------------
CH063_1011     ------------------------------------------------------------
GLRG_06676     ------------------------------------------------------------
CSUB_TS_7      ------------------------------------------------------------


CH063_0494     ------------------------------------------------------------
GLRG_07002     ------------------------------------------------------------
CSUB_TS_8      ------------------------------------------------------------
CH063_0492     ------------------------------------------------------------
GLRG_10556     ------------------------------------------------------------
CSUB_TS_12     ------------------------------------------------------------
CH063_01_0     ------------------------------------------------------------
CH063_0916     ------------------------------------------------------------
CSUB_TS_17     ------------------------------------------------------------
CSUB_TS_15     ------------------------------------------------------------
CH063_0378     ------------------------------------------------------------
GLRG_02475     -------------------------------------------------WTITAHFLLAL
CSUB_TS_2      -------------------------------------------------WTITAHFLLAL
CH063_0666     ------------------------------------------------------------
GLRG_03139     ------------------------------------------------------------
CSUB_TS_4      ------------------------------------------------------------
```

```
CH063_01_3     ------------------------------------------------------------
GLRG_09287     ------------------------------------------------------------
CSUB_TS_11     ------------------------------------------------------------
GLRG_10302     VYMAMYGLDQHAAKDFLLSEIGRIEDEYCERKAVYMAECSPAHHITHYIGLIELCMAGNT
CH063_0097     VYMAMYGLDQYAAKDFLLSEIGRIEDEYCERKAIYMAECSPAHHITHYIGLIELCMAGNT
CH063_0347     LMKQERC-SEAEAKLKVADMVAAEEARWVEAKRKIYQEHPD-LRADLVKFLENLHTALAG
CSUB_TS_16     -------------AIVKVRKMVYEEEQKWVEAKKSLYQKVPNLGRDAIKFLENLHTTLAG
CH063_0282     ------------------------------------------------------------
CH063_1125     ------------------------------------------------------------
GLRG_03419     ------------------------------------------------------------
CSUB_TS_5      ------------------------------------------------------------
CH063_0190     FMRTHGLAMHEAREAVRDRILAYEAE--YLRLKTEFYSQNPNVGADLKKYIEVCGVITAG
CH063_1346     ------------------------------------------------------------
GLRG_00455     ILMRHGGLSEAEAFLRLREKQAELQERHLALLSAL--EAAGPIPESHMLYILAAQYAASG
CH063_0519     VLMRHSGYSESEAFQALADKQAELEDKHLRLVEALREQEGGSLPENQERYIANAQQAVSG
CSUB_TS_1      VLMRHGGYSESEAFKILCVKQAKLEEEHLRLLEALRQQEGGRLPENQELYVANAQQAVSG
GLRG_07956     ------------------------------------------------------------
CSUB_TS_9      ------------------------------------------------------------
CH063_0939     ------------------------------------------------------------
GLRG_03977     ------------------------------------------------------------
CSUB_TS_6      ------------------------------------------------------------
GLRG_02961     ------------------------------------------------------------
CH063_1495     ------------------------------------------------------------
CSUB_TS_3      ------------------------------------------------------------
GLRG_11695     ------------------------------------------------------------
CH063_0640     ------------------------------------------------------------
CSUB_TS_13     ------------------------------------------------------------
CSUB_TS_14     ------------------------------------------------------------
GLRG_08962     ------------------------------------------------------------
CSUB_TS_10     ------------------------------------------------------------
CH063_1011     ------------------------------------------------------------
GLRG_06676     ------------------------------------------------------------
CSUB_TS_7      ------------------------------------------------------------


CH063_0494     ------------------------------------------------------------
GLRG_07002     ------------------------------------------------------------
CSUB_TS_8      ------------------------------------------------------------
CH063_0492     ------------------------------------------------------------
GLRG_10556     ------------------------------------------------------------
CSUB_TS_12     ------------------------------------------------------------
CH063_01_0     ------------------------------------------------------------
CH063_0916     ------------------------------------------------------------
CSUB_TS_17     ------------------------------------------------------------
CSUB_TS_15     --------NETTSSSAKQKEGRSFHHGRFIMNL---------------------------
CH063_0378     ------------------------------------------------------------
GLRG_02475     PLACTALPILLPTVYLWIVDEMALGRGTWAIES---------------------------
CSUB_TS_2      PLACTALPILLPTVYLWIVDEMALGRGTWAIES---------------------------
CH063_0666     ------------------------------------------------------------
GLRG_03139     ------------------------------------------------------------
CSUB_TS_4      ------------------------------------------------------------
CH063_01_3     --------KHMRPLIVLLMSRATALCPKGPRLH---------------------------
GLRG_09287     --------KHMRPLIVLLMARATALCPKGPRLH---------------------------
CSUB_TS_11     --------KHMRPLIVLLMARATALCPKGPRLP---------------------------
GLRG_10302     LWHLSSRRYDPAAPLPRREDIGKVNQVPLDVSE---------------------------
CH063_0097     LWHLSSRRYDPAAPLPRREDIGKVNRGPLDVFE---------------------------
CH063_0347     NDYWSSQCYRHNDWAHVPEQPSENHPKVHELAALGRAVMPAECSGSGLSNAPKSEELGGV
CSUB_TS_16     NDYWSSQCYRHNDWKHVPELPGKCAPRVNDLVG---------------------------
CH063_0282     ------------------------------------------------------------
CH063_1125     ------QPPPPPDPERYATSDLNYTNKTWPAAK---------------------------
GLRG_03419     -------PPPPPDPERYATSDLNFTRKTWPEEK---------------------------
CSUB_TS_5      ---------PPPDPSRYATSDLNFTRKTWPEEK---------------------------
```

```
CH063_0190    NHYWCANCPRHHSWRDQDAPLDPIGRRLSVSSE---------------------------
CH063_1346    ------------------------------------------------------------
GLRG_00455    SEFWSVHVPRYPSKKDLAQPDVEFVNGEFLYRT--------GPATTDIDSIPRPILAKGT
CH063_0519    SELWSVFTTRYPSKADLQQPPVEFVDGKLRYAS---------------------------
CSUB_TS_1     SELWSIFTTRYPSKADLQQAEVVFMDGKFGYVS---EGRKVDGVEPGTNTTASTELEDQL
GLRG_07956    ------------------------------------------------------------
CSUB_TS_9     ------------------------------------------------------------
CH063_0939    ------------------------------------------------------------
GLRG_03977    ------------------------------------------------------------
CSUB_TS_6     ------------------------------------------------------------
GLRG_02961    ------------------------------------------------------------
CH063_1495    ------------------------------------------------------------
CSUB_TS_3     ------------------------------------------------------------
GLRG_11695    ------------------------------------------------------------
CH063_0640    ------------------------------------------------------------
CSUB_TS_13    ------------------------------------------------------------
CSUB_TS_14    ------------------------------------------------------------
GLRG_08962    ------------------------------------------------------------
CSUB_TS_10    ------------------------------------------------------------
CH063_1011    ------------------------------------------------------------
GLRG_06676    ------------------------------------------------------------
CSUB_TS_7     ------------------------------------------------------------


CH063_0494    ------------------------------------------------------------
GLRG_07002    ------------------------------------------------------------
CSUB_TS_8     ------------------------------------------------------------
CH063_0492    ------------------------------------------------------------
GLRG_10556    ------------------------------------------------------------
CSUB_TS_12    ------------------------------------------------------------
CH063_01_0    ------------------------------------------------------------
CH063_0916    ------------------------------------------------------------
CSUB_TS_17    ------------------------------------------------------------
CSUB_TS_15    ------------------------------------------------------RKACQK
CH063_0378    ------------------------------------------------------------
GLRG_02475    ---------------------------GTKLELQLFGSLEIEEATFFLVTNMLIVFGIAA
CSUB_TS_2     ---------------------------GTKLELQLFGSLEIEEATFFLVTNMLIVFGIAA
CH063_0666    ------------------------------------------------------------
GLRG_03139    ------------------------------------------------------------
CSUB_TS_4     ------------------------------------------------------------
CH063_01_3    ------------------------------------------------------------
GLRG_09287    ------------------------------------------------------------
CSUB_TS_11    ------------------------------------------------------------
GLRG_10302    VSKPVDASGSEGILTPASSCLGTKRLRPFWNNQRTEYTTMTPAETNSDH--KKKAKASHE
CH063_0097    VSKPVDSSESDGILTPVSSRLGTKRLRPFWNNQRTEYTTMTPTETSSDHKKKKKAKASHE
CH063_0347    IAESDLDFEANAASLLSFLDKASPQDQQVEVEVGSSASASEANTQSTAELSATLRSNSVS
CSUB_TS_16    LGRALLRNESDLDDADNVAATVHSHKSSPGSPAFHDGQSNHHSSTLSTHSQSSCTTPPSP
CH063_0282    ------------------------------------------------------------
CH063_1125    ------------------------------------------------------------
GLRG_03419    ------------------------------------------------------------
CSUB_TS_5     ------------------------------------------------------------
CH063_0190    ---------------------------DTVEDDCATSPGATTSSSMSQKSSPTTEMTLSE
CH063_1346    ------------------------------------------------------------
GLRG_00455    VATERAGLTNGHTNGSINGSSNGSSNGFSNGITNSHTNGLTNGLANGHANGHARHHTNGN
CH063_0519    -----------------SSSSACENKSHSTGDLATIETTCDSQSITACPSEKEYKRTKAL
CSUB_TS_1     ASLEVASIDEASFTPSISSEPGCDDKDDIKRDAIHEEIDQDSQSITIPRQGHGHGRMKAL
GLRG_07956    ------------------------------------------------------------
CSUB_TS_9     ------------------------------------------------------------
CH063_0939    ------------------------------------------------------------
GLRG_03977    ------------------------------------------------------------
CSUB_TS_6     ------------------------------------------------------------
GLRG_02961    ------------------------------------------------------------
```

```
CH063_1495      ------------------------------------------------------------
CSUB_TS_3       ------------------------------------------------------------
GLRG_11695      ------------------------------------------------------------
CH063_0640      ------------------------------------------------------------
CSUB_TS_13      ------------------------------------------------------------
CSUB_TS_14      ------------------------------------------------------------
GLRG_08962      ------------------------------------------------------------
CSUB_TS_10      ------------------------------------------------------------
CH063_1011      ------------------------------------------------------------
GLRG_06676      ------------------------------------------------------------
CSUB_TS_7       ------------------------------------------------------------


CH063_0494      ----------------HFDDVITEMKKIKKPYYDIIKDITIKMGNGMADYAKNAAFNEN-
GLRG_07002      ----------------HFDDVITELKKIKKPYFDVIKDITIKMGNGMADYAKNAAFNEQ-
CSUB_TS_8       ----------------HFDDVIAELKKIKKPYYDVIKDITIKMGNGMADYAKNAAFNEQ-
CH063_0492      ------------------GNPPAEPICLLLHQ--ALQDLRARSTSGTASSIKFWV-----
GLRG_10556      ------------------GKPPAEPICVLLHQ--ALQDLRARSTASTASSIKFWV-----
CSUB_TS_12      ------------------GNPPAEPICVLLHQ--ALQDLRARSTTSTASSIKFWV-----
CH063_01_0      --------------------YAIDAGLWTACRESRAAMYRRYTPEKWANWYTEP------
CH063_0916      --------------------------------------DRISLACRLLTLLFL-------
CSUB_TS_17      ------------------DRIAFACRLLAILF--LIDNAAGLGWAGLGWTGLDWADNQLT
CSUB_TS_15      QKNITIFETEVTATIRGDDKDTVLGVRSKTTD--PATGEKKDDF--FFGQLTIIADGY--
CH063_0378      -----------------PADALSMSLLRARVF--PSSKYDMQRILGIRQAVARLAKKSRS
GLRG_02475      FDKAVAVCDAFPEKFDKPADALAMSLLRARVF--PSSKYDMQRILGIRQAAARLAKKSRS
CSUB_TS_2       FDKAVAVCDAFPENFDKPADALSMSLLRARVF--PSSKYDMQRILGIRQAVARLAKKSRS
CH063_0666      ---------------------MSVPDSVSLLLDKPLNEEEYFQAATLGWMTELLQAFFLV
GLRG_03139      ---------------------MSVPDSVSYLLGKPLNEEEYFQAATLGWMTELLQAFFLV
CSUB_TS_4       ---------------------MSVPDSVSLLLDKPLSEEEYFQAATLGWMTELLQAFFLV
CH063_01_3      ----------SGQAIGGVDTAISPPNILVDVN--PSSPLTSAAPEPAEVDSDILPSQRR-
GLRG_09287      ----------SAQPMGGVDTAISPPNILVDVN--PSSPLTSPAVEPTEVDSDILPSQRR-
CSUB_TS_11      ----------SGQPIGGVDTAISPPNILVDVN--PSSPLTSPAIEPTEVDSDILPSQRR-
GLRG_10302      TREDLLTVSPCAWPAELDEKNILAPYLYTAAR--PASGARDKLMDALDNWYRVPPGALA-
CH063_0097      TRADLLTVSPCAWPAEPDEKDILAAYLYTAAR--PASGARDKLMDALDNWYRVPPNALA-
CH063_0347      SA-GSASTSDETKARVSDTDIISAPIQYVQSL--PSKGFRTTLIDCLNRWLEVPQQEME-
CSUB_TS_16      RSPAINPHLPCSSTSGFAKSPVTGPIDYIRSL--PSKKVRTQLIDSLNIWFSVPHSALS-
CH063_0282      ------------------EEIIRAPLNYLLDL--PGKEVRSKLMSAFNQWLRIPEEKLE-
CH063_1125      ------------------ERVVAGPFDYLSAL--PGKDFRAQLIQAFNVWLEVPQESID-
GLRG_03419      ------------------EKVIAGPFDYLSAL--PGKDFRSQLIQAFNVWLEVPQESID-
CSUB_TS_5       ------------------DKVITGPFDYLSAL--PGKDFRSQLIQAFNAWLEVPQESID-
CH063_0190      VLSFTTISGXXEQPQRLSDSPLRAPCQYIRSM--PSKGLRHLMAEALDQWLLVDNASLD-
CH063_1346      ------------------------------------------------------------
GLRG_00455      SKEEETLAVKYPEDVKCRSELVLAPYKYLTSM--PSKGIRELFIRALNWWLKVPDDKLA-
CH063_0519      LPDGSDLSTYASRVAAAPDHPVMAPFKYIASL--PSKGVRDTFIDALNWWLKVPDDSLL-
CSUB_TS_1       LPDGPEFPTYASRVAAAPDHAAVAPFKYIASL--PSKGIRDTFIDALNWWLAVPEDSLR-
GLRG_07956      ---------------------LAGWYPYASCE-------RIKALTSFQAWMFIIDDMLDQ
CSUB_TS_9       ---------------------LAGWYPYASCD-------RIEALTSFQAWMFIIDDMLDQ
CH063_0939      --------------------NYARFVSLYYPH--PKWD-QVRILALYIIWLFCWDDAIDQ
GLRG_03977      ---------------------YARFVSLYYPH--PEWD-RVQTLALYIIWLFCWDDAIDQ
CSUB_TS_6       ---------------------YARFVSLYYPH--PEWN-QVRILALYIIWLFCWDDAIDQ
GLRG_02961      --------------YPLPRKTSQLLTRRQLCH--PNHPTEVQGYVGLFTWLVVQYDDIV-
CH063_1495      --------------------------LARLCH--PGHPTEVQGFVGLFTWLVVQYDDIV-
CSUB_TS_3       --------------VVYPSKTSQLLTSRQLCH--PNHPTEVQGFVGLFTWLVVQYDDIV-
GLRG_11695      ----------------------------AEMY--PLHPFNTQLHIGLFTWLGFIIDDLN-
CH063_0640      ------------------HELFRLTTLVQEMY--PLHPFDIQVHIGLFTWLGFIIDDLN-
CSUB_TS_13      ------------------AWHCMRLTLVKDMY--PLHPLEIQVHIGLFTWLGFIIDDLN-
CSUB_TS_14      ----------------------------ADMY--PLHPFEVQVHIGLFTWLGFIIDDLN-
GLRG_08962      ----------------------------------DAPEGRLQTMCDWGNWVFPFDDMFD-
CSUB_TS_10      ------------------------------------------------------------
CH063_1011      --------------------------------------------------MVLWQGKLD-
GLRG_06676      -----------------------NPTLLAS----SLWPDVSKGKLGTLTFMVLWQGRLD-
CSUB_TS_7       -----------------------NPTLLASSGANSLWPNVSKGKLGTLTLMVLWQGRLD-
```

```
CH063_0494     ----------GVATVKEYELYCHYV---------AGLVGDGLTRLFVEGNMANPKLLERP
GLRG_07002      ----------GVATIKEYELYCHYV---------AGLVGDGLTRLFVEGNMANPKLLERP
CSUB_TS_8      ----------GVATIKEYELYCHYV---------AGLVGDGLTRLFVEGNMANPKLLERP
CH063_0492     -------------------------------------QRLIRTRERHMDNRPYASLAALE
GLRG_10556      -------------------------------------QRLIKTREKHMDNRPYASLAALE
CSUB_TS_12      -------------------------------------QRLIKTREKHMDNRPYASLAALE
CH063_01_0     -------------------------------------FTSFLEHMALRPGNRELPAAFKI
CH063_0916     -------------------------------------VDDILEDMSLQEGSLYNEKLILL
CSUB_TS_17      -------------------------------------PQDLLEHMSLEEGKAYNKRLMPL
CSUB_TS_15      ----------ASKFRKEYLPQAPVV------------KSKFYALELIDAPMPSP--GFGH
CH063_0378     FHLASSVFPGRLRIDLTLLYSYCRL------------ADDLVDDAESSEEAAAWISKLDR
GLRG_02475      FHLASSVFPGRLRIDLTLLYSYCRL------------ADDLVDDAATPEEAAVWISKLDR
CSUB_TS_2      FYLASSVFPGRLRIDLTLLYSYCRL------------ADDLVDDAATPEEAAAWISKLDR
CH063_0666     -------------------------------------SDDIMDSSITRRGKPCWYRHEGV
GLRG_03139      -------------------------------------SDDIMDTSITRRGKPCWYRHEGV
CSUB_TS_4      -------------------------------------SDDIMDTSITRRGKPCWYRHEGV
CH063_01_3     -----------LAEITELIHTASLL------------HDDVIDHSVSRRGSPSANLEFGN
GLRG_09287      -----------LAEITELIHTASLL------------HDDVIDHSVSRRGSPSANLEFGN
CSUB_TS_11      -----------LAEITELIHTASLL------------HDDVIDHSVSRRGSPSANLEFGN
GLRG_10302      ----------TIRTIVRIMHNASLM------------LDDVQDNSPVRRGSPSAHVVFGT
CH063_0097     ----------TIRTIVRIMHNASLM------------LDDVHDNSPMRRGSPSAHVVFGT
CH063_0347     ----------CIKKVINSLHDSSLI------------LDDIEDGAKLRRGFPATHVVYGT
CSUB_TS_16      ----------VIKEVVDCLHDSSLI------------LDDVEDGSDLRRGFPATHVVYGT
CH063_0282     ----------VIKRIVMLLHNASLL------------LDDIQDSSTLRRGLPVSHSIFGI
CH063_1125     ----------VITNVVGMLHTASLL------------IDDVEDSSSLRRGLPVAHNIFGV
GLRG_03419      ----------VITNVVGMLHTASLL------------IDDVEDSSSLRRGLPVAHNIFGV
CSUB_TS_5      ----------VITNVVGMLHTASLL------------IDDVEDSSSLRRGLPVAHNIFGV
CH063_0190     ----------RIKNIIDLLHNSSLI------------LDDIEDDSPLRRGLPATHMVFGR
CH063_1346     -------------------------------------MDDFQDDSPLRRGKVAAHTIFGP
GLRG_00455      ----------CIQDVVSYLHQSSLM------------LDDIEDGSKLRRGQPSTHAVYGI
CH063_0519     ----------SIKTVISMLHDSSLMQ--------VEILDDIEDDSTLRRGSPAAHTIYGT
CSUB_TS_1      ----------SIKTIVSMLHDSSLIPEGLKTDKGVTRLDDIEDDSSLRRGSPAAHMIYGT
GLRG_07956      Y---------SIVEKFDFESLHVLLADC---------RDFVERSLDVANTMEKAPLRYKE
CSUB_TS_9      Y---------SLVEKFDFASLHVMLADC---------RDFVERSLKVSNTTENASMQYKD
CH063_0939     Q---------GTGDLSNDILHAKAH------------RDNTIRVLEHFLGLAPPKTKLSV
GLRG_03977      Q---------GTVDLSNNLLRAKTR------------RDNTIRVLEYVLGLAP---KFGS
CSUB_TS_6      Q---------GTGDLSNDLLRAKTR------------RDNTIRVLEHVLGLDN---KFDS
GLRG_02961      ----------GQNDEMAE-------------------AQLFQERFFKGEKQPNAMLE---
CH063_1495     ----------GQNDEMAE-------------------AQLFQKRFFNGETQPNAMLE---
CSUB_TS_3      ----------GQNDEMAE-------------------AQLFQERFFRGETQPNAMLEG--
GLRG_11695      -------------ADLGTD------------------LDNFQSRFFRGDPQPCAILQ---
CH063_0640     -------------AELGSD------------------LDNFQSRFFRGDTQPCVILQ---
CSUB_TS_13      -------------TVLGKD------------------LEQFQSRFHRGEEQPCALLQ---
CSUB_TS_14      -------------TEFGTD------------------LEQFQSRFHCGEKQPGPVLEC--
GLRG_08962      -------------------------------------SGHLRSDLETSRQVLGSLMSNML
CSUB_TS_10      -------------------------------------LDSLMSNMLGKRCYTGTKL----
CH063_1011     ----------DYIEALEYQS-----------------QAKSREFRFRVKEYIAQYLELSE
GLRG_06676      ----------DYIEALEYES-----------------QGKAKEFRSNAKGYIAHYLQLSE
CSUB_TS_7      ----------DYIEALEYES-----------------QAEAKQFRSNAKEYIAQYLQLSE


CH063_0494     ELTESMGQFLQ--KTNIIRDVHED---------------YLDKRRFWPKEIW------GK
GLRG_07002      ELTESMGQFLQ--KTNIIRDVHED---------------YLDKRRFWPKDIW------SK
CSUB_TS_8      ELTESMGQFLQ--KTNIIRDVHED---------------YLDKRRFWPKDIW------SK
CH063_0492     EYAENTYSTLM--Y----ATLAAM------------------PLRSMHVDHL------AS
GLRG_10556      EYAENTYSTLM--Y----ATLAAM------------------PLRSMHVDHL------AS
CSUB_TS_12      EYAENTYSTLM--Y----ATLAAM------------------PLRSMHVDHL------AS
CH063_01_0     VEGDNSQYFTV--LPFYNLFLVQF---------SSFGPYFKSLGKEMPFSSR------KF
CH063_0916     SRGDIVPDRSVPVEWITYDLWNDL------------RTCDKSLADEILEPVF------SF
CSUB_TS_17      ARGDAQPDRTIQVEAMFYNLWASM------------REKDQHLADQVLEPTF------VF
```

```
CSUB_TS_15     VVIGKAFPVLM--YQIGTHETRALIDVPANIPEASP---AAGGVRGYIKNVVMPTLPPQM
CH063_0378     HLSLLYKDPDAVSAPLAAKYAAEN---------FPA--SALSALDMLPASLLPREPLAEL
GLRG_02475     HLSLLYKDPDATSTPLASKYAAEN---------FPP--SALSALDMLPTSLLPREPLAEL
CSUB_TS_2      HLSLLYKDPDATSAPLASKYAAEN---------FPP--SALSALDMLPTSLLPREPLAEL
CH063_0666     GMVAINDAFLL--EAAIYTLLRKF---------FREHASYVDLLELFHEVTF------QT
GLRG_03139     GMVAINDAFML--EAAIYTLLRKF---------FREHASYVDLLELFHEVTF------QT
CSUB_TS_4      GMVAINDAFML--EAAIYTLLRKF---------FREHASYVDLLELFHEVTF------QT
CH063_01_3     KMAVLAGDFLLGRASVALARLRNS-----------------EVIELLATVIA------NL
GLRG_09287     KMAVLAGDFLLGRASVALARLRNA-----------------EVIELLATVIA------NL
CSUB_TS_11     KMAVLAGDFLLGRASVALARLRNA-----------------EVIELLATVIA------NL
GLRG_10302     AQTTNSASYLMIKCVDLARRLGND------------------TLSCLLSELG------QL
CH063_0097     AQTTNSASYLMIKCVDLARRLGED------------------TLSCLLSELG------QL
CH063_0347     SQAINSATFLY--VQAV-EAIHEL---------ENK-----EMMDVLLGHLK------QL
CSUB_TS_16     GQAVNSATFLY--VQAV-EAVHRL---------VKEGGGRLELMDMLLASLK------EL
CH063_0282     AQTINAANYAF---FLAQQEIPKL---------EDP-----RAFEVFTEELL------NL
CH063_1125     AQTINSANYIY---FCTLQELQRL---------KNP-----KTISIFAEELV------HL
GLRG_03419     AQTINSANYIY--FRAL-QEVQRL---------KNP-----KAITIFAEELV------HL
CSUB_TS_5      AQTINSANYIY---FRALQELQRL---------KNP-----KAITIFAEELV------HL
CH063_0190     AQSINTANYMFVQAVQQAQTLSNP-----------------ACLGTLLGELE------CL
CH063_1346     AQAINASTYTIVKAINQVAEFRSQ-----------------SCVHETTEKIM------AL
GLRG_00455     GQTINSANFVF------VQAFARM---------QSLGHGRAEAIDIFIDEVE------NL
CH063_0519     AQCINAANYMV---VMVLVEIQKL---------RSP-----RKLDILSEELE------NL
CSUB_TS_1      AQCINAANYMV---VMVLAELQKL---------RSP-----AKTSILIEELE------SL
GLRG_07956     HDAVVSFYEYAQAVCKSYSDNRPY-------------------RARIAKEAV------AT
CSUB_TS_9      YDAVVSFEEYAQAVCKSYSDNRPY-------------------RARIAKEAI------VT
CH063_0939     ELQNANPELKMIGDKLQMGYSLEQ---------------RQTFMTQMRRYID------NC
GLRG_03977     DTQFSQADYEL---KVIGDELKKA---------YTQ-EQRQVFMSQMRRYIG------NC
CSUB_TS_6      DIEFAQADYEL---KVIGDELKKA---------YTQ-EQRQVFMNQMRRYIG------NC
GLRG_02961     -----GLASLM-------REAPRW---------FDP-----VMANLLQISTL------KF
CH063_1495     -----GLAGLL-------REAPRL---------FDP-----VMANLLQISTL------KF
CSUB_TS_3      -------------LANLMREAPRW---------FDP-----VLANLLQISTL------KF
GLRG_11695     -----CFASVL-------RSMTDY---------YDP-----VVANLIVLSAL------AF
CH063_0640     -----CFASVL-------RSTTDY---------YDP-----VVANLIVLSAL------AF
CSUB_TS_13     -----CFANLL-------RSTTDF---------YEP-----VVANLIVLSAL------TF
CSUB_TS_14     -------------FANVLRSTTDH---------YEP-----VVANLIVLSAL------AF
GLRG_08962     GKGCDIGTKLP-VVQVHDDIFRRL---------SEVGLTYAGVQRRLAKSME------LY
CSUB_TS_10     -PVVQAHNDIFRRLSEVGVTLRCL-------------EDYAGVQRRFARSME------LY
CH063_1011     GPKEPATSSIISTFQPVARMICQQ---------YNK-----DQRRRLKASLF------EY
GLRG_06676     RPTEPAMTPIISDFQPVAKMICQQ---------YNK-----DQRRKLKTSLF------EY
CSUB_TS_7      GPTEPAMTSVISDFQPVAKMICQQ---------YDK-----DQRRKLKASIF------EY


CH063_0494     HVDTWDDLF--KPENQRKALDCSSEMV-LNALKH-------------------AEECL-F
GLRG_07002     HVDTWDDLF--KPENQRKALECSSEMV-LNALKH-------------------TEECL-F
CSUB_TS_8      HVETWDDLF--KPENQRKALECSSEMV-LNALKH-------------------TEECL-F
CH063_0492     HIGKACGIVAVLR-GI---PVLAAPQQ-QLAKSHA------------------GPGGG-R
GLRG_10556     HIGKACGIVAVLR-GI---PVLAAPQQ-QPAKSHA------------------GPGGG-R
CSUB_TS_12     HIGKACGIVAVLR-GI---PVLAAPQQ-QPAKSHA------------------GPGGG-R
CH063_01_0     GLEGIPHIAVEF-------DPSWTLEE-LDQHE--------------------KDSGWEW
CH063_0916     MRAQTDKSR----------LSIKQLGH-YLKYRE-------------------KDVGK-A
CSUB_TS_17     MRAQTEGIR----------TEITELGQ-YLQYRER------------------DVGKAWN
CSUB_TS_15     RPCFEAALA-----DG---KIPPSMPNSYLPASRQ------------------TAKGM-L
CH063_0378     LKGFEMDLS--FDSST---FPITDPED-LELYAA-------------------RVASTVG
GLRG_02475     LKGFEMDLS--FSNSA---FPIADPED-LELYAA-------------------RVASTVG
CSUB_TS_2      LKGFEMDLS--FSNSA---FPITDPED-LELYAA-------------------RVASTVG
CH063_0666     ELGQLCDLLTAPEEQVN--LDNFSMEK-YRFIVVY------------------KTAYYSF
GLRG_03139     ELGQLCDLLTAPEDKVN--LDNFSMEK-YSFIVIY------------------KTAYYSF
CSUB_TS_4      ELGQLCDLLTAPEDKVN--LDNFSMEK-YSFIVIY------------------KTAYYSF
CH063_01_3     VEGEFMQLKNTEQDER---RPVWSEET-LTYYLQKTYL---------------KTASLIS
GLRG_09287     VEGEFMQLKNTERDER---RPVWSEDA-LAYYLQKTYL---------------KTASLIS
CSUB_TS_11     VEGEFMQLKNTERDER---RPVWSEEA-LTYYLQKTYL---------------KTASLIS
GLRG_10302     HLGQSHDLAWTFH------CRAPSMPEYYSHLEQ-------------------KTGGL-F
```

```
CH063_0097    HLGQSHDLAWTFH------CRAPSLPEYYSHLEQ-------------------KTGGL-F
CH063_0347    FCGQSLDLY--WT-FN---RRCPTEDE-YLDMIGQ------------------KTGAL-L
CSUB_TS_16    FQGQSCDLY--WT-HH---RICPTEKE-YLYMVDR------------------KTGAM-M
CH063_0282    HRGQGMDIY--WR-DA---SICPTEEE-YFTMVSN------------------KTGGL-F
CH063_1125    HRGQGMDLH--WR-DT---LTCPTEED-YLEMVGN------------------KTGGL-F
GLRG_03419    HRGQGMDLF--WR-DT---LTCPTEED-YLEMVGN------------------KTGGL-F
CSUB_TS_5     HRGQGMDLF--WRDT----LTCPTEED-YLEMVGN------------------KTGGL-F
CH063_0190    FIGQSWDLYWKFH------LQIPTEKE-YFAMVDS------------------KTGAM-F
CH063_1346    FQGQAMDLFWTYN------SRCPTMNE-YYRMVDN------------------KTGQL-F
GLRG_00455    HKGQSYDLF--WK-DQ---VHCPSVDE-YFMMIDN------------------KTGGL-F
CH063_0519    FLGQSEDLF--WK-YQ---VECPTTEE-YMEMIEN------------------KTGGL-F
CSUB_TS_1     FLGQSEDLL--WK-YQ---VECPSTDE-YMEMIDNSKRTDQTVYIQRANGIDSETGGL-F
GLRG_07956    LDGYQQEALNRYA------GRIPTLEE-YLGYRQ-------------------ASSCI-M
CSUB_TS_9     LDGYRQEALNRYA------GRVPTLEE-YLGYRE-------------------ASSCI-M
CH063_0939    HEEQKMRLQ----------GTLPDIES-YSELRH-------------------GTAAV-W
GLRG_03977    HEEQTMRLQ----------GTLPSIES-YSELRH-------------------GTAAV-W
CSUB_TS_6     HEEQAMRLQ----------GTLPGIDS-YSELRH-------------------GTAAV-W
GLRG_02961    LTCNLLERHKGFM-NM---NITRAGVK-FPDFVR-------------------DLSGINV
CH063_1495    LTCNLLERHDGFR-NM---PVTRSGVK-FPDFLR-------------------DLSGI-N
CSUB_TS_3     LTCNLLERHKGFL-NM---PVTRAGVK-FPDFVR-------------------DLSGINV
GLRG_11695    VNSNAIELRREYQ-TI---ALTKDALS-WPYYFR-------------------DKEGL-P
CH063_0640    VNSNAIELRREYQ-TI---ALTREALS-WPYYFR-------------------DKEGL-P
CSUB_TS_13    VNCNAIELRHDYQ-TM---IPTKETVN-WPYYFR-------------------DMEGL-P
CSUB_TS_14    VNCNAIEPRRDYQ-AI---IPTKQTAN-WPYYFR-------------------DKEGL-P
GLRG_08962    AEGVARHVETFTS------HRIPSLQD-MLQTRR-------------------LSVGV-A
CSUB_TS_10    AEGVARHIETFTS------HHPPSLQY-MLQTRR-------------------LSVGV-A
CH063_1011    IDSTVQETRYVES------GEVPTDLK-YEMLRK-------------------KTGGT-G
GLRG_06676    IDSTVQEIR--YI-DS---GEVPTDLK-YDMLRK-------------------RTAGT-G
CSUB_TS_7     IDSTVQEMR--YI-DS---GEVPTDLK-YDTLRK-------------------KTAGT-G


CH063_0494    YMAGIKDQSVFNFVAIPQSMAIATLELVFRNPAIFSS-------HIKITKGDACQLMSDS
GLRG_07002    YMAAIKDQSVFNFVAIPQSMAIATLELVFRNPAIFNS-------HIKITKGDACQLMSES
CSUB_TS_8     YMAAIKDQSVFNFVAIPQSMAIATLELVFRNPAIFNS-------HIKITKGDACQLMTES
CH063_0492    QDPALL------------------LPLDVMAEANLRE-----------------EDVFRY
GLRG_10556    QDPALL------------------LPLDVMAEANLRE-----------------EDVFRY
CSUB_TS_12    QDPALL------------------LPLDIMAEASLRE-----------------EDVFRY
CH063_01_0    KYAHVY------------------HPRGAKEEGEPMQ----------PLSHERYAEYDSL
CH063_0916    LLSAL-------------------MRFSMKLHITPAE-------------LKSVDNIELN
CSUB_TS_17    RLLSAL------------------MRFAMDLHLSGEE-------------LAEMKEVEQN
CSUB_TS_15    LLGDAFN-----------------MRHPLTGGGMTVA-------------FNDCVILSDL
CH063_0378    QACLEL------------------VFRHCHHGL-----------------PDYMQAYLRN
GLRG_02475    QACLEL------------------VFCHCQHGL-----------------PDYMKAYLRN
CSUB_TS_2     QACLEL------------------VFGHCQHAL-----------------PDYMQAYLRN
CH063_0666    YLPVA-------------------LALHCLNIATPKN-------------LKQAEDILIP
GLRG_03139    YLPVA-------------------LALHCLNIATPKN-------------LKQAEDILIP
CSUB_TS_4     YLPVA-------------------LALHCLNIATPKN-------------LKQAEDILIP
CH063_01_3    KSC---------------------RASAILGGSDAAS-------------VEAAYAYGKN
GLRG_09287    KSC---------------------RASAILGGSDAAS-------------VEAAYAYGKN
CSUB_TS_11    KSC---------------------RASAILGGSDPAS-------------VEAAYAYGKN
GLRG_10302    RMASRM------------------MRASATQNKH----------------LDACKLMSLL
CH063_0097    RMASRM------------------MRASATQNKH----------------LDACKLMSLL
CH063_0347    SMVSDL------------------M-----VAASPRYRKNSPGQ--QPLALTAFSRFSRL
CSUB_TS_16    QLLVGL------------------MQIAATSSDPRYGSLDPTHLEDKDGARQRLLRFTQL
CH063_0282    RLAVRL------------------MQLASES-------------------DRDYVPLVNV
CH063_1125    RLGIKL------------------------MQAESRS-------------PTDCVELVNL
GLRG_03419    RLGIKL------------------------MQAESRS-------------LTDCVELVNL
CSUB_TS_5     RLGIKL------------------------MQAESRS-------------LTDCVELVNL
CH063_0190    RLLTRL------------------MFHARSVVANDKV-------------AQLLDEMCRL
CH063_1346    SIVTR-------------------LMLDNHECISTKT-------------SAALDKFTTL
GLRG_00455    RLCVRL------------------MEFFATGSTRGIS----------------SEFFVKR
CH063_0519    RLCVR-------------------LLQAESTRNDVRN-------------LDP-RPFVRQ
```

```
CSUB_TS_1       RLCVR-------------------LLQAESSRADVHY-------------LDP-RPFVRQ
GLRG_07956      QVAVNLE-----------------FANGISLPEEVME-------------SEEMRELYRA
CSUB_TS_9       QVAVNLE-----------------FANGISLPEEIME-------------SEEMRELYRA
CH063_0939      TLCALIE-----------------YGLSDNIPEDIRH-------------MGQIQTIWSE
GLRG_03977      TLCALVESVIQEIRYSSLVDQYIRFGLSENISEHIRY-------------MEQIQTIWSE
CSUB_TS_6       TLCALVD-----------------------------------------------------
GLRG_02961      AFAVF-------------------CFPKAQYPDVEQY-------------LEAIPDM---
CH063_1495      VAYAVF------------------CFPKAQYPDVGQY-------------LEAIPDM---
CSUB_TS_3       AFAVF-------------------CFPKAQYPDVEQY-------------LEAIPDM---
GLRG_11695      EVYTYF------------------CFYREACPDISRF-------------MPAAPEM---
CH063_0640      EVYTYF------------------CFYKEVCPDISRF-------------MPAAPEM---
CSUB_TS_13      EAYTYF------------------CFYRELCSDISQF-------------MPAAPEM---
CSUB_TS_14      EVYTYF------------------CFYRDLCPDISQF-------------MPAAPDM---
GLRG_08962      PLYHLVE-----------------YAHSLQLPDEVFE-------------DPAIQALERL
CSUB_TS_10      PLYHLVE-----------------YAHSLQIPDEVFE-------------DPAIQALERL
CH063_1011      PLCSLAE-----------------FATGLQWPSFVFD-------------SHPYKMMLQA
GLRG_06676      PLCALAE-----------------FASGLEWPSTVFN-------------SHPFKMMLQS
CSUB_TS_7       PLCALAERR---------------FASGLEWPSTVFN-------------SHPFKMMLQS


CH063_0494      T--QNLFVVCD--VFRKYTR--RIHKKNDPRDPNFVQISVQC--AKIEQFIDSLFPRQDP
GLRG_07002      T--QNLYVVCE--VFRRYIR--KIHKKNDPRDPNFVQISVQC--AKIEQFIDTLFPKQDP
CSUB_TS_8       T--QNLQVVCE--VFRRYIR--KIHKKNDPRDPNFVQISVQC--GKIEQFIDTLFPRQDP
CH063_0492      G--PQAEGFQD--AVFAVAT--RA-------NDHLITAREMLK-SLQAGQGAGHEFEHQG
GLRG_10556      G--PRAEGFQD--AVFKVAT--RA-------NDHLITAREMLK-NLKAGQEAGHEFEHQG
CSUB_TS_12      G--PRAEGFQD--AVFKVAT--RA-------NDHLITAREMLK-NLKAGQEAGHEFEHQG
CH063_01_0      ILASRSVGLPQGVDLWLIDR--RL--RRKPSPVDNEKEKPSFGRYGIYGTEQLVFESQNC
CH063_0916      C--SRHISVVN--DIYSWEK--EL----KASQTGHKEGAALCSSVSVLTSETNLDFAASK
CSUB_TS_17      C--AKHISIVN--DIYSWEK--EL----KQSQVAVEEGSILCSGVKVLADSTGLSIEAAK
CSUB_TS_15      LHPSRVTDLADPVAIKNVLR--EFHWRRKSLTSIINVLAMAL--YALFAANDRQLRALQM
CH063_0378      T--ARQMGLAL--QFVNIAR---------DVAVDAKIGRVYL--PTMWLKEEGLTSEDVL
GLRG_02475      T--ARQMGLAL--QFVNISR---------DIAVDAKIGRVYL--PTTWLKEEGLTPEDVL
CSUB_TS_2       T--ARQMGLAL--QFVNVAR---------DIAVDAKIGRVYL--PTTWLKEEGLTPEDVL
CH063_0666      L--GEYFQIQD--DYLDNFG---LPEHIGKIGTDIMDNKCSW-------LVNQALQIATP
GLRG_03139      L--GEYFQIQD--DYLDNFG--QP-EHIGKIGTDIMDNKCSW-------LVNQALQIATP
CSUB_TS_4       L--GEYFQIQD--DYLDNFG--QP-EHIGKIGTDIMDNKCSW-------LVNQALQIATP
CH063_01_3      L--GLAFQLVD--DMLDYTR--SEKELGKPAGADLELGLATA--PLLFAWK------TMP
GLRG_09287      L--GLAFQLVD--DMLDYTR--SEKELGKPAGADLELGLATA--PLLFAWK------TMP
CSUB_TS_11      L--GLAFQLVD--DMLDYTR--SEKELGKPAGADLELGLATA--PLLFAWK------TIP
GLRG_10302      ---GRLYQLRD--DYQDITS------ESLSTYDDLDEGSFTL--PLIHALQ-RENEHGDV
CH063_0097      ---GRLYQLRD--DYQDITS------ESLSTYDDLDEGSFTL--PLIHALQ-REDERGDV
CH063_0347      X--GLYYQVRD--DYMNIVS--ADYAGKKGYAEDLDEQKFSY--MLVHMAR-RAPEMMD-
CSUB_TS_16      F--GRFFQVRD--DYLNISSCESSYMEKKGFAEDLDEQKFSY--ILVHMY--ARNPEARD
CH063_0282      M--GLIFQVRD--DYLNLQS--TAYTKNKGFGEDLTEGKFSF--PIIHSIR---SNPSNI
CH063_1125      M--GLIFQIRD--DYMNLSS--KEYSENKGMCEDLTEGKFSF--PIIHSIR---ADPSNL
GLRG_03419      M--GLIFQIRD--DYMNLSS--KEYSDNKGMCEDLTEGKFSF--PIIHSIR---SDPSNL
CSUB_TS_5       M--GLIFQIRD--DYMNLSS--KEYSDNKGMCEDLTEGKFSF--PIIHSIR---CEPSNL
CH063_0190      L--GRFFQIRD--DFMNLNS--SEYSDLKGFCEDLDEGKMSY--PMILVLR--QNPEYQD
CH063_1346      L--GRYFQVRD--DYQNLAS--ADYTKQKGFCEDLDEGKYSV--PLIYTLQTQPDNIQLI
GLRG_00455      L--SRYFQIRD--DYQNLMS--DQYAKEKGFAEDLDEGKISL--PLIYTLQSSPYRDAIS
CH063_0519      L--SLYFQIRD--DYQNLVS--DQYAKQKGFAEDLDEGKISL--PMILTLQ---RMRTRP
CSUB_TS_1       L--SLYFQIRD--DYQNLVS--DQYAQQKGFAEDLDEGKISL--PMILTLQ---RVRTRP
GLRG_07956      A--VAVVWIVN--DIVSLRK--EI---REGFVENL-----------VVLLSQGDIQKGID
CSUB_TS_9       A--VAVVWIIN--DIVSLRK--EI---KEGFVENL-----------VVLLSDGKVQKGID
CH063_0939      T--SRAIWITN--DILSLRK--EIPKEGSESIVNAI--------PILMKHKGISPQQAVD
GLRG_03977      T--SRAVWITN--DILSLRK--EIPKEGSESVVNAI--------PIIMKHEGKCPQQAVD
CSUB_TS_6       ---------TN--DILSLRK--EIPEEGSESVVNAV--------PIIMKHEGKCPQQAVD
GLRG_02961      ---ARFIDISN--DVLSFYK--EE---LSGDTRNYVHNRAMA---------------TGR
CH063_1495      ---ARFIDISN--DVMSFYK--EE---LSGDTRNYVHNRAMA---------------TGK
CSUB_TS_3       ---ARFIDISN--DVLSFYK--EE---LSGDTRNYVHNRAMA---------------TGR
GLRG_11695      ---GKFINLTN--DILS-------------------------------------------
CH063_0640      ---GKFINLTN--DILSFYK--EE------------------------------------
```

```
CSUB_TS_13     ---GKFINLTN--DIVSFYK--EE---KAGEVRNYIHKKAVS---------------LGC
CSUB_TS_14     ---ARFINLTN--DILSFYK--EE---KAGEIRNYIHKKAVS---------------LDC
GLRG_08962     G--ADFVILSN--DILSYRK--EE---SEGCPFNM---------VAASQMNGNSAQEAFD
CSUB_TS_10     G--ADLVILSN--DILSYRK--EE---SEGCPFNM---------VAACRMNGKSAQEAFD
CH063_1011     V--AIIVGLTN--DLLSLGK--EL---RKGRILNAVPVRLWN------SREGGDLASVVR
GLRG_06676     V--AIVVGLTN--DLLSLRK--EL---RKGRTLSAVP-------VRLWNNPDGDLEAVVG
CSUB_TS_7      V--AIVVGLTN--DLLSLKK--EL---RKGRTLSAVP-------VRFWNNPDGDLEAVVE


CH063_0494     KKLGEAAKQKE---------RGQPGMDPGEAFILGGMVLL--TLFFVSALMIGTAWYFGA
GLRG_07002     KMLGGSAKQKE---------RGQPGMDPGEAFVLGGLVLL--TLFLLSGLMIGTAWFFGA
CSUB_TS_8      KKLGESAKQKE---------RGQPGMDPGEAFILGGMVLL--TLFFISGLMIGTAWFFGA
CH063_0492     EAEHVYE------------HEEDDTQRDIRRGFGVLLESV--PAQEYLTNLEGTNFDPWA
GLRG_10556     EAEHVYE------------HEEDDTQQDIRRGFAVLLESV--PAQEYLTNLEGTNFDPWA
CSUB_TS_12     EAEHMYE------------HEENDTQRDIRRGFGVLLESV--PAQEYLTNLEGANFDPWA
CH063_01_0     KYYAVPE-----------------------EFIQEFCFYEEPETISTTIWDFLSELSRLA
CH063_0916     RVLWVMC----------------REWELVHRDLVAKRLRSPEPCSQDLQDYMRGLEFQMS
CSUB_TS_17     TCLWSLV----------------REWELKHELLSSEAHIT--SQFSEAQKLYLKGLEYQM
CSUB_TS_15     GCFQYFQ---------------RGHASEPMALMGGLLHQPSKLAYHFFSVAFLAIWLNAL
CH063_0378     SN------------------PTSEGVEKVRRRILAKAFDHYAEARDSMKWIPSEARGPMV
GLRG_02475     KS------------------PNSEGVGKVRRRILAKALDHYGEARDSMKWIPSEARGPMI
CSUB_TS_2      KS------------------PNSEGVGKVRRRILAKAFDHYGEARDSMKWIPSEARGPMV
CH063_0666     EQRKILE-------------DNYGQKDKAKEAVIKKLFDD--MKLKERYHEFEEKRANEI
GLRG_03139     EQRQILE-------------ENYGQKDKAKEAIVKKLFDD--LKLKERFEAFEEKRANEI
CSUB_TS_4      EQRQILE-------------ENYGQKDKAKEAVVKKLFDD--LKLKERFEAFEEKRANEI
CH063_01_3     ELGALVG-------------RKFEQEGDVARARELVLQSD---GIEQTRALAQDYSERAI
GLRG_09287     ELGALVG-------------RKFEQEGDVARARQLVLQSD---GIEQTRALAEDYSERAI
CSUB_TS_11     ELGALVG-------------RKFEQEGDVARARQLVLQSD---GIEQTRALAEDYSERAI
GLRG_10302     QLHSILQSARAARSASTSSNNNGRLSVETKLLIREVLEES--GSLEYTRGVIRGLYDETR
CH063_0097     QLHSILQSARAARSASASSNNDGKLSLETKLLVREMLEEA--GSLEYTRVVIRGLYNETR
CH063_0347     QVEGMFR-AMR---------RGDADPLESKRYIVSLLHRS--GSLEATRELLLEWQ-RGI
CSUB_TS_16     KVEGVFRAARQMG-------KEKHDGIQWKRYILGLLEES--GALNATRQVLVEWHKEMM
CH063_0282     QLSSILK--------------QRTTDVDVKLFAVAYIEST--GSFEHCRKTLAELMAQAK
CH063_1125     QLINILR--------------QKTNDIEVKRFAVSYMEST--GSFEYTKQVVAVLVDRAR
GLRG_03419     QLINILA--------------QKTNNVEVKRFAVARMEST--GSFEYTKQVVGVLIDRAR
CSUB_TS_5      QLINILA--------------QKTNDVEVKRFAVARMEST--GSFEYTKQVVGVLIDRAR
CH063_0190     LIMGIFR-QQTANAAKGGSAESARLPHETKRYMLGLLKSS--DAMALTLKKLQELEAAME
CH063_1346     SLLSTGK-------------KTGTLTREQKECILEILEKS--GGLAYTRVVLLRLHDQAS
GLRG_00455     RVFKRRD-------------DGGEMGLEMKVFIIREMRNT--GALDKTYDLLRSMQSDLM
CH063_0519     EIMGIMK-----------HKKPGPMSMEMKQFILTEMRKS--GALETTHSLLQTMQEDLI
CSUB_TS_1      EIMGIMK-----------HKKPGPMPFEMKQFIVAEMKKS--GALEATHKLLKEMQENLI
GLRG_07956     GAVARLE-------------REVAAVNEASEAAERRFADT--PHKHHVSTLSRNCKNMCM
CSUB_TS_9      GAVARLE-------------REIAAVNEAAEAAEKRFAST--PHKDNISLLARNCKNMCM
CH063_0939     DLLAELA--------------TSVTVFEEAAGILEQSAGK--GGQELMKTYCDACRCMVT
GLRG_03977     ALLAELA----------------TSVAAFEAAAIGLEEAADKGERELLKTYCDACRCMVT
CSUB_TS_6      ALLAELV----------------TSVTAFEAAAIDLEEAVGKEGKELMKTYCDACRCMVT
GLRG_02961     PVLVELE-------------VIKNEVVEAANRAAVILE----GRGQYEQSMHDSVRGLLA
CH063_1495     PVLTELE-------------EVKNEVVEAANRASTILK----GRGRYEQSLHESVRGLLA
CSUB_TS_3      PVLTELE-------------VLKDEVVESANRAAKILK----GRGKYEQSMHDSVRGLLA
GLRG_11695     ------------------------------------------------------------
CH063_0640     ---------------------------KAGERCISILVVT--------------------
CSUB_TS_13     SPLAVLE-------------STISETALAYNQTLSILKDS---------RSYNELWYAYA
CSUB_TS_14     TPLGVLE-------------STISETTLAYNRTLAILKDK---------HPYSELWYAYA
GLRG_08962     IVGSLLE-------------ESYIEWDEVISRMPCWDVDVDSEVERYIKGIQNVVQANI-
CSUB_TS_10     MVGSLLE-------------ESYHEWHEVMSRLPDWDVNVDGEVERYIKGIQDVVQA---
CH063_1011     SIVDDIE-------------KAIRHFDACERRLITKSRED----ADTTRKIAATLKTICT
GLRG_06676     EVVDDIE-------------KAIRQFDICERRLVNQSQAD----SDTTRQIAATLKTIFS
CSUB_TS_7      EVFDDIE-------------KAVRQFDVCERRLINQSQTD----GDTTRQIAATLKTICT


CH063_0494     RFDSLWKTDSVY---WPGTQP--GSA-PIDHKEL----------------------
GLRG_07002     RFDSLWKTDGVY---WPSTEP--GSAQPIAHKEL----------------------
```

```
CSUB_TS_8      RFDTLWKTDGVY---WPSTEP--GSAQPIDHKEL----------------------
CH063_0492     V-KASWKLPW-------------RLWRATSKYQI----------------------
GLRG_10556     V-KASWKLPW-------------RLWRATSKNQI----------------------
CSUB_TS_12     V-KASWKLPW-------------RLWRATSKNQI----------------------
CH063_01_0     GKHDAEYHGIDE-----------EDANPHTLGVLVCEME-----------------
CH063_0916     G-NEAW-----------------SEMTPRYHSV-----------------------
CSUB_TS_17     SGNELW-----------------SRTTPRYLVVD----------------------
CSUB_TS_15     D-LMSSGVFGFL-----------KAPLAVIDGIL----ILWRASVVFLPVMWRELN
CH063_0378     V-AVESYMEIGRVLMRNGG----AAADGSGRATVPKSRRIWVAWSTLMAS------
GLRG_02475     V-AVESYMEIGRVLMRNGGS---AAADGSGRATVPKSRRIWVAWSTLMAA------
CSUB_TS_2      V-AVESYMEIGRVLIRNSG----AAADGSGRATVPKSRRVWVAWSTLMAA------
CH063_0666     K-VMIDNIDESE-----------GLKKGIFEAFL---AKIYKRSK-----------
GLRG_03139     K-EMINNVDESE-----------GLKKGIFEAFL---AKIYKRSK-----------
CSUB_TS_4      K-DMINNVDESE-----------GLKKGIFEAFL---AKIYKRSK-----------
CH063_01_3     --AAISEFPDS------------EAKDGLVEMAV---KALKRTK------------
GLRG_09287     --AAISGFPDS------------EAKDGLVEMAV---KALKRTK------------
CSUB_TS_11     --AAISGFPDS------------EAKDGLVEMAV---KALKRTK------------
GLRG_10302     --AVLNALENESGP---------GGKNWMLRLLT---FQLKI--------------
CH063_0097     --AMLTEIENEAGS---------GGKNWMLRLLT---FQLKV--------------
CH063_0347     N-EEIERLEGDF-----------GAPNPTLRLLM---ESLRIDV------------
CSUB_TS_16     --AEIGELEREF-----------GAGNGMLRLLV---EMLRI--------------
CH063_0282     --AIIEGMEGDS-----------SESLSVMNQIL---TMLGLDGNEVPR-------
CH063_1125     --KLADKIDDGR-----------GLTKGAHAILD---KMALVDRTTS---------
GLRG_03419     --KAVQDIDEGR-----------GLTKGMYKILD---KMALADSATS---------
CSUB_TS_5      --KAAEDIDEGR-----------GLTKGVHKILD---KMALSD-------------
CH063_0190     --EKISDLEERF-----------GETNPVMRILL---SRLSVRDVTL---------
CH063_1346     --AALKKLELSF-----------GSSNPEMKVLL---ELLRI--------------
GLRG_00455     --DELKRLEKAF-----------GAPNASLELVL---RKLWID-------------
CH063_0519     --TELRGLERDF-----------GSKNPMLEMVL---RKLWIS-------------
CSUB_TS_1      --TELRRLEREF-----------GSKNPTLELVL---RKLWIN-------------
GLRG_07956     A-NWLWSMRTPRYCLWDIKPDAAGGYNWTVDLEV----------------------
CSUB_TS_9      S-NWLWSMKTPRYCLWDVKREADGGYNWIVDAEP---ENH----------------
CH063_0939     G-SIQFTLESSR---YKLEGC--LNEDGSLDILL----------------------
GLRG_03977     G-SILFSLESSRYKLEDC-----LNEDGSLDILL----------------------
CSUB_TS_6      G-SIQFTFESSRYKLEGC-----LNEDGSLDILL----------------------
GLRG_02961     MHTANPRYKLKD---LGLAE---KHPLAPFEDQI---GELFERMRE----------
CH063_1495     MHTANPRYRLGD---LGLAE---EHPLGPFEDQI---GELFDRMKA----------
CSUB_TS_3      MHTANPRYKLKDLGLAE------EHPLAPFEDQI---GELFERMKAK---------
GLRG_11695     --------------------------------------------------------
CH063_0640     -----------------------TLQK-----------------------------
CSUB_TS_13     MGYVAMHLR--------------TRRYCLSELGI----------------------
CSUB_TS_14     MGYVAMHLN--------------SRRYYLSELGV----------------------
GLRG_08962     --GWSFHLKKDL-----------GADGPQVRKTW---RIDVLVNPPYLSNPGEADY
CSUB_TS_10     --NLSWRYF--------------GADGPKVRKTW---RIDVLANPPYLSDSGGAEY
CH063_1011     G-NLTWSLACKRY----------NVGKPAADGSL---RQELRAGE-----------
GLRG_06676     ---------------H-------PVGTVSANQLL---MEV----------------
CSUB_TS_7      G-NLTWSLASKR---Y-------SVGEPASDGSL---RQELRVGKQSD--------
```

PART 4: KS domains of PKS and Hybrids sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment


GLRG_10537      -EPIAIVGFACRLPGG-N-YSPPQFWEFLKRGEIAYNKV-PESRFNH----AAHYD----
CS_23           --PIAVVGMACRLPGH-S-NSPTALWEFLQRGGVAKNEP-PPSRFNL----DGHCD----
CH063_1191      -EPIAIIGVGMRLPGR-X-HNAADYWDLLVNGKSGRCRV-PESRYNV----DNWYG----
CS_16           -EPIAIIGMGMRLPGH-I-QNATDYWDLLVNGKSGRCPV-PKSRYNV----ENWYG----
GLRG_06162      -EPIAIIGMGMRLPGH-V-HNATDYWDLLVKGNSGRCPV-PKSRYNV----NNWYG----
GLRG_09075      ---VAVCGLGLRAPGG-I-RNATDYWDLLVNGRDARGPI-PDTRYNI----DGFND----
CH063_1187      -----------RLPQG-V-RDTNAFWEVLVNGKDLQSHV-PESKYNH----QGFAN----
CH063_1411      ---VAICSIGLRLPQG-V-RDTNAFWEVLVNGKDLQSHV-PESKYNH----QGFAN----
CS_5            -DPVVICGMAMRLPGG-V-SNSEQLWELLLQKRCASVPI-PGSRFNL----EGYYS----
GLRG_11893      -DPVVICGMAMRLPGG-V-SNSEQLWELLLQKRCASIPI-PGSRFNV----EGYYS----
CS_19           -DPIAIVGMAVRLPGG-V-HTTSDFWDMLINKKTGLTKI-PKERWDS----EGFYS----
CH063_0663      -DPIAIVGMAVRLPGG-V-RNTEDFWNLMMDKRSGLIPI-PKERWNS----EGFYS----
GLRG_00918      -EPIAVCGMAARLPGD-V-QTPAEFWDMLISKRNGLVDWPPSGRFDA----AGFQS----
CH063_0821      ------------------------------------------------------------
CS_13           --PIAICGIGVRLPRD-I-CTPEEMYEFLLSGLDARSPP-DESRYNA----RAFHY----
CH063_0053      -EPIAICGIGLRLPGG-I-DTPTSLYNFLVNRKDARSKP-DKPRYSS----QTHHF----
CH063_1006      AEPIAVCGMALRLPGG-I-STPEQFWQFLVDKRDARGPI-PETRFSA----ASYYS----
GLRG_07171      -EPIAVCGMALRLPGG-I-STPEEFWQFLIDKKDARGLI-PQTRFNA----SSYYS----
CS_22           TEPIAVCGMALRLPGG-I-STPEEFWQFLVDKKDARGPI-PQSRFNA----SSYYS----
CS_14           --PIAICGMGLRLPGG-S-STPQEFWNFLVNKGDARGRV-PTSRYNV----AAYHE----
CS_26           --PIAICGMACRLPGG-L-HTPQQLWDFLLAKGDARSRV-PESRYNV----DAFYD----
GLRG_11435      --PIAICGMACRLPGG-L-HTPQQLWDFLLAKGDARSRV-PESRYNV----DAFYD----
GLRG_08640      -EPIAINGMGCRLPGQ-V-DSESSFWQTLVEKRTGQTPKAPESRFDI----DAHYH----
GLRG_10827      -EPIAVVGMGCRLPGD-V-SSASDFWDMLMNKRSGQNPKVPANRFNV----EAHYH----
GLRG_11425      ------------------------------------------------------------
CH063_0013      --PIAVIGMGCRLPGD-V-SSPSDFWKLMMEKRSGQTPKVPSSRFNI----DAHFH----
GLRG_08482      -EPVAIVGMGCRWPGG-V-HDSSQFWEFLCNKTNGWKEF-DEPRFST----SGFYH----
CH063_1130      ------------------------------------------------------------
GLRG_03511      ---IAIIGMACRLPGG-A-DNPESLWSMLAEGRDGRKEI-PKDRWDW----KSFYH----
CS_18           --PIAIIGIGCRLPGG-S-DNPNKLWDLLSEGRSGWREI-PADRWNK----DSFYH----
GLRG_05714      --PIAVIGMGCRLPGG-S-DNPDKLWDMLSEGRSGWREI-PADRWNK----DSFYH----
CH063_0922      -MPIAIIGMGWRGPGE-A-TNVQNFYELLAAAREARVAA-HKGKWNH----EAFYH----
CH063_1282      ------------------------------------------------------------
CS_28           DEDIAIIGMACRFPGD-A-TSPSKLWDLLMDGRSAWSEV-PASRWNQ----DAHYH----
GLRG_11563      --------MACRFPGD-A-TSPSKLWDLLMEGKSAWSEV-PASRWNQ----DAHYH----
C.P.NRPS_6      ---IAIIGVAHRFPGG-A-NTPAKLWEVLNSRRDLSRKP-DPDRLNL----EKFYH----
CH063_0617      ---IAIIGLAHRFPGG-A-NTPGKLWEVLNSRKDLSREP-GVDRLRL----DKFYS----
C.P.NRPS_1      ---IAVIGIGCRFPGA-S-SFT-ELGDVLSNPPDLSQTI-PSDRFNI----DGYHH----
GLRG_09842      ---IAVIGIGCRFPGA-S-SFP-ELADALHEPADLSQTI-PSDRFSI----DGFHH----
GLRG_09715      ---VAIVGTSCRFPGA-A-HSPSKLWQILRDPPDLRKKI-PSTKFNQ----YSFHH----
C.P.NRPS_8      ---IAIIGSACRFPGE-S-VTPSKLWELLRDPRDVGGPL-PASRFNA----EGFHH----
GLRG_10367      ---IAIIGSACRFPGE-S-VTPSKLWELLRDPRDVGGPL-PVSRFNA----EGFHH----
CS/1            ---IAIIGSACRFAGG-V-DSPATLWELLSNPRDVRTEI-PSSRFSA----DGFYH----
CH063_0489      ------------------------------------------------------------
CH063_0392      ---VAIVGSACRFPGG-A-SSPSKLWDLLRQPRDVLRTI-PEDRFNP----EGFYH----
GLRG_07434      ---IAIVGSGCRFPGG-A-SSPSKLWELLREPRDVLCEI-PPSRFDV----NGFHH----
CS/20           ---IAVIGSGCRFPGG-C-TTPSRLWDLLENPTDIVSEI-TPDRFNV----DRFFH----
CH063_0306      ---IAVVGSGCRFPGG-S-TTPSKLWDLLENPNDIVSEI-TPDRFNV----DRYYH----
CH030B          ---IAIIGTGCRFPGG-S-DSPSKLWELLKDPRDLLKEI-PEDRFSV----EGFYH----
CS_30           -EPIAIVGTGCRFPGS-C-NSPSKLWELLEKPRDLLKEI-PDDRFNA----DGFYH----
CS_29           -EPIAIIGSACRFPGG-V-DSPSKFWECLQQPRDLRSEV-PKQRFSA----EKFFN----
CH063_0199      ---------GCRFPGA-A-NTPSKLWDLLKEPRDVQSRI-PKERFDV----DTFYH----
C.P.NRPS_2      ---IAIIGRACRFPGENS-TSPSGLWDLLLAPRDLTKPIPPESRFNP----DGFYH----
CH063_0288      -EPIAIVGTGCRFAGD-V-TSPSKLWDLLTNPRDLSQEV-PSSRFHA----QGFYH----
CS_4            -EPIAIVGTGCRFAGQ-V-NSPSKLWDLLTNPRDLTQEV-PKNRFNA----QGFYH----
GLRG_08212      -EPIAIVGTGCRFAGQ-V-NSPSKLWELLINPRDLTQEV-PKNRFNA----QGFYH----
```

```
CS_15          MEDIAIIGMACRFPGN-A-TSPEKLWETMVQKESAWSEF-PKDRLNI----DGYYH----
CH063_1445     -EDIAIIGMACRFPGN-A-TSPEKLWEMMVQKESAWSEF-PKNRLNI----DGYYH----
GLRG_02583     -EDIAIIGMACRFPGN-A-TSPEKLWQMMVQKESAWSEF-PKDRLNI----DGYYH----
CS_32          -MPIAVVGMACRLPGS-V-SSPAEFWELCSRARSGFTPV-PKERFNH----EAFYH----
GLRG_03360     -MPIAIVGMACRLPGS-V-SSPAEFWELCSRARTGFTPV-PKERFNH----EAFYH----
CH063_1222     -MPIAIVGMACRLPGS-V-SSPAEFWELCSRARTGFTPV-PKERFNH----EAFYH----
CH063_0134     ---VAVIGMSCKFSGE-A-TTPEKLWQLVVDGRDGWTTI-PKSRFNA----DAFYD----
CS_21          ------------------------------------------------------------
GLRG_09267     -EPIAIIGMSSKFSGD-A-TDNEKLWEMLAGGRSGWTPF-PSSRFRL----QGIYH----
CH0190         --PVAIVGMGCRFSGD-S-TSPQKLWELLEQGRSTWSKI-PTSRFNV----GGVFH----
CH063_0873     -MPIAIVGMGCRFAGN-V-TGPERLWELLKSGQSGWSEI-PESRFAT----SGILH----
CS_9           -EPIAVIGLDAKLPCD-G-DTVEKFFQFLVDGRSARKPV-PSDRYNA----EAFWH----
CH063_1490     -EPIAIVGLSARLPGD-G-DTPERFFESLLAGRSARTEV-PQDRYNA----DAFWH----
GLRG_01860     -EPIAIVGVSARLPGD-G-DTPERFYDSLLAGRSARTEV-PRERYNA----DAFWH----
CS_25          -EPIAIVGLSARLPGD-G-DTPERFFDSLLAGRSARTEV-PRERYNA----NAFWH----
CS_34          DEPVAIIGMACRFPGE-A-TNPSKFWELLKDGRDAFSEA---DRFNP----EAFYH----
GLRG_11770     --PIAVVGLACRLPDE-A-SSPEEFWKLLLEQRDTLSV--PHNRWNT----DAFHH----
CH0399         ------------------------------------------------------------
CS_11          --PIAIVGLSCRFPGD-A-TDPSKFWDMLKSGRDGYTP--KTNRYNE----DAFHH----
GLRG_11840     --PIAVVGLSCRFPGD-A-TDPSKFWDMLKNGRDGYTP--QTNRYNE----DAFHH----
CH063_0781     --PIAVVGLACRLPGD-A-TCPTEYWNLLKEGRDAYST--DTDRYNT----SAFYH----
CH063_1357     --PIAIIGIGLRAPGD-A-SDPEKFWQMLQDARSARSEI-PKDRYNV----DGFYH----
GLRG_10317     --PIAIIGIGVRAPGD-A-SDPEKFWQMLLDARSARSEI-PKDRYNV----DGFYH----
CH063_1471     ---IAIVGIAGRFPGD-A-ENPQKLWDMLAEGRSALSDV-PGDRFNV----DAFYH----
CH063_0506     -EPLAIIGLSVKFPQD-A-TSPEEFWEMLREGRSAASKV-PEDRFNV----DAFYH----
GLRG_11835     --PIAIIGMSMKLPGD-I-KSSQALWDILKNAQCLTGKI---------------------
CS_3           -EPIAIIGLNLKFPGN-A-TSVQAFWDMLENGLNASCEV-PASRFNI----DAFYH----
GLRG_11878     -EPIAIIGLNLKFPGE-A-TSMQAFWAMLENGLNASCEV-PAGRFNI----DAFYH----
CH063_0051     ------------------------------------------------------------
CH0135         -EPIAIIGAACRLPGE-A-TSLGGLWDMISHGRSAHGKV-PPSRWDA----DTWYH----
CS_43          -DDIAVIGMACHVAGA---QDLEQYWKLMLEGKSQHKELVPNDRFAM----ETSYR----
GLRG_09268     -NDIAVIGMACHVAGA---QDLEQYWKIMLEGKSQHRELVPNDRFAM----ETSYR----
GLRG_11857     -NDIAVVGMSAKVAGA---DSLEAFWDVLCAGVSQHSEV-SPDRFGGGFGMDKSMF----
CH063_0213     --DIAVVGMACKVAGS---DGLEEFWDLLCNGEPQHQEL-KDQRFSF----QTPFR----
CH063_0477     ---IAIVGASCRLPGA---NSLDELWEVISQGQTRLEKL-RTDRFDL----GGSHR----
CH063_0313     ---VAIVGMAGRFPGA---NSVDELWELMSAGRSTVTS--PPERVGL----NQLAE----
CS_44          ---IAVVGMAGRFPGA---NSVDELWELLLDAQSMVDR--APDRVGL-------------
GLRG_11575     ---VAIVGIAGRFPGA---NSVDELWELLLNAKSMVDK--APTRVGL----DQL------
CH063_1347     ---IAVVGMACRFPRS---ENLEAFWRLLADGQTGLGTM-PADRFDP----ATLTR----
C.PKS          ---IAVVGMGCKFPGA---DSVEKFWSMLDSGESKLSSP-PSGRFPA--------H----
CS_45          DDSIAIVGMACRLPGA---DNLDEFWQLLLSGSSMLGRL-PGERFAT-------------
GLRG_11798     -DSIAIVGMACRLPGA---DNLEEFWQLLLSGSSMLRRL-PGERFAT----TGLRR----
GLRG_11778     --RVAIVGMGGRYPRS---GDLEEFWQHICQGKALHTEI-PKDRFSL----DDYYS----
CH0398         ---IAIVGMGGRFPAC---DDLEGLWDIISEGQIVHSEV-PRDRFNV----DELYD----
CS_37          --RIAIVGVSGRGPMC---DNISEFWDIIMSKTDTCTEV-PKDRFDI----DEFYC----
GLRG_11836     --RIAVVGVSGRGPMC---GNINEFWDIIMSKTDTCTEV-PKDRFDI----DEFYS----
CH063_1255     ---IAVVGMAGRAPEC---DDLEQYWQVIETGRDVAREI-PPGRFDA----DDLFQPQPG
CS_40          --RIAVVGMGGRGPDSHANDSLDQFWNLIRQGQDVARTI-PEDRFDA----DSLFP---P
CS_38          -----VVGMSCRLPGG-A-NDTELFWKLMMDKRDVHTTV-PPDRFDL----STHFD----
CS_36          DSKLAIVGMACRMPGD-A-DTPERFWELLAQGRDTHTVV-PPDRFDL----GAHFD----
GLRG_08632     ----AIVGMSCRLPGG-A-NDCKLFWDLLKEGRDTHTEV-PADRFDL----QAHFD----
CH063_0351     ----AIVGMAGRFPDA---ASHEKLWELLSKGLDVHRVV-PADRFPV----ATHYD----
CS_35          ----AIVGMAGRFPDA---ASHEKLWELLSKGLDVHRVV-PADRFPV----ATHYD----
GLRG_04203     ----AIVGMAGRFPDA---ASHEKLWELLSKGLDVHRVV-PPDRFPV----ATHYD----
GLRG_11956     ----AIIGMSGRFPG--A-KHNEAFWDVLRQGLDLHKPAPA-LHWDV----KTHVD----
CH063_1223     ----AILSASGRFPGN-A-DTMDQFWDILYNGIDTHEMV-PPSRWDA----STHVG----
GLRG_08620     --PIAILAASGRFPGR-A-DTMDAFWDILIRGVDTHEMV-PSSRWSS----SAHVG----
CH063_0250     KSPIAILAVSGRFPGN-A-DTMDAFWDILHNGVDTHEMV-PASRWNA----SAHVG----
CS_39          --DIAIVGYSCRLPGA---ETGEDFWRILAEGRDVHGRV----SFSQ----QVLHN----
GLRG_03507     ---IAIVGMSGRFPGG---DNLQAFWDILVQGKDLHKKV-PKDRFDV----DLHCD----
GLRG_03509     ---IAIVGMSGRFPGG---DNLQAFWDILVQGKDLHKKV-PKDRFDV----DLHCD----
```

```
GLRG_10537   ----GSL--KPRTMRQP--GGMFLKD-ID--LAD-FDAKFFELSATE-------------
CS_23        ----KSR--RPRTMKSP--GGMFIED-VD--PEL-FDGQFFNISRVD-------------
CH063_1191   -----PG--RALHVPTE--FGYFLEE-LN--LAH-VDASFWSFTKQE-------------
CS_16        -----PG--RVSHVPTK--FGYFLEE-LN--LAH-VDPSFWSFTKQE-------------
GLRG_06162   -----PG--RVSHVPTK--FGYFLEE-LN--LAH-VDPSFWSFTKQE-------------
GLRG_09075   ----SLG--GKNTIKTR--FGYFLTD--D--LTR-IDTSLFSMSRKE-------------
CH063_1187   ----TLG--KKGAIETQ--CGYFVNK--D--LGS-LDTSLFSMSQKE-------------
CH063_1411   ----TLG--KKGAIETQ--CGYFVNK--D--LGS-LDTSLFSMSQKE-------------
CS_5         ----RHP--RPGTMHFE--KAYILDH-VD--LDR-FDSCHFPFSRKE-------------
GLRG_11893   ----RHP--RPGTMHFE--KAYILDH-VD--LDR-FDTCHFPFSRKE-------------
CS_19        ----ETP--KWGTIQNR--DCYLLSD-VD--LKK-FDTSFFTCGQAE-------------
CH063_0663   ---PVA---KWGTVQNK--EAYMFSQ-ADNDLTK-FDASYFTCGEKE-------------
GLRG_00918   ----RLP--AKNTLQAP--QAYLLNH-VS--LSD-FDTSFFPVGAKE-------------
CH063_0821   ------------------------------------------------------------
CS_13        -QASGGN--EAGAEKPLPSQGYWLNW-KE--ISS-FDASLFNLPRKE-------------
CH063_0053   ---TESG--TTRPLPTE--EGYWLTH-DD--VTK-FDPSMFSMNPKE-------------
CH063_1006   ----EAA--KPGHIKTQ--HGYFLDDSVD--LAA-LDTTFFRMPKSE-------------
GLRG_07171   ----EAA--KPGHIKTQ--YGYFLDESVD--LAA-LDTTFFRMPKSE-------------
CS_22        ----EAA--KPGHIKTQ--HGYFLDDSVD--LAA-LDTTFFRMPKSE-------------
CS_14        ----TTK--RPTTVAAE--YGYFLSE--DVKLGA-MDTTRFSMSRAE-------------
CS_26        ----TSG--RPGTVKTE--YGYFLDY--D--ISK-TDGGFSSFVRGA-------------
GLRG_11435   ----TSG--RPGTVKTE--YGYFLDY--D--ISK-SDGGFSSFVRGA-------------
GLRG_08640   ---ERLD--RPGFFNVP--GGHFLDG--H--PEN-FDPTFFNTTPVK-------------
GLRG_10827   ---RNKE--RPGSLPVM--GGYFINE--D--LEN-FDPGLFNITPVE-------------
GLRG_11425   --------------------------------------SFFGITPIE-------------
CH063_0013   ---SDND--RPGSFGVL--GGYFLNE-----TLQEFDPAFFGITPVE-------------
GLRG_08482   ---PNPD--RPGSISMK--GAFLAKE--D--ARL-FDHAFFGMTGLE-------------
CH063_1130   ------------------------------------------------------------
GLRG_03511   ---KDTN--AKEAINFS--HGYFLDQ--E--ISA-FDARFFGVPGAE-------------
CS_18        ---PDPE--AKEAVNFK--GLYFLDQ--D--IAA-FDARFFNIHPHE-------------
GLRG_05714   ---PDPE--AKEAVNFK--GLYFLDQ--D--IAA-FDARFFNIHPHE-------------
CH063_0922   ---PESS--RKGTSNVE--AGHYFKD--D--LSL-FDAPFFSMTDSE-------------
CH063_1282   ----------------------------D--IHA-FDAGFFNVRRDE-------------
CS_28        ---PSQE--RAGSNTVK--GGHFLRD--GEQNGKQFDAAFFNITKRE-------------
GLRG_11563   ---PSQE--RAGSNTVK--GGHFLRD--DEQNGKKFDAAFFNITHKE-------------
C.P.NRPS_6   ---TNGE--YHGKSNVT--KSYFLEE--D--PRL-FDTAFFNVSPLE-------------
CH063_0617   ---ENGE--HHGKSNVT--KSYFLDE--D--PRL-FDTTFFNISPLE-------------
C.P.NRPS_1   ---DDGS--HHGSTNVR--KSYFLRD--N--VRK-FDNNFFNVTAVE-------------
GLRG_09842   ---EDGK--YHGNTNVR--KSYFLRD--G--IRK-FDNNFFNITAVE-------------
GLRG_09715   ---PESV--RHGTSNVT--HSYVLEE--D--VGE-FDYGFFNLNPKE-------------
C.P.NRPS_8   ---PDPL--HHGHSNVKDMRAYFLSG--EGVEHR-FDAGFFGIRSAE-------------
GLRG_10367   ---PDPL--HHGHSNVKDMRAYFLSG--EGVEHR-FDAGFFGIRSAE-------------
CS/1         ---VDSQ--YHGHSNVR--HAYVLEQ--D--PGV-FDAQFFGVKPVE-------------
CH063_0489   ------------------------------------------------------------
CH063_0392   ---PDGT--HHGTSNVT--QSYLLEE--D--PRV-FDSGFFNIKPVE-------------
GLRG_07434   ---TDSQ--FPGRTNVK--HSYLLQE--D--VAK-FDAQFFNITASE-------------
CS/20        ---PDHK--HHGTSNVR--HSYMLQE--N--FKL-FDAKFFGVKPQE-------------
CH063_0306   ---PDHK--HHGTSNVR--HSYLLQE--N--FKL-FDAKFFGVKPQE-------------
CH030B       ---PKNS--HHGTTNVR--HSYLLDE--D--LRR-FDASFFGIKATE-------------
CS_30        ---PKSS--HHGTSNVR--HSYLLDE--D--LRT-FDASFFGIKAIE-------------
CS_29        ---SNGM--HHGSTNSR--YGYFLKE--P--LGA-FDAPFFNIHAGE-------------
CH063_0199   ---PDGT--HHGRTNAS--YAYFLKE--D--LHA-FDAPFFNIQAGE-------------
C.P.NRPS_2   ---PNSE--HHGSSNIK--ESYFIEE--D--PRL-FDAGFFNIAPRE-------------
CH063_0288   ---ADGE--YHGATNAT--QGYFLEQ--D--HRV-WDAGFFNITPKE-------------
CS_4         ---PDGE--YHGATNAT--QGYFLEQ--D--HRV-WDAGFFNITPKE-------------
GLRG_08212   ---PDGE--YHGATNTT--KGYFLEQ--D--HRV-WDAGFFSITPKE-------------
CS_15        ---PSGD--RQGSISFR--GAHFLKD--D--IAA-FDASFFSVMAED-------------
CH063_1445   ---PSGD--RQGSISFR--GAHFIKD--D--IAA-FDASFFSVMAED-------------
GLRG_02583   ---PSGD--RQGSISFR--GAHFIKD--D--IAA-FDASFFSVMAED-------------
CS_32        ---PNPG--KTGAYHAE--GGNFLNV--D--LAS-FDAPFFGLTEKE-------------
GLRG_03360   ---PNPG--KTGAYHAQ--GGNFLDV--D--LAS-FDAPFFGLTEKE-------------
```

```
CH063_1222    ---PNPG--KTGAYHAE--GGNFLDV--D--LAA-FDAPFFGLTEKE-------------
CH063_0134    -----------------------------------TDHAKIGMQQIT-------------
CS_21         ----EED------------------------VGL-FDAAFFGYSSEA-------------
GLRG_09267    ---PNNE--RLNSTHVK--GAHFLQE--D--IGL-FDAAFFGYSSEA-------------
CH0190        ---PDGQ--RVGSTHVK--GGHFLDE--D--PAL-FDASFFNMTSEV-------------
CH063_0873    ---PQKD--KIGSTNVR--GGYFLGT--D--PGA-FDASFFNLSTEV-------------
CS_9          ---PDNH--RDGVIGGK--EGHFMEA--N--VKA-FDAPFFAMTPAE-------------
CH063_1490    ---PDAN--RSGATRVR--HGHFLKG--S--ISA-FDAPFFSITPTE-------------
GLRG_01860    ---PNAD--RSGATRVR--HGHFLKG--S--ISA-FDAPFFSITPTE-------------
CS_25         ---PDAD--RSGTTRVR--HGHFLKG--S--ISA-FDAPFFSITPTE-------------
CS_34         ----ANSSGRQNVLPCK--GAHFLQG--S--PYA-FDAAFFNITPGE-------------
GLRG_11770    ---PGRN--KPQTIAAR--SAHFLKR--D--VMA-FDAAFFNINASE-------------
CH0399        -----------------------------------FDAAFFNIGHAE-------------
CS_11         ---PGGNNKRQNVLPVK--GGYMLKQ--D--PYV-WDAAFFNITAAE-------------
GLRG_11840    ---PGGNNKRQNVLPVK--GGYMLKQ--D--PYV-WDAAFFNITAAE-------------
CH063_0781    ----ANPKDRQNVLPTK--GGHFIKE--D--PYA-FDTSFFKLTAAE-------------
CH063_1357    ---PDPE--RLGSIQQR--HAHFLKQ--D--FKV-FDAPFFSVTPKE-------------
GLRG_10317    ---PDPE--RLGSIQQR--HAHFLKQ--D--FKV-FDAPFFSVTPKE-------------
CH063_1471    ---PHNE--RQGTINVR--KAHFMNR--D--ISA-FDAPFFNMPIAE-------------
CH063_0506    ---PDPN--RLDSLRVR--DAHFMKE--D--PRA-FDAPFFNMSPAE-------------
GLRG_11835    --------------RVQ--NAHFMNE--D--PRA-FDAPFFSMSPAE-------------
CS_3          ---PDPA--RLDSIRVR--KASFMAE--D--PRA-FDAAFFNMSPAE-------------
GLRG_11878    ---PDPA--RLDSIRVR--KAHFMAE--D--PRA-FDAAFFNMSPAE-------------
CH063_0051    ------------------------------------------------------------
CH0135        ---PDPD--RKGSISVK--HGYFLEQ--D--VSL-FDAPFFSMTAKE-------------
CS_43         -----PG--QKDDERKW--YGNFIDD-----HDA-FDYKFFKKSPRE-------------
GLRG_09268    ---------PSEKEGTRKWYGNFLDD-----HDS-FDYKFFKKSPRE-------------
GLRG_11857    ----RTA--APNRSRKW--FANFLSG-----PDE-FDHRFFGKSPRE-------------
CH063_0213    ---------PADAKRKW--FANLVDG-----ADH-FDHKFFKRSPRE-------------
CH063_0477    ----AAQ--DRDWVTKREFYGNFIDD-----VRG-FDNAFFGISPRE-------------
CH063_0313    ---------DNSQVRWR---GNFIED-----HDA-FDHKFFNKSARE-------------
CS_44         ----DQL--VEDVSQVKW-WGNFLAD-----HDS-FDHKFFNKSARE-------------
GLRG_11575    ---------DEDVSQVKW-WGNFLDD-----YDS-FDHKFFNKSSRE-------------
CH063_1347    ---------EPRLANF---YGNFIDA-----PDA-FDHRFFGISGRE-------------
C.PKS         ----EHQ--RDTEKTVY--FGNYLDD-----IAS-FDNRFFKKSGRE-------------
CS_45         ----TGL--RRSPKGDKPFLGNFVRD-----AFA-FDHKFFNKSSRE-------------
GLRG_11798    ---------SPKEDTPF--LGNFVRD-----AFA-FDHKFFNKSSRE-------------
GLRG_11778    ----ASS--SPFTTNTK--YGCFLDN-----PGL-FDARFFSLSPRE-------------
CH0398        ---PKTK--LHLSTSTK--YGCFIRQ-----PGL-FDARFFQVSPRE-------------
CS_37         TEHGRGE--KKCTMTTK--YGCFMNN-----PGH-FDSRFFHISPRE-------------
GLRG_11836    --AEHGRGEKKCTTTTR--YGCFMDK-----PGH-FDSRFFHISPRE-------------
CH063_1255    HGENSSCSRATCVSTCK--LGCFLDN-----PGH-YDARFFRVSPRE-------------
CS_40         HAASSSSSSPCRSSTD--YGCFLRD-----PGH-FDARFFRLSPRE-------------
CS_38         ---PTGQ--TENATQTP--FMNYMEN-----PGH-FDAGFFNISPKEVWQPPHFLPSTYF
CS_36         ---PTGE--TENAIGTQ--FGNFISN-----PGH-FDAGFFNMSPRE-------------
GLRG_08632    ---PTGQ--VPNSTPTP--WGNFIDH-----PGL-FDAAFFNMSPRE-------------
CH063_0351    ---ITGK--AVNTSHSQ--YGCWIEN-----PGY-FDPRFFNMSPRE-------------
CS_35         ---ITGK--AINTSHSQ--YGCWIEN-----PGY-FDPRFFNMSPRE-------------
GLRG_04203    ---ITGK--AINTSHSQ--YGCWIEN-----PGY-FDPRFFNMSPRE-------------
GLRG_11956    ---PTGK--IKNTSATP--FGCWLDD-----PGA-FDAKFFNISPRE-------------
CH063_1223    ----DTS--IKNVSGTG--YGCWLHQ-----AGQ-FDTTFFNISPRE-------------
GLRG_08620    ----DTS--KKNVSGVG--FGCWLHQ-----ASQ-FDITFFNMSPRE-------------
CH063_0250    ----DPS--VKNVSGTG--FGCWLHQ-----APL-FDAAFFNMSPRE-------------
CS_39         ----MPESRCAGFGGNPP-LGCFYQG-----LGL-FDCRLFNMSPRE-------------
GLRG_03507    ---PTGK--TPNSTLSA--FGCFLDK-----PGH-FDNLMFNMSPRE-------------
GLRG_03509    ---PTGK--TPNSTLSA--FGCFLDK-----PGH-FDNLMFNMSPRE-------------


GLRG_10537    ---------------ATAMDPAQRQMLEVVYEGLEN------------------------
CS_23         ---------------CIAMDPQQRQMLEVAYECLEN------------------------
CH063_1191    ---------------AELMDPRQRLFLEVAYEALES------------------------
CS_16         ---------------AELMDPRQRLFLEVSYEALEN------------------------
```

```
GLRG_06162     ---------------AELMDPRQRLFLEVAYEALEN------------------------
GLRG_09075     ---------------VERCDPQQRLLLEVAREALED------------------------
CH063_1187     ---------------LERCNPQQRQFLEVVRQALEN------------------------
CH063_1411     ---------------LERCNPQQRQFLEVVRQALEN------------------------
CS_5           ---------------VEQMDPAQRMLLEITWECLEN------------------------
GLRG_11893     ---------------VEQMDPAQRLLLEITWECLEN------------------------
CS_19          ---------------AERMDPMQRQLLEITRECLDS------------------------
CH063_0663     ---------------IERMDPMQRQLLEIARECLDN------------------------
GLRG_00918     ---------------ASRMDPMQRQLLEVAYECAEN------------------------
CH063_0821     ------------------------------------------------------------
CS_13          ---------------IDRLDPQQRVLLRVVWEALES------------------------
CH063_0053     ---------------LSKLDPQQRLLLQVVWEALES------------------------
CH063_1006     ---------------VERADPQQRLLLELTRECLES------------------------
GLRG_07171     ---------------VERADPQQRLLLELTRECLES------------------------
CS_22          ---------------VERADPQQRLLLELARECLES------------------------
CS_14          ---------------VEFADPQQRRLLEVVMEAFED------------------------
CS_26          ---------------TEREDPHQRQMTEVARECIED------------------------
GLRG_11435     ---------------TEREDPHQRQMTEVARECIED------------------------
GLRG_08640     ---------------AQWLDSQQRRTLEMCYESLVS------------------------
GLRG_10827     ---------------AMWMDPQQRKLCEVVYEAFES------------------------
GLRG_11425     ---------------AMWMDPQQRKLLEVDYDSFES------------------------
CH063_0013     ---------------ATWMDPQQRKLLEVVYEAFES------------------------
GLRG_08482     ---------------VETMDPSQRKLLEVSYEAIEN------------------------
CH063_1130     ------------------MDPQQRKLLQVVYECLQS------------------------
GLRG_03511     ---------------APGIDPQQRLLLEVTYEAVEN------------------------
CS_18          ---------------AHCLDPQQRILLETTYEALEN------------------------
GLRG_05714     ---------------AHCLDPQQRILLETTYEALEN------------------------
CH063_0922     ---------------AACLDPQQRLLLETCYEALEN------------------------
CH063_1282     ---------------AKAMDPQQRITAECAFEALES------------------------
CS_28          ---------------TETMDLQQRIVMENVYEAMES------------------------
GLRG_11563     ---------------TETMDLQQRIVMENVYEAIES------------------------
C.P.NRPS_6     ---------------AEAMDPQQRILLETVYEATES------------------------
CH063_0617     ---------------AEAMDPQQRLLLETVYEATES------------------------
C.P.NRPS_1     ---------------ASSIDPQQRIVLETVHEAFEN------------------------
GLRG_09842     ---------------ASAMDPQQRLLLETVYEAFEN------------------------
GLRG_09715     ---------------AESMDPQQRLLLETVYEGIEA------------------------
C.P.NRPS_8     ---------------AAVLDPQVRLLMETTYEALEA------------------------
GLRG_10367     ---------------AAVLDPQVRLLMETTYEALEA------------------------
CS/1           ---------------AKALDPQQRLLLEVVYEGLEA------------------------
CH063_0489     ------------------MDPQHRMLLEVVYEAMEH------------------------
CH063_0392     ---------------AHATDPQHRILLETVYEALES------------------------
GLRG_07434     ---------------AVAMDPQQRLLLETVYEGLEG------------------------
CS/20          ---------------AIAMDPQQRLLLETVYESLEA------------------------
CH063_0306     ---------------AIAMDPQQRLLLETVYESLEA------------------------
CH030B         ---------------ANSIDPQQRMLMETVYEGLES------------------------
CS_30          ---------------ASSIDPQQRLLLETVYEALES------------------------
CS_29          ---------------AKSMDPQQGLLLEATYDALSS------------------------
CH063_0199     ---------------AESMDPQQRLLLETVYEAVSN------------------------
C.P.NRPS_2     ---------------AEAIDPQQRLLLETTYEAMEN------------------------
CH063_0288     ---------------AEAIDPQQRIILEVVYEALES------------------------
CS_4           ---------------AEAIDPQQRIILEVVYEALES------------------------
GLRG_08212     ---------------AEAIDPQQRIILEVVYEALES------------------------
CS_15          ---------------AKAIDPQQRFLLEVSYEALEN------------GKRTPFLLGKQF
CH063_1445     ---------------AKAIDPQQRFLLEVSYEALEN------------------------
GLRG_02583     ---------------AKAIDPQQRFLLEVSYEALEN------------------------
CS_32          ---------------AISMDPQQRLLLECTFEALEN------------------------
GLRG_03360     ---------------AISMDPQQRLLLECTFEALEN------------------------
CH063_1222     ---------------AISMDPQQRLLLECTFEALEN------------------------
CH063_0134     ---------------ANRTDCFV----------------------------------------
CS_21          ---------------AASLDPQYRLQLESVYEALES------------------------
GLRG_09267     ---------------AASLDPQYRLQLESVYEALES------------------------
CH0190         ---------------ARDMDPHHRLMLEVVYEALES------------------------
```

```
CH063_0873   ---------------ASTLDPQIRMTLEVVFEAMES------------------------
CS_9         ---------------AANLDPQQRMLLECVYTAMEN------------------------
CH063_1490   ---------------ARSIDPQQRGMLESVYKALEN------------------------
GLRG_01860   ---------------ARSIDPQQRGMLESVYKALEN------------------------
CS_25        ---------------ARSIDPQQRGMLESVYKALENGQGSRIHAPTWGGRTCKTSEKKKK
CS_34        ---------------ALATDPRQRLALEVAYEALEN------------------------
GLRG_11770   ---------------AVALDPQQRFVLEVTYEALES------------------------
CH0399       ---------------ATALDPQQRFTLELTFEALES------------------------
CS_11        ---------------AIAFDPKQRIAMEVTYEALEN------------------------
GLRG_11840   ---------------AIAFDPKQRIAMEVTYEALEN------------------------
CH063_0781   ---------------AAVLDPKQRILLEVTYEALEN------------------------
CH063_1357   ---------------AKAMDPTQRMLLEAAYEGFEN------------------------
GLRG_10317   ---------------AKAMDPTQRMLLEAAYEGFEN------------------------
CH063_1471   ---------------AKAMDPQQRMALECTYEALEN------------------------
CH063_0506   ---------------ASVLDPQQRGLLEGAYHCLEN------------------------
GLRG_11835   ---------------ACILDPQQRGLLEGAFHTFEN------------------------
CS_3         ---------------ASILDPQQRGLLEGAYHTFEN------------------------
GLRG_11878   ---------------ASILDPQQRGLLEGAYHTFEN------------------------
CH063_0051   ------------------MDPMQRLGLEIAYEAFEN------------------------
CH0135       ---------------AAGMDPMKRLLLEVSYEGFES------------------------
CS_43        ---------------ALHMDPQQRLMLQVAYQAVAQ------------------------
GLRG_09268   ---------------ALHMDPQQRLMLQVAYQAVAQ------------------------
GLRG_11857   ---------------SAAMDPQQRHMLQVAYQAVEQ------------------------
CH063_0213   ---------------SATMDPQQRWMLQIAYQAVEQ------------------------
CH063_0477   ---------------AAYMDPQQRILLETAFEAMDS------------------------
CH063_0313   ---------------ATACDPQQRKLLEVVYEALES------------------------
CS_44        ---------------ALACDPQQRKLLEVVYEALES------------------------
GLRG_11575   ---------------AMACDPQQRKLFEVVYEALES------------------------
CH063_1347   ---------------AKSMDPQQRLALQVAYEALES------------------------
C.PKS        ---------------AVSLDPQQRLLLEVSYQALES------------------------
CS_45        ---------------ARSMDPQHKLILQVAYEALES------------------------
GLRG_11798   ---------------ARSMDPQHKLILQVAYEALES------------------------
GLRG_11778   ---------------AMEVDPAHRLFLLAVFEALES------------------------
CH0398       ---------------ALQMDPAHRLFLMSTYEALEM------------------------
CS_37        ---------------AMLMDPCHRQFLMSTYEALEM------------------------
GLRG_11836   ---------------AMLMDPCHRHFLMSTYEALEM------------------------
CH063_1255   ---------------ALLMDPGSRLFQMACHEALES------------------------
CS_40        ---------------ATLLDPCSRLFLMAAHAAVEM------------------------
CS_38        PPLFQYLYKDLPLGGAEQMDPMHRLALVTAYEALEM------------------------
CS_36        ---------------AQQTDPMQRLALVTAYEALEM------------------------
GLRG_08632   ---------------AEQTDPMHRLALVTAYEALEM------------------------
CH063_0351   ---------------AFQTDPM-RMALTTAYEALEM------------------------
CS_35        ---------------AFQTDPMQRMALTTAYEALEM------------------------
GLRG_04203   ---------------AFQTDPMQRMALTTAYEALEM------------------------
GLRG_11956   ---------------AAQIDPAQRLALMTAYEAIEQ------------------------
CH063_1223   ---------------ASRIDPAQRLALMTAAEALEK------------------------
GLRG_08620   ---------------ASQVDPAQRLALLTTAEALEQ------------------------
CH063_0250   ---------------AAQVDPAQRLALLTAAEALEK------------------------
CS_39        ---------------AAQTDPGQRLLLMATYEALQL------------------------
GLRG_03507   ---------------AAQTDPCQRLLLMAAYEALET------------------------
GLRG_03509   ---------------AAQTDPCQRLLLMAAYEALET------------------------


GLRG_10537   ----AGI---PLDKV-----NGQSVGCYVGNYASD--Y----GDMG--------ARDPED
CS_23        ----AGV---PLELL-----SGTKTGVIVGTNFID--Y----AAIQ--------NRDPED
CH063_1191   ----SGS-----TSW-----RGSDVGVYVGTMGED--W----SLLE--------CHDQQC
CS_16        ----SGS-----KSW-----KGNNVGVYVGTMGDD--W----ATLE--------SRDERN
GLRG_06162   ----SGS-----KSW-----KGNDVGVYVGTMGED--W----TTLE--------SRDERN
GLRG_09075   ----AGE-----TKY-----RGERIGCYVGTFGDD--W----AQIA--------GKETQH
CH063_1187   ----AGE-----VNY-----RGKAVGCYVGTFSED--W----LQMA--------AREQQH
CH063_1411   ----AGE-----VNY-----RGKAVGCYVGTFSED--W----LQMA--------AREQQH
CS_5         ----AGE-----MNW-----QGKAVGCYVGSFAED--W----TTEM--------ERDTQY
```

```
GLRG_11893     ----AGE-----VNW-----QGKAVGCYVGSFAED--W----TTEM--------ERDTQY
CS_19          ----AGE---VGTST-----MDKRVGCYVGTFSSD--W----QDDQ--------SMDPQA
CH063_0663     ----AGE-----TNW-----RGEEIGCYVGNFSSD--W----QDDL--------SMDPHA
GLRG_00918     -AGAGGL---AHDSA-----TRKDVGVFVGVFGED--W----LQEN--------VMDSQL
CH063_0821     ------------------------------------------------------------
CS_13          ----AGE-----ADW-----RGKNIGCYVGSFGDD--W----RELH--------ARDELD
CH063_0053     ----AAE-----TDW-----QGQRIGCYVGSFGDD--W----REMH--------AVDTQD
CH063_1006     ----AGE-----TEY-----RGKTIGTFVGCFGED--W----LETL--------TRDSEV
GLRG_07171     ----AGE-----TDY-----RGKTIGTFVGCFGED--W----LETL--------TKDSEV
CS_22          ----AGE-----TAY-----RGKTVGTFVGCFGED--W----LETL--------TRDSEV
CS_14          ----AGE-----AQF-----RGKKIGCYFGNMNED--W----GEMM--------NRDPLW
CS_26          ----AGE-----TNY-----RGGLIGCYVGSFGED--W----CEMF--------AKDVQQ
GLRG_11435     ----AGE-----TNY-----RGGLIGCYVGSFGED--W----CEMF--------AKDVQQ
GLRG_08640     ----AGI---TLERI-----SGSNTAVFVGSIMAD--YHTNVEQMS--------TWNTDF
GLRG_10827     ----GGI---SLDAI-----AGSRTGVFAASFTAD--W----QQMS--------MKEPSF
GLRG_11425     ----AGV----------------TLELYSRLPTDD--------------------VQEPSF
CH063_0013     ----AGL---TLDQL-----SGSDTACFMATFTAD--F----QQMS--------FKEPSF
GLRG_08482     ----AGE---TWDSV-----SGSRTGVFVGNFCLD--H----WMIQ--------SRDWDN
CH063_1130     ----SGT---TMESL-----SGSSTGVFVANFSVD--F----QPMQ--------ARDPDY
GLRG_03511     ----AGI---PIEDL-----RGSDTSVHMAMFARD--Y----DRMG--------YRDTSQ
CS_18          ----AGQ---TIAGI-----KGSDTSVHVGAYATDCSF----ERMG--------YKDTAR
GLRG_05714     ----AGQ---TIAGI-----KGSDTSVHVGAYATD--F----ERMG--------YKDTAR
CH063_0922     ------------GNA-----AGSDTSVFVSTFGAD--Y----TDLL--------QKDPDT
CH063_1282     ----AGW---TLRDV-----AGRNVAVFAAHQGST--Y----AGHA--------AEDLLT
CS_28          ----AGL---QLDDV-----KGSKTSVFAAVFTDD--V----RSIL--------QEDPDL
GLRG_11563     ----AGL---RLEDV-----KGSKTSVFAAQFTDD--V----RSIL--------QEDPDL
C.P.NRPS_6     ----SGT---PIDKF-----QGSRCSVFVGVMTGD--Y----ETIQ--------YRDTED
CH063_0617     ----AGV---SVDRL-----RGSRCSVFVGVMTGD--Y----ETIQ--------YRDTED
C.P.NRPS_1     ----AHV---SMSQL-----KGSDTAVYVGLMCGD--Y----EYIL--------NNDLDT
GLRG_09842     ----ADV---STPQL-----KGSDTAVYVGLMCGD--Y----ENIL--------LRDIDS
GLRG_09715     ----AGY---AMSDL-----RGSPTGVYVGQMTDD--Y----YDIL--------NKDVQC
C.P.NRPS_8     ----AGQ---TMEAL-----RGSDTGCYVGMMISE--Y----EQSM--------MRDPES
GLRG_10367     ----AGQ---TMEAL-----RGSDTGCYVGMMISE--Y----EQSM--------MRDPES
CS/1           ----AGI---PAEDL-----RGSNTGVYVGLMSND--Y----ESLM--------LRDVDT
CH063_0489     ----GGF---TLSSL-----QNSDTAVYVGMMCTD--Y----NITL--------GLDSTF
CH063_0392     ----AGL---PMEGM-----AGSQTAVYVGQMWCD--Y----NDHL--------QRDMDS
GLRG_07434     ----AGL---TIEGL-----KGSNTGVYVGVMYVD--Y----ESLQ--------FRDLQH
CS/20          ----AGI---PIEDI-----KGSKTGVFVGNMGGD--Y----SELL--------GQDLDA
CH063_0306     ----AGI---TIQHL-----KGSKAGVFVGNMGGD--Y----SELL--------GQDLDA
CH030B         ----AGL---SIKQL-----QGSDTAVYVGVMSAD--Y----TEML--------ARDVEK
CS_30          ----AGL---SVKQL-----QGSDTAVYVGVMSAD--F----TDML--------ARDIEK
CS_29          ----AGL----VQDL-----QGSDTAVYVGLMTHD--F----ETIK--------TQDMHH
CH063_0199     ----AGM---RMQDL-----QGSSTAVYVGMMTHD--Y----ETTS--------TRDLES
C.P.NRPS_2     ----AGL---SLQGM-----KGSSTSVYVGIMSAD--Y----TETQ--------LRDPES
CH063_0288     ----AGY---NIEKY-----AGRKVAVFAGLMTAD--Y----DTLC--------QRDDIT
CS_4           ----AGY---NLERY-----AGRKVAVFAGLMTAD--Y----DSLC--------QRDDIT
GLRG_08212     ----AGY---NLERY-----AGRKVAVFAGLMTAD--Y----DTLC--------QRDDIT
CS_15          VLTLAGL---RMEDL-----RGSPTAVYVGSFVKG--Q----SAPCCILITPAETRQTTS
CH063_1445     ----AGL---RMEDL-----RGSPTAVYVGSFVKD--Y----EQVC--------LRDMDW
GLRG_02583     ----AGL---RMEDL-----RGSPTAVYVGSFVKD--Y----EQIC--------LRDMDW
CS_32          ----AGI---PKHTI-----VGKDVGVFVGGSFAE--Y----ESHL--------FRDSDT
GLRG_03360     ----AGI---PKHTI-----VGKDVGVFVGGSFAE--Y----ESHL--------FRDSDT
CH063_1222     ----AGI---PKHTI-----IGKDVGVFVGGSFAE--Y----ESHL--------FRDSDT
CH063_0134     ----AGV---PLQKL-----AGSDTSVFVGSFSRD--Y----HDSL--------LRDPDT
CS_21          ----AGI---PMSAI-----AGSGTSVFAGVFVHD--Y----RDGL--------LRDADN
GLRG_09267     ----AGL---PISTI-----AGSNASVFTGVFVHD--Y----RDGL--------LRDADN
CH0190         ----AGI---PLESA-----AGTNTAVFGGIMYKD--Y----HDSL--------NRDPKS
CH063_0873     ----AGI---KLPDI-----QGSDTSVFAGCMVRD--Y----HDTL--------ARDPHT
CS_9           ----AGY---TMADM-----DGSLTGVYAGAFMWD--F----RDLL--------IKDVDV
CH063_1490     ----AGI---PLDTV-----AGSRTGVYVGCFTAD--Y----NDHI--------TKDLDI
GLRG_01860     ----AGI---PLDTV-----AGSQTGVYVGCFTAD--Y----NDHI--------AKDLDI
CS_25          KKKRAGI---PLDTV-----AGSRTGVYVGCFTAD--Y----NDHI--------SKDLDI
```

```
CS_34         ----SGL---PPHRV-----AGTRTACFMGAAANMAEY----KDGI--------VRDFGN
GLRG_11770    ----AGL---TMAGV-----SGSRTGCYVGFSSCD--Y----RDSI--------LRDTET
CH0399        ----AGL---PIDVV-----AGTKTGCFMGSAAAD--Y----RDTI--------NRDPDS
CS_11         ----AGM---SLQKV-----AGTQTACYMGTSMSD--Y----RDSI--------VRDFGN
GLRG_11840    ----AGM---SLQKV-----AGTQTACYMGTSMSD--Y----RDSI--------VRDFGN
CH063_0781    ----AGL---PLPRV-----AGTRTACIIGTAWSD--Y----RDAL--------VRDFEQ
CH063_1357    ----AGL---RLEDV-----SGTQTSCYIGTFTND--F----VNLQ--------AQSNEA
GLRG_10317    ----AGL---RLEDV-----SGTQTSCYIGTFTND--F----VNLQ--------AQSNEA
CH063_1471    ----AGI---AMENV-----DGSNTSCFVGCFTRD--Y----SDML--------ACDRED
CH063_0506    ----AGI---SIPNV-----TGSNTSVFVACFGRD--Y----DASI--------ARDVES
GLRG_11835    ----AGY---PIDKI-----AGTNTSVFCASFGRD--N----EAIV--------SRDPEF
CS_3          ----AGI---AMECA-----AGSKTSVFCASFGRD--S----DAIV--------ARDPAF
GLRG_11878    ----AGI---AMECA-----AGSKTSVFCASFGRD--S----DAIV--------ARDPAF
CH063_0051    ----AGI---PMQKL-----AKSTTAVYSGAMTND--Y----QMLA--------EADPYR
CH0135        ----AGV---LMDDL-----MNSQTGVYVGCMTND--Y----EQLS--------THDIYD
CS_43         ----SGY---YHTSQ-----PDSRVGCYIGVVGSD--Y----------------ENNISH
GLRG_09268    ----SGY---YHKPQ-----PDSRVGCYIGVVGSD--Y------EN--------NISHTS
GLRG_11857    ----AGY---FNSKQ---DETNGRVGCFVGCSTCD--Y------EQ--------NIACHD
CH063_0213    ----SGY---FNAVS-----PERSVGCYIGLRGND--Y----------------YDNVAC
CH063_0477    ----SGY---LRHHR---RERGDPVGCFIGASYTE--Y------LE--------NTSAYT
CH063_0313    ----SGY-HSADSRV-----DPTDYGCYIGAVMNN--Y----------------ASNVSC
CS_44         ----AGY---LGADA---RSDCTDYGCYIGAVTNN--Y------VT--------NVSCHP
GLRG_11575    ----SGQ---LGADA---LSDCSDYGCYIGAVMNN--Y----------------VTNVSC
CH063_1347    ----AGYAGGAGTDGDGIKSTGPNVGCYLGVGSVD--Y------DA--------NVAARD
C.PKS         ----AGF---FGPRN-----PDVDVGCFIGVCASD--Y------ND--------NVASHP
CS_45         ----AGY---FSRGG-----QPRDVGCYIGVAASD--Y----EDNV--------ASHAPT
GLRG_11798    ----AGY---FSRGG-----QPRDVGCYIGVAASD--Y----ED----------NVASHA
GLRG_11778    ----AGY---SRRPH--TKTAACRFATFVGQIADD--W----------------AEITRL
CH0398        ----AGY---GRGRF--ASPHDTRIATFFGQASED--W----RDIM----------HISG
CS_37         ----AGYSNGKTKST-----DPNRIATFYGQVTDD--W----HDQ---------SHPTLG
GLRG_11836    ----AGYSDGKTKST-----DPNRIATFYGQVTDD--W----HDQ---------SHPTLG
CH063_1255    ----AGYSCGTTRAV-----DPSRISVLFGQSNVD-GY----ET----------AHHEKG
CS_40         ----SGYSDGPTRTV-----DPARIAVVYGQSNED-------GYMT--------SHHERG
CS_38         ----SGY---VPNRT--RSTAGPRIGTYYGQASDD--W----RE----------LNASQN
CS_36         ----AGF---VPNRT--PSSHTTRVGTYYGQASDD--Y----RE----------VNASQK
GLRG_08632    ----SGY---VPNRT--PSTILERVGTFYGQASDD--W----RE----------LNAAQN
CH063_0351    ----CGY---VPNRT--PSTRLDRIGTFYGQTSDD--W----RE----------INAAQE
CS_35         ----CGY---VPNRT--PSTKLDRIGTFYGQTSDD--W----RE----------INAAQE
GLRG_04203    ----CGY---VPNRT--PSTKLDRIGTFYGQTSDD--W----RE----------INAAQE
GLRG_11956    ----AGI---IPDAT--PSTRRDRVGVFYGVTSND--W----MET---------NS-AQN
CH063_1223    ----AGI---VPDRT--PSTMRHRVGVWYGSCSMD--W----LE----------TNSAQD
GLRG_08620    ----AGI---VPDRT--PSTQKHRVGVWFGSTSND--W----ME----------TNSAQS
CH063_0250    ----AGI---VPNRT--ASTQKHRVGVWFGATSND--W----ME----------TNSAQD
CS_39         ----AGY---SPGTT--SSTDGRRVGTFFGQTTDD--W----RE----------YNVRDD
GLRG_03507    ----AGY---RYDAK----PDRGNVGSFVGLTTDD--W----RE----------YNISQE
GLRG_03509    ----AGY---RYDAK----PDRGNVGSFVGLTTDD--W----RE----------YNISQE


GLRG_10537    RS--PGH----GLGIARTILANRISHFLNIHGP----------------------S----
CS_23         RA--DSI----TIGLASSILSNRVSHFLNVTGP----------------------S----
CH063_1191    LN--QVR----PDVYGDYILANRASYEFDLTGP----------------------S----
CS_16         LN--PVR----PDVFGDYILANRASYEFDLTGP----------------------S----
GLRG_06162    LN--AIR----PDVFGDYILANRASYEFDLTGP----------------------S----
GLRG_09075    QGGLGYV----ATGNGDLMLSNRVSYEYDLRGP----------------------S----
CH063_1187    AG--GYL----LTGHGDLMIANRASYEYDLKGP----------------------S----
CH063_1411    AG--GYL----LTGHGDLMIANRASYEYDLKGP----------------------S----
CS_5          FD--HY-----PAGKWDFMLSNRISYVYDFRGP----------------------S----
GLRG_11893    FD--HY-----PAGKWDFMLSNRISYVYDFRGP----------------------S----
CS_19         SG--VYR----GSGYLDFFQANRISYEYNWTGP----------------------S----
CH063_0663    SG--LYR----GSGYLDFLQPNRVSYEYGWTGP----------------------S----
GLRG_00918    AG--LYR----GTGFLDFLQANRVSFAMGWQGP----------------------S----
CH063_0821    ------------------------------------------------------------
```

```
CS_13           PG--GYR----LTGYMDYALSNRISSVLDLRGP----------------------S----
CH063_0053      DG--MYR----LTGYMDFVQANRISHAFDLRGP----------------------S----
CH063_1006      VG--QYK----ITGYGDFMLSNRLAYEYDLKGP----------------------S----
GLRG_07171      VG--QYK----ITGYGDFMLSNRLAYEYDLKGP----------------------S----
CS_22           VG--QYK----ITGYGDFMLSNRLAYEYDLKGP----------------------S----
CS_14           HG--PNK----IDGYQDWMLANRISYEFGLTGP----------------------S----
CS_26           HG--LYR----VAGYGDFMLSNNVAYEMDLKGPRYVAHPRSFLLVHISQTNQDLSS----
GLRG_11435      HG--LYR----VAGYGDFMLSNNVAYEMDLKGP----------------------S----
GLRG_08640      CH--NCA----ATGVDIRIISNRIGHMFNLRGP----------------------S----
GLRG_10827      RH--TMA----ATGVDPGILSDRISYVFNLNGP----------------------S----
GLRG_11425      RH--ALA----ATGVDPGIISNRISHAFNLKGT----------------------S----
CH063_0013      RH--SLA----ATGVDPGLLSNRVSHIFNLRGP----------------------S----
GLRG_08482      PR--PYA----FTGAGTSILANRISYIFNLQGP----------------------S----
CH063_1130      IH--RYQ----ASGSGATVMSNRISHVLNLQGP----------------------R----
GLRG_03511      MH--KQH----STGSGDAILANRISYLLDLKGT----------------------S----
CS_18           TP--KAH----MIGTGIAILSNRISYVFDLHGP----------------------S----
GLRG_05714      TP--KAH----MIGTGIAILSNRISYVFDLHGP----------------------S----
CH063_0922      MP--VYQ--ATNSGYSRAIISNRLSHFFDFRGS----------------------S----
CH063_1282      TS--AYS----ASGTAGCMLANRISYLFDLRGP----------------------S----
CS_28           SV--KYK----PIGTSAAILAARVSWFYDLRGS----------------------S----
GLRG_11563      SV--KYK----PIGTSAAILAARVSWFYDLRGA----------------------S----
C.P.NRPS_6      LS--QYA----ASGTSRAILANRISYFFDLSGS----------------------S----
CH063_0617      LS--QYT----ASGTSRAILANRISYFFDLAGS----------------------S----
C.P.NRPS_1      AP--RYQ----ATGIARSIISNRISYIWDLHGP----------------------S----
GLRG_09842      AP--RYQ----ATGVGRSIMANRISYIWDLHGP----------------------S----
GLRG_09715      AP--QYT----ATGSSRAIIANRVSYFFDWRGP----------------------S----
C.P.NRPS_8      TG--TYH----ILGTARSLMSNRLSYFFDWHGP----------------------S----
GLRG_10367      TG--MYH----ILGTARSLMSNRLSYFFDWHGP----------------------S----
CS/1            MA--NYH----AVGTQRSILANRISYFYDWHGP----------------------S----
CH063_0489      IP--QYT----ATGVSPSNASSRISYYFNWHGP----------------------S----
CH063_0392      LP--TYA----GVGTARSILSNRISHFFDWHGP----------------------S----
GLRG_07434      VP--TYL----AIGTARSIVSNRISYFFDWHGP----------------------S----
CS/20           FP--TYF----APGTARSILSNRISYFFDLHGP----------------------S----
CH063_0306      FP--TYF----APGTARSILSNRISYFFDLHGP----------------------S----
CH030B          YP--TYF----ATGTARSILSNRLSYAFDWRGP----------------------S----
CS_30           FP--TYF----ATGTARSILSNRLSYFFDWHGP----------------------S----
CS_29           MP--TYF----ATGVAASIASNRLSYVFDWHGP----------------------SRLIL
CH063_0199      IP--TYS----ATGVAVSIASNRISYFFDWHGP----------------------S----
C.P.NRPS_2      VS--QYW----VTGSSRALTSNRLSYFFDWRGP----------------------S----
CH063_0288      SN--QYF----ATGNSRAILSNRISYFFNFRGP----------------------S----
CS_4            AN--QYF----ATGNSRAILSNRISYFFNFRGP----------------------S----
GLRG_08212      SN--QYF----ATGNSRAILSNRISYFFNFRGP----------------------S----
CS_15           RF--AYAIWTGSPSTGNAIMSNRVSYTYDFKGP----------------------S----
CH063_1445      QP--QYA----ATGTGNAIMSNRVSYTYDFKGP----------------------S----
GLRG_02583      QP--QYA----ATGTGNAIMSNRVSYTYDFKGP----------------------S----
CS_32           IP--MHQ----ATGCAHAMQSNRLSHFFDLRGP----------------------S----
GLRG_03360      IP--MHQ----ATGCAHAMQSNRLSHFFDLRGP----------------------S----
CH063_1222      IP--MHQ----ATGCAHAMQSNRLSHFFDLRGP----------------------S----
CH063_0134      LP--RTT----LTGNGVAMMSNRISHFFDLRGP----------------------S----
CS_21           LP--RLM----ATGTGVPMMSNRVSHFFDLRGA----------------------S----
GLRG_09267      LP--RLM----ATGTGVPMMSNRVSHFFDLRGA----------------------S----
CH0190          LP--RYF----ITGNAGAMVANRISHFFDLRGP----------------------S----
CH063_0873      LP--RYF----MTGNAATMAANRVSHFYDLRGP----------------------S----
CS_9            PM--TYT----ACGTIASTLAGRVSWFYGLRGP----------------------A----
CH063_1490      PN--KYS----ALGTVASMLSNRVSWFFDFRGP----------------------S----
GLRG_01860      PN--KYS----ALGTVASMLSNRVSWFFDFRGP----------------------S----
CS_25           PN--KYS----ALGTVASMLSNRISWFFDFRGP----------------------S----
CS_34           QP--QHL----IMGISEELMSNRLSHFFDLHGP----------------------S----
GLRG_11770      SP--RYT----DLGTHAEMLSNRTSWFYNLKGP----------------------S----
CH0399          NP--RYS----LIGVTTEMISNRTSWFYDLKGP----------------------S----
CS_11           YP--KYH----LLGTSDEMISNRISHFLDIHGP----------------------S----
GLRG_11840      YP--KYH----LLGTSDEMISNRISHFLDIHGP----------------------S----
```

```
CH063_0781     WP--RLY----LMGVSDEMVSNRLSHYFDLRGP----------------------S----
CH063_1357     PS--IYH----ATGLSSSLASNRLSWFYNLKGP----------------------S----
GLRG_10317     PS--IYH----ATGLSSSLASNRLSWFYNLKGP----------------------S----
CH063_1471     LP--LYH----GTGTGSAIMSNRISWFFNMKGP----------------------S----
CH063_0506     MS--RYH----ATGSGSSMLANRISHAFDLRGP----------------------S----
GLRG_11835     QS--RYQ----GTSSGSTMLSNRVSYFYDLHGP----------------------S----
CS_3           QS--RYQ----ATASGSSMLSNRISHFFDLRGP----------------------S----
GLRG_11878     QS--RYQ----ATASGSSMLSNRISHFFDLRGP----------------------S----
CH063_0051     LG--VNS----AAGTGKSMLSNRISWFFDLQGP----------------------S----
CH0135         IG--DVA----ASGMSEAMTANRVSWFFGLRGP----------------------S----
CS_43          TSPNAFS----AMGALRCYIAGKVSHNFGWTGP----------------------S----
GLRG_09268     PN--AFS----AMGALRCYIAGKVSHYFGWTGP----------------------S----
GLRG_11857     AS--AFS----ATGQLRAFIAGKVSHYFGWTGP----------------------G----
CH063_0213     HPPNAFT----ATGNLQAFVAGKISHQFDWTGP----------------------A----
CH063_0477     PS--AFT----ATGTIRAFLSGKISYHFGWTGP----------------------S----
CH063_0313     HPPTAYA----TMGTSRAYFSGAISHYFGWTGP----------------------A----
CS_44          PT--AYA----TTGTGRSFLSGAVSHHFGWTGP----------------------A----
GLRG_11575     HPPTAYA----TTGTGRSFLSGAVSHHFGWTGP----------------------A----
CH063_1347     AN--AFA----ATGTLRAFISGR-------------------------------------
C.PKS          AN--AFS----ALGTLRAFVPGRISHYFGLTGP----------------------S----
CS_45          ----AFS----VTGAVRAFVSGKISHFFGLSGP----------------------S----
GLRG_11798     PT--AFS----VTGAVRAFVSGKISHFFGLNGP----------------------S----
GLRG_11778     RGGDAFS----LTGIQRSFAAGRVNYHFKWAGP----------------------A----
CH0398         GD--AYS----LTGLQRSFGPGRVNYSMRWSGP----------------------T----
CS_37          CD--AYT----LPGVQRAFGPGRVAWQFNWEGP----------------------T----
GLRG_11836     CD--PYT----LPGVQRAFGPGRVAWQFNWEGP----------------------T----
CH063_1255     CD--AYT----LQALARPFAGGRVAFHYGWEGA----------------------T----
CS_40          CD--AYT----LQTVQRAFAPGRVAHHFGWEGQ----------------------T----
CS_38          IG--TYA----VPSGERGFANGRINYFFKFSGP----------------------S----
CS_36          IG--TYG----IPGTERAFGNGRINYFFNFQGP----------------------S----
GLRG_08632     IG--THA----VPGGERAFANGRINYFFKFGGP----------------------S----
CH063_0351     VD--TYY----ITGGVRAFGPGRINYHFGFSGP----------------------S----
CS_35          VD--TYY----ITGGVRAFGPGRINYHFGFSGP----------------------S----
GLRG_04203     VD--TYY----ITGGVRAFGPGRINYHFGFSGP----------------------S----
GLRG_11956     ID--TYF----IPGGNRAFIPGRINYCFKFSGP----------------------S----
CH063_1223     ID--AYF----IPGGARAFIAGRVNYFFKFTGP----------------------S----
GLRG_08620     ID--TYF----IPGGNRAFIPGRINYHFKFRGP----------------------S----
CH063_0250     ID--TYF----IPGGNRAFIPGRINYHFKFSGP----------------------S----
CS_39          ID--MFY----VPGTIRAFGPGHLNYYFKWDGP----------------------S----
GLRG_03507     ID--MYF----VTGGLRSFGSGRLNYFFKLEGP----------------------S----
GLRG_03509     ID--MYF----VTGGLRSFGSGRLNYFFKLEGP----------------------S----


GLRG_10537     -----VTLDTA--------------SSSLVGLDLACQALRARSIDTGIVAAANVY-----
CS_23          -----MTIDTAC-------------SASLVSVDVACRYLDSFQADGMIVGGANMW-----
CH063_1191     -----LVVRTAC-------------SASMVALHQACRDLHSGDCSSALVGGANLI-----
CS_16          -----VVVRTAC-------------SASLVALHQACQDLHSGDCSSALVGGVNLI-----
GLRG_06162     -----VVVRTAC-------------SASLVALHQACQDLHSGDCSSALVGGVNLI-----
GLRG_09075     -----MTIKTGC-------------SASLAALHEAFRSIQNGDATGAIIGGTSLI-----
CH063_1187     -----MVIKTGC-------------SASLVGLHEACRALQNGDCVGAIVGGVNLI-----
CH063_1411     -----MVIKTGC-------------SASLVGLHEACRALQNGDCVGAIVGGVNLI-----
CS_5           -----MTIRTGC-------------SGTLVALNLAEQAVKSGECSSAIIAGCSMF-----
GLRG_11893     -----MTIRTGC-------------SGTLVALNLAEQAVKSGECSSAIIAGCSMF-----
CS_19          -----MLIKTGC-------------SSSMVALHLAAEAVQSGSCTSAMALGCNLI-----
CH063_0663     -----MLVKTGC-------------SSSMVALHLAAEAVQNGACKSAMALGCNLI-----
GLRG_00918     -----IVVKTGC-------------SASLVALDMACHSLRAGECSAAVVLGVNLI-----
CH063_0821     ------------------------------------------------------------
CS_13          -----MTVRAAC-------------AASGLALHLACQAIRAGECDSAIVAGSNII-----
CH063_0053     -----MTVRTAC-------------SAAGLAVHLACQAIRAGECDSAVVAGSNLM-----
CH063_1006     -----MTIRTGC-------------SSALIGLHEACMSISHGQCDAAVVAGSNLI-----
GLRG_07171     -----MTIRTGC-------------SSALIGLHEACMSIHHGQCNAALVAGSNLI-----
CS_22          -----MTIRTGC-------------SSALIGLHEACMSIHHGQCNAALVAGSNLI-----
```

```
CS_14          -----MTIRTAC-------------SSALTCVNEACAAIQRGVCDAAVVAGGNLM-----
CS_26          -----MTIRTGC-------------SAALVGLHEACTAIQRGDCTSAIVGGVNLI-----
GLRG_11435     -----MTLRTGC-------------SAALVGLHEACTAIQRGDCTSAIVGGVNLI-----
GLRG_08640     -----FTINTAC-------------SSSIYTINDACRALRMRDCDAIITAGFNLI-----
GLRG_10827     -----IVCNTAC-------------SSSIYALHNACIALRNDECDGAVVGGTNLV-----
GLRG_11425     -----IVVNIAC-------------SSSVYAVHNACNALRNKECQAAVFCGVNLI-----
CH063_0013     -----IVVNTAC-------------SSSVYALHNACNALRTHECSAAVVGGSNLI-----
GLRG_08482     -----LTVDTAC-------------SSSMYALHLAVNSIRAGDCDAAIVASANWI-----
CH063_1130     ---------------------------------------YAGDCESAIVASANLI-----
GLRG_03511     -----NTLDTGC-------------SGGLVAVHQACRTLKANEASIALAGASQLL-----
CS_18          -----STVDTGC-------------SGSMVALHQACHGLRARESKMAIVAGTQLV-----
GLRG_05714     -----STVDTGC-------------SGSMVALHQACHGLRARESKMAIVAGTQLV-----
CH063_0922     -----VTIDTAC-------------SGGLTALHLACQSIRAGEVRQAVCGGSSVI-----
CH063_1282     -----------------------------------AASVRLGECEAALVSAANLL-----
CS_28          -----FTLDTAC-------------SGSIVALHTGAQDLRAGLSEMSIITGVNII-----
GLRG_11563     -----FTLDTAC-------------SGSIVALHTGAQDLRAGLSDMSIITGVNII-----
C.P.NRPS_6     -----ICLDTAC-------------SSSLVAMHLAVQDLRNGTAETAIVAGANLI-----
CH063_0617     -----VCLDTAC-------------SSSLVAMHLAVQ-LRGGAADTAVVAGTNLI-----
C.P.NRPS_1     -----MTIDTAC-------------SSSLVALHQAVQALRLGETNLAVVAGSNLL-----
GLRG_09842     -----MTIDTAC-------------SSSLVALHQGVQALRLGETQLAVVAGSNLL-----
GLRG_09715     -----VNIDTAC-------------SSSLVALHQAVQALRGGEIDLAVAAGVNLI-----
C.P.NRPS_8     -----MTIDTAC-------------SSSLVAVHQAVQLLRSGESRVAIAAGSNMI-----
GLRG_10367     -----MTIDTAC-------------SSSLVAVHQAVQLLRSGESRVAIAAGSNMI-----
CS/1           -----MVVDTAC-------------SSSLVALHLAVQALRTGEARTALACGSNLL-----
CH063_0489     -----MTIDTAC-------------SSSLVAVHQAVDQLRTGRSRVAVACGTNLM-----
CH063_0392     -----ATIDTAC-------------SSSLVAVYNAVMTLRNGDAPVAVVAGANII-----
GLRG_07434     -----LTVDTAC-------------SSSLVAVHQAVQALRSGEVRVALAAGSNLL-----
CS/20          -----VTIDTAC-------------SSSLVAVHQAVQSLRLGETPVAIACGTNLL-----
CH063_0306     -----VTVDTAC-------------SSSLVAVHQAVQSLRVGETPVAIACGTNLL-----
CH030B         -----MTIDTAC-------------SSSLIAMHQAVQALRSGESKVAIAAGTNLL-----
CS_30          -----MTIDTAC-------------SSSLIALHQAVQVLRSHQSKVAVAAGSNLI-----
CS_29          TSHPQMTIDTAC-------------SSSLVAVHHAVQQLRSGQSRVAVAAGVNMI-----
CH063_0199     -----MTIDTAC-------------SSSLAAVHLAVQQLRSGQSTMAVAAGANLI-----
C.P.NRPS_2     -----MTIDTAC-------------SSSMAALHLAVQSLRNGECQVSCVAGANLL-----
CH063_0288     -----MTIDTAC-------------SSSLVALHQAVLSLRSGECEMACVAGANLM-----
CS_4           -----MTIDTAC-------------SSSLVALHQAVLSLRSGECEMACVAGANMM-----
GLRG_08212     -----MTVDTAC-------------SSSLVALHQAVLSLRSGECEMACVAGANMM-----
CS_15          -----MTIDTGC-------------SGSLVAVHLAAQALRAGDCSLALAAGAGLI-----
CH063_1445     -----MTIDTGC-------------SGSLVAVHLAAQALRAGDCSIALAAGAGLI-----
GLRG_02583     -----MTIDTGC-------------SGSLVAVHLAAQALRAGDCSLALAAGAGLI-----
CS_32          -----FTTDTAC-------------SASLVALHLACQSLRAGESTTAIVGGCHLN-----
GLRG_03360     -----FTADTAC-------------SASLVALHLACQSLRAGESTTAIVGGCHLN-----
CH063_1222     -----FTADTAC-------------SASLVALHLACQSLRAGESTSAIVGGCHLN-----
CH063_0134     -----LTTDTGC-------------SASLAALHLAVNSIRNGESRMAIVGASNLL-----
CS_21          -----MTIETAC-------------SSGMVALHQGIQSLRTGEADMSIIGGANLT-----
GLRG_09267     -----MTIETAC-------------SSGMVAMHQGIQSLRTGEADMSIVGGANLT-----
CH0190         -----VTIDTAC-------------STTLTALHLACQSLRAGESDMAVVVGANLM-----
CH063_0873     -----MTIDTGC-------------STTLTALHLACRSLQAGESGSAIVTGANLM-----
CS_9           -----LSVDTAC-------------SSSMVALHQAVVGLKSRDCNIAVACGTNVL-----
CH063_1490     -----VTVDTAC-------------SSSLVAVHEACMSLKLREISMV-------------
GLRG_01860     -----VTVDTAC-------------SSSLVAVHEACMSLKLREISMAVVGGCNLI-----
CS_25          -----VTVDTAC-------------SSSLVALHEACMSLKLREIPMAVVGGCNLI-----
CS_34          -----ATVETAC-------------SSSLVAVHLACQSLRAGESNMAIAGGVNLL-----
GLRG_11770     -----MTVSTAC-------------SSSLVAVHMACQSLLAGETDLAVAGGVNLM-----
CH0399         -----ITLTTAC-------------SSSLVAVHLACQSVLSGESTMAVAGGVNLM-----
CS_11          -----ATIETAC-------------SSSHVATHLACQSLQSGESEMAIAGGVGMT-----
GLRG_11840     -----ATIETAC-------------SSSHVATHLACQSLQSGESEMAIAGGVGMT-----
CH063_0781     -----LTVETAC-------------SSSLVAIHQACESLRSGQAEMAVAGGINMI-----
CH063_1357     -----ITIDTAC-------------SSSLTAFHLACQSIRSGESEMSIVAGANLM-----
GLRG_10317     -----VTVDTAC-------------SSSLTAFHLGCQSIRTGEAEMSVVAGANLM-----
CH063_1471     -----ISLDTAC-------------SSSMAALHLGCQSLRTGETTMSVVGGTNLM-----
CH063_0506     -----ITVDTAC-------------SAGLSAFHLACQSVRSGESDLSLVCGGNTY-----
```

```
GLRG_11835     -----ITVDTAC-------------SSGLYALHLGCQSVLSGESDMSLICGANVY-----
CS_3           -----LTVDTAC-------------SSGLYAVHLAYQSILAGESDMSLVCGSNTY-----
GLRG_11878     -----LTVDTAC-------------SSGLYAVHLACQSIMAGESNMSLVCGSNTY-----
CH063_0051     -----LTVDTAC-------------SSSLYAVHLACQSIRTGESDQALVTGHNLI-----
CH0135         -----LTLDTAC-------------SSSLYALHLACQSLKSGETKMGLVAGVNLI-----
CS_43          -----MTVDTAC-------------SASTIAMNLACQAIKSGDCSAALVGGTNFC-----
GLRG_09268     -----MTLDTAC-------------SASTIAMNLACQAITSGDCSAALVGGTNFC-----
GLRG_11857     -----VTIDTAC-------------SSSAVAVHQACQAILAGDCDAALAGGAFVM-----
CH063_0213     -----MTIDTAC-------------SSSLVAIHQACRAIIAGDCNAALAGGAHIM-----
CH063_0477     -----EVIDTAC-------------SASLVAVHRACRAIAAGECPVALAGGVNII-----
CH063_0313     -----TTIDTAC-------------SSSLVAIHTACRAIAAGECSRAVAGGTNVI-----
CS_44          -----LTIDTAC-------------SSSLVAIHMACRAIATGECSRAVAGGANII-----
GLRG_11575     -----LTVDTAC-------------SSSLVAIHMACRAIATGECSRAVAGGTNII-----
CH063_1347     ------------------------------------------------------------
C.PKS          -----VALDTAC-------------SSSAVAIDAACKAILQGDCKSALAGGVSIF-----
CS_45          -----LVFDTAC-------------SSSAVAIHTACQAIRSGDCSMALAGGVNVI-----
GLRG_11798     -----LVFDTAC-------------SSSAVAIHTACQAIRSGDCSMALAGGVNVI-----
GLRG_11778     -----VTLDTAC-------------SSSMTALDQACTFLLSGQCDMAVAGGTNIL-----
CH0398         -----YSVDSAC-------------ASSLSSVNLACSSLLARDCDMAVAGAANLI-----
CS_37          -----YALDSAC-------------ASTTSCVHLACMSLQTGDIDMAVAGAGNIL-----
GLRG_11836     -----YALDSAC-------------ASTTSCVHLACMSLQTGDIDMAVAGAGNIL-----
CH063_1255     -----YSVDQAC-------------STSLSLIHLACRDLLAGDHDMVVAGAANVL-----
CS_40          -----WSVDSAC-------------STSSSILHLATRLLASGEVDMVVAGASNVI-----
CS_38          -----FNMDTACLSLHSYLPPRSRGSSGLAAVNAACSALWAGEVDTALAGGLNVI-----
CS_36          -----FNVDTAC-------------SSGLAAVHMACSALWAGEADTVVAGGLNIIGRTDL
GLRG_08632     -----FNIDTAC-------------SSGLAAVNAACSALWAGEADTVIAGGLSVI-----
CH063_0351     -----LNIDTAC-------------SSSAAALNVACNSLWVKDCDTAIVGGLSCM-----
CS_35          -----LNIDTAC-------------SSSAAALNVACNALWVKDCDTAIVGGLSCM-----
GLRG_04203     -----LNIDTAC-------------SSSAAALNVACNALWVKDCDTAIVGGLSCM-----
GLRG_11956     -----YAVDTAC-------------SSSLAGIHLACNALWRGEIDTAIAGGTNVL-----
CH063_1223     -----FTVDTAC-------------SSSLAALHLACNALWRGEVDTAV------------
GLRG_08620     -----YTIDTAC-------------SSSLAALHLACNALWGGEVDTAIVGGTNIL-----
CH063_0250     -----YTIDTAC-------------SSSLAALHLACSALWRGEVDTAIVGGTNVL-----
CS_39          -----YSFDTAC-------------SSSSVAVQMACASLLARDCDLAVAGGLNTL-----
GLRG_03507     -----YVLDTAC-------------SSSAASIELACASLLGRDCDMALAGGANVM-----
GLRG_03509     -----YVLDTAC-------------SSSAASIELACASLLGRDCDMALAGGANVM-----


GLRG_10537     -----MSPEHLI-DAGGVGGAHSLTALCHTFDAAAD--GYVKAEAVNTVIIKRLSDA---
CS_23          -----LTPEHNE-EVGMMNITQSGSGRCHSFDAKAD--GHVKAEGINAVYLKRLDDA---
CH063_1191     -----LTPKDTT-VMHQNG-VLSPSGSCKSFDADAD--GFARGEGVSAIYIKKLSDAVRD
CS_16          -----LTPKETV-AMHQNG-VLSSSGFCKSFDADAD--GFVRGEGVSAIYIKKLSDACTA
GLRG_06162     -----LTPKETA-SMHQNG-VLSPSGSCKSFDADAD--GFARGEGVSAIYIKRLSDAVRD
GLRG_09075     -----MTPTTTA-AFTSEG-ILSPDGSCKSFDAKAD--GFARAEAINAIYIKPLSSAIRD
CH063_1187     -----MGPTTTA-AMTEEG-MLSPEGSCKTFDAAAD--GYGRGEAINAVYIKLLDDAIRD
CH063_1411     -----MGPTTTA-AMTEEG-MLSPEGSCKTFDAAAD--GYGRGEAINAVYIKLLDDAIRD
CS_5           ----LSGPIATKGSFGTVGATVSPQGICKTFDADCD--GLGRGEAVNAIYVKRLSHA---
GLRG_11893     ----LSGPIATKGSFGTVGATVSPQGICKTFDADCD--GLGRGEAVNAIYVKRLSHAVRD
CS_19          -----TSVLTSV-IFTETG-VLSPSGKCKTFDILAD--GYGRGEAVNAVYVKKLSKA---
CH063_0663     -----TSVITSI-VFTETG-VLSPSGKCKTFDLLAD--GYGRGEAVNAVYVKRLSDALRD
GLRG_00918     -----TCPLMTL-LYTAQG-LLSPSGICKTFDAAAD--GYGRGEAVNAVFLRRISDAKRT
CH063_0821     ------------------------------------------------------------
CS_13          -----LSPDFGC-FMAEHG-ILSPDASCRTFDAQAN--GYARAEAVNCVFLKSLDSA---
CH063_0053     -----LSPGFTK-LMAEQS-VLSPDASCKTFDAGAN--GYARAEAVNCLFVKRLHCAIRD
CH063_1006     -----MAPGLYV-SMSEQG-VLSPSGSCRTFDAGAD--GYARGEAVNVVYVKRLSDA---
GLRG_07171     -----MAPGLYA-SMSEQG-VLSPNGSCRTFDAGAD--GYARGEAVNVVYVKLLSDALRD
CS_22          -----MAPGLYV-SMYEQG-VLSPNGSCRTFDARAD--GYARGEAVNVVYVKRLSDA---
CS_14          -----LAPGQTQ-QMTEKG-ILSPEGSCKTFSADAD--GYARGEAFTAVFLKPLDAA---
CS_26          -----LAPGMTQ-AMTEQG-VLSPDGSCKTFSADAN--GYGRGEAVCAVYLKPLTAA---
GLRG_11435     -----LAPGMTQ-AMTEQG-VLSPDGSCKTFSADAN--GYGRGEAVCAVYLKPLEAAIRD
GLRG_08640     -----KKDDQHM-NTAKLG-ILSPKFHMPHIRCFCGWLRYGKVEGAGTLFLKRLLDA---
GLRG_10827     -----FTIDQQM-NTAKIG-VLSPRGTCFTFDESAD--GYGRAEGVGAVYLKRLSDAVRD
```

```
GLRG_11425      -----LAVDQHM------------------------------NTGCCAVYLKRLSDAIRD
CH063_0013      -----LTVDQHM-NTAKLG-VLSPTSTCHTFNSYAN--GYGRAEGVGAIYLKRLSDAIRD
GLRG_08482      -----ADPGVQI-ALDKLG-ALSASSRCHTFDARAE--GYARGEGYGAIYLKRPSLAVAD
CH063_1130      -----MSPEPHI-GAAKSG-VLSPTGTCHTFDSSAD--GYGRAEGVNAIYVKRLSAALRD
GLRG_03511      -----LTPDQ---SMTMFN-LTNKDGRCYTFDDRGA--GYARGEGLGVLVLKRLEQAIAD
CS_18           -----LTPDQII-PMSSVG-MTNPDGKCYVFDSRGS--GYARGEGVVTLILKRLDDA---
GLRG_05714      -----LTPDQII-PMSSVG-MTNPDGKCYVFDSRGS--GYARGEGVVTLILKRLDDAVKD
CH063_0922      -----LHPDQFI-SLSAM------------------------------------------
CH063_1282      -----NGPELWS-MLDTVG-VLSPEGKCFSYDHRAS--GFGRGEGSACLVVKPLAAALAD
CS_28           -----ESPEFMF-RASGLG-MVSPDGKCYSLDARAN--GYGRGEGVGTLILKPVSAA---
GLRG_11563      -----ESPEFMF-RASGLG-MVSPDGKCYSLDARAN--GYGRGEGVGTLILKPVSAAIRD
C.P.NRPS_6      -----FGPDMYI-SESKLR-MLSPTGKCQMWDAMAD--GYARGEGVAALLLKPLSKALRD
CH063_0617      -----FGPDMYI-SESKLR-MLSPTGKCRMWDAQAD--GYARGEGVAALLLKPLSKAVRD
C.P.NRPS_1      -----LGPECYI-TESNLN-MLSPNGTSRMWDAEAN--GYARGEGVAAVVLKRLSDAVAD
GLRG_09842      -----LGPEWYV-TESNLN-MLSPTGTSRMWDADAN--GYARGEGVAAVVLKRLSDAVTD
GLRG_09715      -----LGPEKYI-YESQLS-MLSPTGRSRMWDSEAD--GYARGEGFAAVIIKRLSQAIAD
C.P.NRPS_8      -----MDPVTYV-GESKLQ-MLSPDGRGRMWDADVN--GYARGEGVATIVMKRLCDAVAD
GLRG_10367      -----MDPVTYV-GESKLQ-MLSPDGRGRMWDADVN--GYARGEGVATIVMKRLCDAVAD
CS/1            -----LGPENFV-GYSKMK-MLSPDGKSMTWDRDAN--GYARGEGVATVVLKTLSAALED
CH063_0489      -----INVSPYL-TESKLN-MLSPTGRSRMWDADAD--GYARGDGIATVLLKRLSDAIAD
CH063_0392      -----LTPEPYI-GESKLK-MLSPDSRSRMWDEDAK--GYARGDGVAALVLKPLSAAVAD
GLRG_07434      -----LGPEPYI-HESKLK-MLSPDGRSRMWDQGAN--GYARGDGVAAVVLKTLSAALAD
CS/20           -----LGPSQYI-AESKLQ-MLSPNGRSRMWDASAD--GYARGEGFASVVLKTLSAALAD
CH063_0306      -----LGPSQYI-AESKLQ-MLSPNGRSRMWDASAD--GYARGEGFASVVLKTLSAALAD
CH030B          -----LGPEQYI-AESKLQ-MLSPSGRSRMWDADAD--GYARGEGVAAIVMKPLSQALAD
CS_30           -----LGPEQYI-AESKLQ-MLSPAGRSRMWDAEAD--GYARGEGVAAVVLKTLTQA---
CS_29           -----MSPLSYI-SESKLS-MLSPTGRSRMWDAAAD--GYARGEGVAAVVLKTLQQA---
CH063_0199      -----LGPMTFV-LESKLN-MLSPSGRSRMWDAGAD--GYARGEAVCSVVLKTLSQALKD
C.P.NRPS_2      -----LAPDSFI-GASNLH-LLSPDGKSKMWDITAD--GYARGEGICAMFLKPLSQALRD
CH063_0288      -----ITPEQFI-SEASLH-MLSPTGKSRMWDVSAD--GYARGEGFAALFVKPLSQALRD
CS_4            -----ITPEQFI-TEASLH-MLSPTGKSRMWDVSAD--GYARGEGFAALFVKPLSQA---
GLRG_08212      -----ITPEQFI-SEASLH-MLSPTGKSRMWDVSAD--GYARGEGFAALFVKPLSQALKD
CS_15           -----FTPNTMM-PMTALN-FLSPDGKCFAFDARAN--GYGRGEGIGFVVLKKLSDA---
CH063_1445      -----FTPNTMM-PMTALN-FLSPDGKCFAFDSRAN--GYGRGEGIGFVVLKKISDAVRD
GLRG_02583      -----LTPNTMM-PMTALN-FLSPDGKCFAFDARAN--GYGRGEGIGFVVLKKLSDAVRD
CS_32           -----MLPEFWI-SFSSCR-LLADSGRSIAFDQRGT--GFGRGEGCGMIILKPLDQA---
GLRG_03360      -----MLPEFWI-SFSSCR-LLADSGRSIAFDQRGT--GFGRGEGCGMVILKPLDQAIRD
CH063_1222      -----MLPEFWI-SFSSCR-LLADSGRSIAFDQRGT--GFGRGEGCGMLILKPLDQA---
CH063_0134      -----LNPDPFI-IMSSLG-VLGADGRCFAWDSRAN--GYGRGEGMATLLLKPLRDAMAD
CS_21           -----LNPDMFK-ALGSAG-FLSGDGKCYAFDSRAS--GYGRGEGVATIVIKRLGDALAA
GLRG_09267      -----LNPDMFK-ALGSAG-FLSGDGKSYALDSRAS--GYGRGEGVATIIIKRLGDALAA
CH0190          -----LNSDVFV-TMSNLG-FLSPDGISYAFDSRAN--GYGRGEGVAAIVLKALPKALAD
CH063_0873      -----LNPDVFV-SMSTIG-FLSPDGISYSFDHRAN--GYGRGDGIAAVILKRLDHAVAA
CS_9            -----LTPEMGL-ELNALG-VLDPEGASKSFDKRAN--GYGRGEGIAAVVLKRLSDA---
CH063_1490      ------------------------------------------------------------
GLRG_01860      -----LTPEMTL-KLDAAG-VLGPDGKSYSFDHRAN--GYARGEGFGAVVLKRVSDAVRD
CS_25           -----LTPEMTL-KLDAAG-VLGPDGKSYSFDHRAN--GYARGEGFGALVLKRVSDA---
CS_34           -----LSPDTFM-QLQNLS-VLSPEGRSRSFDENGR--GYGRGEGCGIVVLKR-------
GLRG_11770      -----LNPECGI-YLSSLT-MISPEGHCKSFDASGD--GYGRGEGCGMVVLKRLDDALRD
CH0399          -----LNPDYSL-YLSNMT-MLSKEGGCKSFDAHGD--GYARGEGCGIVVLKKLSNAIRD
CS_11           -----LVPESTM-QLNNLG-FLSPLGQSRAFDSKAG--GYARGEGCGVIVMKRLDKA---
GLRG_11840      -----LVPESTM-QLNNLG-FLSPLGQSRAFDSKAG--GYARGEGCGVIVMKRLDKAIAD
CH063_0781      -----LGPEATM-ELNSLG-VLNPDGRSLSFDEAAN--GYGRAEGCGIVVLKPLDQA---
CH063_1357      -----FGPDMSI-LLGAAK-ILSPEGKSKMWDANAN--GFARGEGFGVTILKPLDAALRD
GLRG_10317      -----FGPDMSI-LLGAAK-ILSPDGKSKMWDAKAN--GFARGEGFGVTILKPLEAALRD
CH063_1471      -----LLPDIMG-AMTRLH-FLSPDGKCQSFDHKGN--GYARGEGAGFCILKPLHLALKD
CH063_0506      -----LTPESLTIPLDDAG-FLSPDGRCYSFDHKAN--GYARGEGFGCVAVKKLSHAIRD
GLRG_11835      -----MTPECMSIPLSNAG-FLGPDGRSYSFDHRAN--GYGRGEGFAFILIKLLHNAARD
CS_3            -----ITPECMSLPLSNAG-FLSPDGRSYSFDKKGN--GYGRGEGYAFVLLKPLHKA---
GLRG_11878      -----ITPECMSLPLSNAG-FLSPDGRSYSFDKKGN--GYGRGEGYAFVLLKPLHKAIED
CH063_0051      -----VNPTLFS-QLSAMH-MVSPDGISHSFDAAAN--GYGRGEGIAGLVLKRLSSALAD
CH0135          -----LHPNFMH-QLSSMH-MLSPEGISHSFDHRAN--GYGRGEGIGCLVLKRLKDALRD
```

```
CS_43          -----SAPLFFQ-NLAAGS-FLSTTGQCKPFDKDAD--GYCRGEAIGAVYLKKLSKAL--
GLRG_09268     -----SAPLFFQ-NLAAGS-FLSTTGQCRPFDKDAD--GYCRGDAIGAVYLKKLSKAIAD
GLRG_11857     -----CGPQWFQ-DLAGAS-FLSPTGQCKPFDAGAD--GYCRGDGVAAVFLKRMDKALAD
CH063_0213     -----TSPLWFQ-NLAGAS-FLSTTGQCKPFDAAAD--GYCRGDGVGALFLKRHSQAIAD
CH063_0477     -----TGVNNYF-DLAKAG-FLSKTGQCKPFDDSAD--GYCRADGVGLVVLKSLRQAVSD
CH063_0313     -----TCPHDYR-DLKAAG-FLSPTGQCKPFDADAD--GYCRGEAVGVVVLKSLDAAIEE
CS_44          -----TSPHDYR-DLKAAG-FLSPTGQCKPFDAGAD--GYCRGEAVCVVVLKSLAAAL--
GLRG_11575     -----TSPHDYR-DLKAAG-FLSPTGQCKPFDAGAD--GYCRGEAVCVVVLKSLASAIAE
CH063_1347     -------------NLAAGS-FLNTRGSSGAFDDGAS--GYCRGEGAGILVLKPLSRAVAD
C.PKS          -----TSPFFYQ-NLAAAS-FLSPTGASKSFDASAD--GYCRGEGVGLVVLKRLSDAM--
CS_45          -----TSPTLHQ-NLAAAN-FLSPTGESKSFDARAD--GYCRGEGAGVLMLKRYSAA---
GLRG_11798     -----TSPTLHQ-NLAAAN-FLSPTGESKSFDARAD--GYCRGEGAGVLMLKRYSAAVAD
GLRG_11778     -----TSPSNFN-ILGKAG-FTSTTGGSKSYRADAD--GYCRGEFVGAFVLKRYEDAVAA
CH0398         -----TSPVAYH-MLDKAG-FLSKTGGCKVYRSDAD--GYCRGEFVGAFVLKRLEDAIAD
CS_37          -----NYPHSFA-SLSKSG-VLSDTGNCKTFRDDAD--GYCRGDFVGSVVLKRLEDAL--
GLRG_11836     -----NYPHSFA-SLSKSG-VLSDTGNCKTFRDDAD--GYCRGDFVGSVVLKRLEDAEAD
CH063_1255     -----ASPHGWC-LLSKAG-VLSDTGNCRTFRDDAQ--GYCRGEPPHAV-----------
CS_40          -----SSPHGWC-ALSKSG-VLSATGNCKPFRDDAD--GYCRGEFVGAVVLKRLEDAL--
CS_38          -----TDPDNFC-QLGKGH-FLSLTGQCKVWDEAAD--GYCRADGVGSVVIKRLDDAFAD
CS_36          VSPTKTSPDIYC-MLGKGH-FLSQTGQCKVWDAGAD--GYCRSDGVGSVVIKRLEDAL--
GLRG_08632     -----TNPDNYA-MLGNGH-FLSRTGQCKVWDEGAD--GYCRADGIGSVVIKRLEDAVAD
CH063_0351     -----TNPDIFA-GLSRGQ-FLSKTGPCATFDNGAD--GYCRADGCASVIVKRLDDAIAD
CS_35          -----TNPDIFA-GLSRGQ-FLSKTGPCATFDNGAD--GYCRADGCASVIVKRLDDAIAD
GLRG_04203     -----TNPDIFA-GLSRGQ-FLSKTGPCATFDNGAD--GYCRADGCASVIVKRLDDAIAD
GLRG_11956     -----TNPDFTA-GLDRGH-FLSRTGNCKTFDDGAD--GYCRGEGIGTIILKRLDDAIAD
CH063_1223     -----------------------------TFDDAAD--GYCRGEAAVTLVLKRLGDARRD
GLRG_08620     -----TNPDMTA-GLDRGH-FLSHTGNCKTFDDSAD--GYCRGEAVVTIILKRLDDAMVD
CH063_0250     -----TNPDMTA-GLDRGH-FLSRTGNCKTFDDEAD--GYCRGEAVVTIILKRLEDA---
CS_39          -----TGPSMFA-GLGEGR-FLSTTGSCKTFDSGAD--GYCRADGVGVIALKRLDSAY--
GLRG_03507     -----TGPNLWA-GLSRAG-FVSPTGSCKTFDETAD--GYCRGEGVGIIVLKRLEDAIQA
GLRG_03509     -----TGPNLWA-GLSRAG-FVSPTGSCKTFDETAD--GYCRGEGVGIIVLKRLEDAIQA


GLRG_10537     -----------------------------------------------------------
CS_23          -----------AVIRGTAANASGRTA-----------------GLANPSPDAQAAVTRM
CH063_1191     G------DPVRAVIRSTCVAGNGRTP-----------------GLTTPNPVIHERLMRR
CS_16          G--------------------NGRTL-----------------GLTTPNPEVHERLMRR
GLRG_06162     G------DPIRSVIRSTCTAGNGRTL-----------------GLTTPNPEIHERLMRR
GLRG_09075     G------NAIRAVIRATGANSDGHSQ-----------------GLFTPNQVAQEALMRK
CH063_1187     N------NPIRAIIRASGTNSDGNSS-----------------TLLTPNSAMHEALMRK
CH063_1411     N------NPIRAIIRASGTNSDGNSS-----------------TLLTPNSAMHEALMRK
CS_5           -----------AVIRSTATQHDGSGS-----------------GMLIPNGHAQESLIRH
GLRG_11893     G------NPIRAIIRSTATQHDGSGS-----------------GMLIPNGHAQESLIRH
CS_19          -----------AIIRGSGTNNDGRSR-----------------GIMTPNDVLQEKLIRQ
CH063_0663     G------NPVRAIIRASATNNDGRSN-----------------GIMSPNTYLQEALIRK
GLRG_00918     G------DPIRALVRASATGHDGRKS-----------------GQLKPDTMGQEVLIRK
CH063_0821     -----------ALVRASATGHDGRKV-----------------GQLKPDAAAQELLIRK
CS_13          -----------AIIRGSATNADGRTV-----------------GMSTPNAGAHETLIRS
CH063_0053     G------NPIRAVIRGSATNSDGKTL-----------------GLTTPSAQAQQQLIRD
CH063_1006     -----------AVIRGTSSNADGRTP-----------------SLTIPSFESHEAMIRQ
GLRG_07171     G------NPIRAVIRGTSSNADGRTP-----------------SLTIPSFESHEAMIRQ
CS_22          -----------AVIRGTSSNADGKTP-----------------SLTIPSFESHEAMIRQ
CS_14          -----------AVIRAAVANSDGKTQ-----------------GITQPNGAAHEAMIRL
CS_26          -----------AVIKGTATNFDGRTN-----------------GISLPSTEAHEALIRQ
GLRG_11435     R------NPIRAVIKGTATNFDGRTN-----------------GISLPSTEAHEALIRQ
GLRG_08640     -----------------------------------------------------------
GLRG_10827     G------DPIRAVLRGSSVNHNGRVS---------------SASISYPGVNGQAYVMAD
GLRG_11425     GDPIRDRDPIRAVMRSSATNNNGKVP---------------VVGITHPNRDAQVEVIRH
CH063_0013     G------DPIRGVIRSSATNNNGKAP---------------AVGITYPGFDGQRNVMRH
GLRG_08482     H------SPIRAMIRGTAINSNGRTG-----------------GITRPSARGQEMVIRE
CH063_1130     G------NPIRAIIKGSAVNASGRTP-----------------GISLPSGNMQEVVIRK
GLRG_03511     G------DAIHAIVMESGSNHDGKTG-----------------GIFLPNSDAQESLARR
```

```
CS_18          ----------HAIIRNSGLNQDGKTV------------------GLTLPNPIAQANLMRL
GLRG_05714     G------DKVHAIIRNSGLNQDGKTV------------------GLTLPNPIAQANLMRL
CH063_0922     ------------------------------------------------------------
CH063_1282     G------DTVRAVVRNTAASHSGRIA-----------------GGITMPSQDAQEALARR
CS_28          -----------AVLRGTGVNSDGRGT-----------------GGITLPNKAAQESLIRD
GLRG_11563     G------NVIRAVLRGTGVNSDGRGT-----------------GGITLPNKAAQECLIRD
C.P.NRPS_6     N------DPIQAVIRNSGVNSDGRTP------------------GITMPSAMAQAKLTTE
CH063_0617     G------DPIHAVVRNSGVNSDGRTP------------------GITMPSAAAQARLTAE
C.P.NRPS_1     G------DFIECVVRETGVGQDGHTN------------------GITTPGAKSQASLIRR
GLRG_09842     G------DFIECVVRETGVCQDGRTN------------------GITTPSSESQASLIRH
GLRG_09715     N------DHIECIIRETGVNQDGRSE------------------GLTVPSSEAQAALIRS
C.P.NRPS_8     G------DDIECVIRETGVNQDGKAR------------------GITMPSASAQSALIRD
GLRG_10367     G------DDIECVIRETGVNQDGKTR------------------GITMPSASAQSALIRD
CS/1           G------DHIECIIRETGTNQDGATP------------------GITMPSAESQRALIRD
CH063_0489     G------DAIEGIIRNTGFTHDGRTL------------------GITMPSGQAQADLIRR
CH063_0392     G------DHIESVIRECLVNQDGRTA-----------------EGITVPSSEAQAELIRR
GLRG_07434     G------DHIEAIIRETGVNQDGRSR------------------GITMPSAAAQASLIRA
CS/20          G------DHIECIIRETGVNQDGRTK------------------GITMPSPLAQANLIRD
CH063_0306     G------DHIECIIRETGVNQDGRTK------------------GITMPSPLAQANLIRD
CH030B         G------DSIECVIRETGLNQDGKTP------------------GITMPSSSAQAALIRS
CS_30          -----------CIIRETGINQDGKTP------------------GITMPSAKAQASLIRS
CS_29          -----------SIIRETGVNQDGKTT------------------GITMPSHLAQEVLIRD
CH063_0199     G------DTIECVIRETGINQDGRTT------------------GITMPNHDAQEALIRA
C.P.NRPS_2     G------NRVDALIRETGINSDGRTQ------------------GITQPSATAQTALIRK
CH063_0288     G------DSVISIIRETGVNSDGRTQ------------------GITMPSPLAQAELIRD
CS_4           -----------SIIRETGVNSDGRTQ------------------GITMPSPIAQAELIRD
GLRG_08212     G------DSVISIIRETGVNSDGRTQ------------------GITMPSPIAQAELIRD
CS_15          -----------AVIRGTHVNQDGLTTGELSASKAHNPSKLTVQQGITLPSKEAQVANIRA
CH063_1445     N------DTIRAVIRGTHVNQDGLTT------------------GITLPSKEAQVANIRS
GLRG_02583     N------DTIRAVIRGTHVNQDGLTT------------------GITLPSKEAQVANIRA
CS_32          -----------AVIRGTGINQDGKTP------------------GITMPSGAAQEKLMRQ
GLRG_03360     N------DSIRAVIRGTGINQDGKTP------------------GITMPSGAAQEKLMRQ
CH063_1222     -----------AVIRGTGINQDGKTP------------------GITMPSGAAQEKLMRQ
CH063_0134     G------DPIHAVIRETAINQDGKTP------------------TITSPSSEAQEELIRA
CS_21          G------DPIRGVIRGSALNQDGKTE------------------TITTPSLEAQEALIRA
GLRG_09267     G------DPIRGVIRGSALNQDGKTE------------------TITTPSLEAQEALIRA
CH0190         H------DPVRTIIRETALNQDGKTP------------------AITAPSDMAQEQLIRE
CH063_0873     G------DPIRAIIRGSALNQDGKTP------------------TITTPSQEAQERLMGA
CS_9           -----------AVIRNTGCNHDGHSP------------------GLTAPAKEAQAELMRL
CH063_1490     ------------------------------------------------------------
GLRG_01860     G------DVIRAVIRNSSTNQDGRSP------------------GITQPTKAGQVALIKH
CS_25          -----------AVIRNSSTNQDGRSP------------------GITQPTKAGQVALIKH
CS_34          -----------AIIRATGCNSDGWTK------------------GMALPSGESQMQLIKD
GLRG_11770     G------DPIRAVIPGTGVNSDGFTQ------------------GFTMPSSESQAALIRD
CH0399         G------DPIRAVIRGSGVNADGFTQ------------------GFTMPSATAQADLICD
CS_11          -----------AIIRGSGVNQDGWTQ------------------GVTMPSGEAQAALIKY
GLRG_11840     G------DTIRAIIRGSGVNQDGWTQ------------------GVTMPSGEAQAALIKY
CH063_0781     -----------AVIRGSGVNSDGWTK------------------TISTPSGQAQADLIRQ
CH063_1357     G------DTVRAVVLASAANEDGRTP------------------GISLPNSEAQQALIRT
GLRG_10317     G------DTIRAVVLASAANEDGRTP------------------GISLPNSEAQQALIRK
CH063_1471     G------DVIRGVIRNTA------------------------------------------
CH063_0506     G------NVIRAVVRATGVNQDGKTP------------------SITQPSLEAQIALIRK
GLRG_11835     G------DVIRAVIRATGVNQDGRTP------------------SITQPNSKAQVDLIRR
CS_3           -----------AVIRATGVNQDGRTP------------------SITQPSSQAQADLIRS
GLRG_11878     G------DPIRAVIRATGVNQDGRTP------------------SITQPSSQAQADLIRS
CH063_0051     G------DVIRAVVRATGVNSDGKTP------------------GITVPSEDAQAELIRR
CH0135         G------DTIRAVIRGTGTNADGKTP------------------GITQPSSEAQAELIRT
CS_43          -----------GVISATAVNQNQNST------------------PIFVPNPLSLTDVFRT
GLRG_09268     G------DQVLGVVSATAVNQNQNST------------------PIFVPNPLSLTDVFRT
GLRG_11857     G------DEILGVIGATVVQQNQNST------------------PIVVPNEASLSDLFKT
CH063_0213     G------DQVLGVIPATAVQQNLNTT------------------PIFVPNAPSLSDLFSS
CH063_0477     G------DHIMGVIPSVATNQGGIGA-----------------PGITVPDGICQKALYST
```

```
CH063_0313     N------DHILGVVVGSATNQNPKEG------------------PIVVPNSKAQASLLRK
CS_44          -----------GVVVGSATSQNDQNG------------------PIIVPNAKSQASLLRK
GLRG_11575     G------DEIIGVVVGSATSQNDNEG------------------PIVVPNAKSQASLLTK
CH063_1347     G------DRVLGVLAGSAVNQNSNRS------------------PITVPDSSSQGSLYRR
C.PKS          -----------GTILATAVRQSSNKV------------------PITVPYSPSQTQLYRK
CS_45          -----------GVIAGSAVNQNANTA------------------PITVPVSSSQTSLYMR
GLRG_11798     G------DNIFGVIAGSAVNQNANTA------------------PITVPVSSSQTSLYMR
GLRG_11778     N------DNILAVILSSARNHSGDAV------------------SITHSDPEAQQELMAE
CH0398         N------DNILSVIASSARNHSGTAT------------------SITRSDHRAQELLMAE
CS_37          -----------AVIAGSGRNHSGNSP------------------SITTSDAGAQERLFRK
GLRG_11836     N------DKILAVIAASGRNHSGNSP------------------SITTSDAGAQERLFRK
CH063_1255     -----------------------------------------G------------------
CS_40          -----------AVIAASGKNQSGNAA------------------SITTPDAGAQERLFRR
CS_38          N------DTILATILAANTNLSAEAV------------------SITHPHAPTQADNYKK
CS_36          -----------GTIIAGATNHSAESI------------------SITQPHAGAQRDNYRQ
GLRG_08632     N------DNIIATVLAGATNHSAYAA------------------SITQPHAGAQKSNYAQ
CH063_0351     K------DNVLAVILGTA-NHSADAI------------------SITHPHGPTQSILSRH
CS_35          K------DNVLAVILGTATNHSADAI------------------SITHPHGPTQSILSRY
GLRG_04203     K------DNVLAVILGTATNHSADAI------------------SITHPHGPTQSILSRH
GLRG_11956     H------DPILGVVLGAYTNHSAESE------------------SITRPHVGAQRAIFSK
CH063_1223     N------DPIEACLLSVATNHNAEAE------------------SITRPNTAAQRELFRG
GLRG_08620     K------DPIQACIRGIATNHNAEAE------------------SITRPHSEAQVELFEH
CH063_0250     -----------ACILGVATNHNAEAE------------------SITRPHAAAQKDLFEY
CS_39          -----------GVIKSWATNHSSRAI------------------SITQPHADTQARLFLK
GLRG_03507     G------DNIQGVIRGIATNHSANAL------------------SITQPHGPTQKKLYKQ
GLRG_03509     G------DNIQGVIRGIATNHSANAL------------------SITQPHGPTQKKLYKQ


GLRG_10537     ------------------------------------------------------------
CS_23          AYKNAGI-SDF----WATQFLE--------------------CHGTGTLAGDPVEVKGAA
CH063_1191     GHEIAGI-ADL----SKTAMVE--------------------CHGTGTPVGDPLEVSAVA
CS_16          GHKLAGI-TDL----SKTAMVE--------------------CHGTGTPVGDPLEVSAVA
GLRG_06162     GHRLAGV-TDL----SKTAMVE--------------------CHGTGTPVGDPLEVSAVA
GLRG_09075     VYRDAGL--DP----SQTAYVE--------------------CHGTGTATGDPIETSAVG
CH063_1187     VYLDAGL--DP----SKTAFVE--------------------CHGTGTPTGDPLETNAVG
CH063_1411     VYLDAGL--DP----SKTAFVEVSQMCLYVLGGRGATQLERQCFCEGTPTGDPLETNAVG
CS_5           TYDIAGI-RDF----SKTAFFE--------------------CHGTGTVVGDPIEVAAVE
GLRG_11893     TYDIAGI-RDF----SKTAFFE--------------------CHGTGTIVGDPIEVAAVE
CS_19          VYSQAGI-DDL----SKTAFFE--------------------CHGTGTSTGDPLEASAVA
CH063_0663     TYEKAGVPF------NKTAFFE--------------------CHGTGTPTGDPLEVAAVA
GLRG_00918     AYALAGI---PDSQFGETAWIE--------------------CHGTGTAVGDPIELRSVA
CH063_0821     AYALAGI---PDSQFGDTAWIE--------------------CHGTGTAVGDPIEMRSVA
CS_13          AYRMAGI---W--DFGETAMVE--------------------CHGTGTTVGDAIETCAVA
CH063_0053     AYRQAGI---PESDMWQTAMVE--------------------CHGTGTAVGDTIEACTVG
CH063_1006     AYKMAAI---GGRDIADTGFVE--------------------CHGTGTAAGDPIETTAIA
GLRG_07171     AYKMAAI---SGQHIAATGFVE--------------------CHGTGTAAGDPIETTAIS
CS_22          AYKMAAI---GGQHVADTGFVE--------------------CHGTGTAAGDPIETTAIS
CS_14          AYQQAGI-TDF----SKTAYFE--------------------CHGTGTAVGDPIETGAVA
CS_26          AYAHAGI-TDF----SETTFVE--------------------CHGTGTPIGDPIEANAVA
GLRG_11435     AYAHAGI-TDF----SDTTFVE--------------------CHGTGTPIGDPIEANAVA
GLRG_08640     ------------------------------------------------------------
GLRG_10827     AYKRSGN-LDP----MLTGYFE--------------------CHGTGTAVGDPIEVEAIG
GLRG_11425     AYHSAGN-LDP----RLTGFFE--------------------CHGTGAAIGDPLEVHAVL
CH063_0013     AYQRGGL--DP----MLTGYFE--------------------CHGTGTAIGDPLEVHAVS
GLRG_08482     AYRSAG---DL--PFRDTSYFE--------------------CHGTGTYVGDPIEVAALG
CH063_1130     AYRDAGL--DF----ADTDYVE--------------------CHGTGTPVGDPIEVDAIG
GLRG_03511     VYANAGL--DP----RETLYVE--------------------AHGTGTAAGDNAEVSSIS
CS_18          VYKNAGL--DP----ADTVYVE--------------------AHGTGTQAGDNAEISSIA
GLRG_05714     VYKNAGL--DP----ADTVYVE--------------------AHGTGTQAGDNAEISSIA
CH063_0922     ------------------------------------------------------------
CH063_1282     VHGEVGL--DP----KDTGFVE--------------------GHGTGTAVGDPIDAAAIA
CS_28          VYARNNL--DL----DHTGFIE--------------------GHFTGTPAGDPIEASAIA
```

```
GLRG_11563      VYARHNL--NP----DETGFLE--------------------GHFTGTPAGDPIEASAIA
C.P.NRPS_6      TYRAAGL--DPSKAEDRCQYFE--------------------AHGTGTQAGDPTEAKGIF
CH063_0617      TYRAAGL--DPLNPDERCQYFE--------------------AHGTGTQAGDPTEARGIY
C.P.NRPS_1      VYAKAGL--DVIK--DGCQYFE--------------------AHGTGTPAGDPVEAEAIH
GLRG_09842      VYAKAGL--DVVK--DGCQYFE--------------------AHGTGTPAGDPVEAEAIH
GLRG_09715      TYSKCGL--DWRTTEDRCQYFE--------------------AHGTGTQAGDPKEARAIR
C.P.NRPS_8      TYRRAGL--DPHRPADRPQYFE--------------------AHGTGTPAGDPVEAEAIH
GLRG_10367      TYRRAGL--DPHRPADRPQYFE--------------------AHGTGTPAGDPIEAEAIH
CS/1            TYRRASL--DPRDPADRCQYFE--------------------AHGTGTQAGDPIEAEAIY
CH063_0489      TYEEAGL--DLANRANWPQFFE--------------------AHGTGTPVGDPQEASALA
CH063_0392      TYAKSGL--DATNVDDRCQYFE--------------------AHGTGTGAGDPREARAIA
GLRG_07434      TYAKAGL--DPESKTDRCQYFE--------------------AHGTGTPAGDPVEAEAVH
CS/20           TYKRAGL--DLNKPSDRPQYFE--------------------AHGTGTPAGDPVEAEAIS
CH063_0306      TYKRAGL--DLSKPSDRPQYFE--------------------AHGTGTPAGDPVEAEAIS
CH030B          TYARAGL--DLSKPSDRPQFFE--------------------AHGTGTPAGDPIEAAAIR
CS_30           TYARAGL--DLCKASDRPQFFE--------------------AHGTGTPAGDPIEAAAIS
CS_29           TYARAGL--DLSRIGDRPQFFE--------------------AHGTGTPAGDPQEAEAIT
CH063_0199      TYARAGL--DINNPQERCQFFE--------------------AHGTGTPAGDPQEANAIA
C.P.NRPS_2      TYRNAGL--DLGRAEDRPQYIE--------------------AHGTGTQAGDPQEASAIS
CH063_0288      TYRRSGL--DALSPTDRCQYFE--------------------AHGTGTQAGDPREAQAIS
CS_4            TYRRSGL--DPQSPADRCQYFE--------------------AHGTGTQAGDPREAQAIS
GLRG_08212      TYRRSGL--DPHSPADRCQYFE--------------------AHGTGTQAGDPREAQAIS
CS_15           LYTKHNL--DM----KQTSFVE--------------------CHGTGTQAGDFRELKAIS
CH063_1445      LYSKHNL--DM----KQTAFVE--------------------CHGTGTQAGDFRELKAIS
GLRG_02583      LYAKHDL--DM----KQTAFVE--------------------CHGTGTQAGDFRELKAIS
CS_32           IYRNAGL--DP----SDCGYVE--------------------AHGTGTKVGDPIEATAIH
GLRG_03360      VYRNAGL--DP----NDCGYVE--------------------AHGTGTKVGDPIEATAIH
CH063_1222      IYRNAGL--DP----NDCGYVE--------------------AHGTGTKVGDPIEATAIH
CH063_0134      CYRRAGL--DP----AKTPYVE--------------------AHMTGTPTGDPIEAKAIS
CS_21           CYQNSGL--NP----NDTQYFE--------------------AHGTGTQAGDTVEARAIA
GLRG_09267      CYENAGL--SP----KDTQYFE--------------------AHGTGTQAGDTIEARAIA
CH0190          CYNRAGI--DT----SQTSYIE--------------------AHGTGTPTGDPLEISAIS
CH063_0873      CYRSCGL--DP----GETGFVE--------------------AHGTGTPTGDPIEVTAIA
CS_9            TYAQAQL--DP----SETRFFE--------------------AHGTGTNVGDPTEASAIS
CH063_1490      -----------------------------------------------------------
GLRG_01860      VYDRANL--DP----SLTRFAE--------------------AHGTGTPVGDPIEASALA
CS_25           VYDRANL--DP----SLTRFAE--------------------AHGTGTPVGDPVEASALA
CS_34           VYDTFGL--DY----SLTQYVE--------------------AHGTGTKAGDPTEAHAIY
GLRG_11770      VYEGAEL--DM----SDTQFVE--------------------CHGTGTKAGDPIETRAIY
CH0399          VYKTAGL--DM----GLTTYVE--------------------AHGTGTKVGDPIEARAIY
CS_11           VYESNKL--DY----GGTQYVE--------------------AHGPGTQAGDPTETGAIY
GLRG_11840      VYESNKL--SY----GDTQYVE--------------------AHGPGTQAGDPTETGAIY
CH063_0781      VYESNGL--DY----DSTQYVE--------------------AHGTGTKVGDPIELGAIH
CH063_1357      AYENAGV--DP----AETGYVE--------------------AHGTGTQAGDPLEARAIL
GLRG_10317      AYRDAGV--DP----AETGYVE--------------------AHGTGTQAGDPLEARAIL
CH063_1471      ------------------------------------------AHGTGTPAGDPVEAAALA
CH063_0506      TYEQAGL--GF----STTEYVE--------------------AHGTGTAVGDPIEARAIG
GLRG_11835      TYQAGGL--DL----AHTEYVE--------------------AHGTGTPVGDPIEASGIG
CS_3            TYEAAGL--SL----DDTGYVE--------------------AHGTGTPVGDPLEAAAIS
GLRG_11878      TYRTAGL--SL----DDTGYVE--------------------AHGTGTPVGDPLEAAAIS
CH063_0051      VYEGADL--DM----ARTAYFE--------------------AHGTGTPKGDPIEVRAIR
CH0135          TYEAAGL--SL----SDTQYFE--------------------AHGTGTALGDPIELSAIG
CS_43           VISKSGL--GI----KDISVVE--------------------AHGTGTAVGDPAEYDSIR
GLRG_09268      VISKSGL--GV----KDISVVE--------------------AHGTGTAVGDPAEYDSIR
GLRG_11857      VLDKAGV--QP----GQVSVIE--------------------AHGTGTAVGDPAEYGAIR
CH063_0213      VTTRAAI--AP----SQISVVE--------------------AHGTGTAVGDPAEYNGIR
CH063_0477      LLRKSGI--KA----DQVSYVE--------------------AHGTGTQVGDPIEISSIR
CH063_0313      VMDMSGV--LP----EDVTYVE--------------------AHGTGTGVGDPIEVQSLR
CS_44           VMEVSNV--VP----GDVTYVE--------------------AHGTGTDVGDPIEVSSIR
GLRG_11575      VMNMSNV--VP----EDITYVE--------------------AHGTGTEVGDPIEVSSIR
CH063_1347      VLDEAGI--QP----HEVTYVE--------------------AHGTGTKVGDPIEYESVR
C.PKS           ILGAAGI--EA----EEVTYVE--------------------AHGTGTPVGDPLEFGAIK
```

```
CS_45         VLQRAHM--DP----KSVSYVE--------------------AHGTGTPKGDPIECASIR
GLRG_11798    VLQRAHM--DP----KSVSYVE--------------------AHGTGTPKGDPIECASIR
GLRG_11778    VLRKARL--EP----SDVSYVE--------------------MHGTGTQVGDYAEMMAVS
CH0398        TLRKACL--EP----RDISYVE--------------------MHGTGTQVGDVAEMTAVA
CS_37         VLRNAHV--SA----SDISYVE--------------------MHGTGTQVGDPAEMGAVG
GLRG_11836    VLRNAHV--SA----DDISYVE--------------------MHGTGTQVGDPAEIGAVG
CH063_1255    ---------------EDVTIED--------------------------------------
CS_40         VLRTAGV--AP----ADVSYVE--------------------AHGTGTPVGDPCEMTAVS
CS_38         VTAKAGI--SP----LDVSYVE--------------------LHGTGTQARDREEARSMS
CS_36         VMDKAGV--RP----LDVSYVE--------------------LHGTGTQVGDAVESESVL
GLRG_08632    VTQAAGI--NP----LDVSFVE--------------------LHGTGTQVGDAIESESVC
CH063_0351    ILDEAGV--DP----LDVDYVE--------------------MHGTGTQAGDGTEMVSVT
CS_35         ILDEAGV--DP----LDVDYVE--------------------MHGTGTQAGDGTEMVSVT
GLRG_04203    ILDEAGV--DP----LDVDYVE--------------------MHGTGTQAGDGTEMVSVT
GLRG_11956    ILNEASV--DP----FSISYVE--------------------MHGTGTQAGDAGEMSSVL
CH063_1223    LLADAVV--RP----NDISYVE--------------------MHGTGTQAGDAGETSSIV
GLRG_08620    LLAEANI--SP----IDISYVE--------------------MHGTATQVGDAAETTSVV
CH063_0250    ILTEANV--SS----NDISYVE--------------------MHGTGTQAGDAGETTSVV
CS_39         VLSEANL--GP----ADITYVE--------------------MHGTGTPVGDTAELTSVL
GLRG_03507    VLRKANL--TP----DQIQYVE--------------------MHGTGTQAGDVTEMNSVV
GLRG_03509    VLRKANL--TP----DQIQYVE--------------------MHGTGTQAGDVTEMNSVV


GLRG_10537    ------------------------------------------------------------
CS_23         SVFAAG---------------RRPGQELVIGS-------------------IKSNIGHSE
CH063_1191    NIWG--------------------QDGIYIGS-------------------V--------
CS_16         NIWG--------------------ESGIYIGSVRTTATMCNFLTNTNIKCQVKPNIGHGE
GLRG_06162    NIWG--------------------ENGIYIGS-------------------VKPNIGHGE
GLRG_09075    EVFG--------------------EKGVYIGS-------------------VKPNVGHSE
CH063_1187    NVFGEH--------------------GVYIGS-------------------VKPNVGHSE
CH063_1411    NVFGEH--------------------GVYIGSTDDVAVKKSSFS-------VKPNVGHSE
CS_5          RIFGEY--------------------GILIGS-------------------AKPNFGHSE
GLRG_11893    RIFGEH--------------------GILIGS-------------------AKPNFGHSE
CS_19         RIWGPH-------------------GGILMGA-------------------VKPNVGHSE
CH063_0663    RIWKDN-------------------EGVMIGA-------------------VKPNVGHSE
GLRG_00918    NIFG--------------------QQGIIVGS-------------------VKPNVGHSE
CH063_0821    NVFG--------------------RDGIVVGS-------------------VKPNVGHSE
CS_13         RVFG--------------------EKGVVIGS-------------------TKPALGHSE
CH063_0053    TVFG--------------------EKGMLIGS-------------------VKPNLGHSE
CH063_1006    KVFGSA--------------------GVYIGS-------------------CKPNIGHSE
GLRG_07171    KVFGDA--------------------GVYIGS-------------------CKPNIGHSE
CS_22         KVFG--------------------DVGVYIGS-------------------CKPNIGHSE
CS_14         RVFG--------------------EHGIHITS-------------------VKPNVGHTE
CS_26         RVFGGP------------------ERGPYIGS-------------------VKPNLGHSE
GLRG_11435    RVFGGP------------------ERGPYIGS-------------------VKPNLGHSE
GLRG_08640    ------------------------------------------------------------
GLRG_10827    RAMNHS-------------RRPGVDDKLMIGA-------------------VKTNVGHGE
GLRG_11425    LAMNDQ---------------RSDEEPLLIGA-------------------VKTNIGHSG
CH063_0013    DIMNAT--------------RTEADGPLHIGA-------------------VKTNIGHSE
GLRG_08482    RVFAPE---------------RSADDPLLVGS-------------------VKSNVGHGE
CH063_1130    RCFSPR-----------------QGPPLIIGS-------------------VKTNVGHSE
GLRG_03511    KVFGRE---------------AGRKSDLPIGS-------------------IKANIGHLE
CS_18         EVFCPE---------------GQREDGLYVGS-------------------IKSNIGHLE
GLRG_05714    EVFCPE---------------GQREDGLYVGS-------------------IKSNIGHLE
CH063_0922    ------------------------------------------------------------
CH063_1282    SVYASP---------------RSTSDPVYLGS-------------------VKSNIGHLE
CS_28         AVFQRG-----GPRAAAD------KPSLYVGA-------------------VKANIGHVE
GLRG_11563    AVFQRG-----GPRAAAD------RPSLYVGA-------------------VKANIGHVE
C.P.NRPS_6    ESFFST-----EGAPG------AEVSPLYVGS-------------------IKTVVGHLE
CH063_0617    DSFFAS-----SADSKDTNVDMTARRPLLVGS-------------------IKTVVGHLE
C.P.NRPS_1    SVFIDP-----SRDKRPT-------GPLYVGS-------------------IKTVIGHTE
GLRG_09842    SVFTDS-----SRDGVRT-------DPLYVGS-------------------VKTVIGHTE
```

```
GLRG_09715     DAFFPE-----NVQQELE------KDTLYVGS-------------------VKTVIGHLE
C.P.NRPS_8     SAFFGG--------RPTKSNEERGRDTLLVGS-------------------IKTVCGHAE
GLRG_10367     SAFFGG--------RPTKTNEERGRDTLLVGS-------------------IKTVCGHAE
CS/1           RSFFDD-----DSGDADS---------LYVGS-------------------LKSIIGHTE
CH063_0489     SAFYPD-----GDNAYEQ------DDRLLVGS-------------------IKTLVGHTE
CH063_0392     TAFFGP-----NHRSDDP------SDKLYVGS-------------------IKTVIGHTE
GLRG_07434     TAFFGN-----GNLDADR---------LLVGS-------------------VKTIIGHTE
CS/20          SAFFGP-----ESGFQRKS---DADPKLYVGS-------------------IKTVIGHTE
CH063_0306     SAFFGA-----ESGFQRRP---DADPKLYVGS-------------------VKTVIGHTE
CH030B         EAFFGA-----DSHFVPR----GPDDTLYVGS-------------------IKTVIGHTE
CS_30          DAFFGE-----DVKSDSK----DKHSTLFVGS-------------------IKTVIGHTE
CS_29          KAFFGD-----EVLDTLQ------PQDLLIGS-------------------VKTVIGHSE
CH063_0199     SAFFGQ-----GDNSNAG------DDPLFVGS-------------------IKTIVGHTE
C.P.NRPS_2     QTFF-----------PPN----EEHDKLFVGS-------------------VKTVIGHTE
CH063_0288     TAFFGD-----EASAAAD----DDEQKLVVGX-------------------VKTVIGHTE
CS_4           TAFFGD-----KASTVAE----SDDQKLVVGS-------------------IKTVIGHTE
GLRG_08212     TAFFGD-----KASTVAE----SEDQKLVVGS-------------------IKTVIGHTE
CS_15          ETLA---------EGRKT------ENPVFVGS-------------------VKTNIGHLE
CH063_1445     ETFGDA---------------RTTENPVFVGS-------------------VKTNIGHLE
GLRG_02583     ETLAEG---------------RTTENPVFVGS-------------------VKTNIGHLE
CS_32          NVIGQN---------------RSAKDPLFIGS-------------------VKSNIGHLE
GLRG_03360     NVIGQN---------------RSTKDPLFIGS-------------------VKSNIGHLE
CH063_1222     NVIGQN---------------RSSKDPLFIGS-------------------VKSNIGHLE
CH063_0134     CVFGKG---------------RGVSNPVLVGS-------------------IKTNLGHLE
CS_21          SVFKCT------------------SEPLLIGS-------------------VKTNLGHTE
GLRG_09267     SVFSSS------------------SEPLLIGS-------------------VKTNLGHTE
CH0190         RAFQGQ--------------------PLHVGS-------------------VKANIGHTE
CH063_0873     NIIASA--------SPPG------SDPLLLGS-------------------VKSAIGHTE
CS_9           EMFTPY---------------RSAEEPLHIGA-------------------LKSNIGHTE
CH063_1490     ------------------------------------------------
GLRG_01860     EMFAPH---------------RSQDEPLFIGA-------------------LKSNVGHLE
CS_25          EVFAPH-------RSP--------DEPLFVGA-------------------LKSNVGHLE
CS_34          NTIGRA----------------TRNQKLIMGS-------------------VKPNIGHLE
GLRG_11770     ETLGKK---------------ASSSQPLIIGS-------------------VKPSIGHLE
CH0399         ETLGRN---------------VKRGQKLMIGS-------------------VKPNIGHLE
CS_11          RTIGLK-----GLETNPS------RKKLWVGS-------------------VKPNIGHLE
GLRG_11840     RTIGLE-----GLKTNPS------RKKLWVGS-------------------VKPNIGHLE
CH063_0781     DTIGRN---------------ASESRKLYVGS-------------------VKANIGHLE
CH063_1357     KTIGDL---------------PGRTSELYVGS-------------------VKS------
GLRG_10317     NTIGDL---------------PNRKSDLYVGS-------------------VKTNIGHLE
CH063_1471     ATFGSA---------------RRAGDPVYMGS-------------------VKSNVGHLE
CH063_0506     EVFRTS-----------------RKTPLHVGS-------------------VKSNIGHLE
GLRG_11835     AVFRQH-----------------RTRPVHIGS-------------------IKSNIGHLE
CS_3           QVFKTA----------------MRNEPLVVGS-------------------VKSNIGHLE
GLRG_11878     QVFQTA----------------SRSEPLVVGS-------------------VKSNIGHLE
CH063_0051     ATIAAA-------------RKLYHAGPLYVGS-------------------VKPNIGHTE
CH0135         AILGAA--------RSPE------TEPLYVGS-------------------IKANVGHTE
CS_43          QVFGGS--------------ARSGLKPLQIGS-------------------VKGLIGHTE
GLRG_09268     QVLGGP--------------ARSGLKPLQIGS-------------------VKGLIGHAE
GLRG_11857     QVFGAS---------------SKRQTPLSLGS-------------------VKGSIGHTE
CH063_0213     KVLGGP----------------GRDSPLVLGS-------------------VKSLVGHTE
CH063_0477     QVFGGS----------------QRTKPLYIGS-------------------LKANIGHSE
CH063_0313     DAFGGP----------------LRTSTLHFSS-------------------IKGNIGHAE
CS_44          EAFGGP----------------SRKSKLRFAS-------------------IKGNIGHVE
GLRG_11575     EAFSGP----------------SRKSKLRFAS-------------------IKGNIGHVE
CH063_1347     STFSGP----------------LRTGPVYLGS-------------------VKDNIGHAE
C.PKS          EVFASG----------------SRRVPLHFAS-------------------VKGNIGHTE
CS_45          EVFGNQ-----------------PSRKLYFSS-------------------VKGSIGHLE
GLRG_11798     EVFGNQ-----------------PSRKLYFSS-------------------VKGSIGHLE
GLRG_11778     KALGRR----------------RRPLPLPLGS-------------------IKANAGHGE
CH0398         NVFGNA----------------SRSVPLSVGT-------------------IKANIGHSE
CS_37          NLLKHT---------------PRADGPVPVGG-------------------VKANFGHGE
```

```
GLRG_11836     NLLKHK---------------SRADGPVPVGG-------------------VKANFGHGE
CH063_1255     ------------------------AEPLTVGS-------------------VKANIGHSE
CS_40          NVLARP-------------GVEDDTHRLAVGA-------------------VKANIGHSE
CS_38          DVFAPA-----GPGPRRKK-----ENRLRLGA-------------------VKSNIGHGE
CS_36          NFFAPL-----GQRQRP-------EQRLHMGA-------------------VKSNIGHGE
GLRG_08632     DFFAPL-----SPRRRA-------DQPLHLGA-------------------VKSNVGHGE
CH063_0351     NVFAPA------DRKRPA------DRPLYLGA-------------------VKSNVGHGE
CS_35          NVFAPA------DRKRPA------DRPLYLGA-------------------VKSNVGHGE
GLRG_04203     NVFAPA------DRKRPA------DRPLYLGA-------------------VKSNVGHGE
GLRG_11956     EVFAPPPSQVVKVRKSDE--------PVYIGS-------------------VKPNVGHGE
CH063_1223     NTLSPL--------TARGSCLRPASSPLYLGA-------------------AKANVGHSE
GLRG_08620     TTLSPF--------TARGSSVRPATHPLHIGA-------------------VKSNVGHGE
CH063_0250     TTLSPL--------TARGTSVRPATSPLHIGA-------------------VKSNVGHGE
CS_39          SVFGRH---------------RSSGNPLYIGA-------------------AKANVGHGE
GLRG_03507     STFASG---------------REPTNPLYVGG-------------------IKANVGHGE
GLRG_03509     STFASG---------------REPTNPLYVGG-------------------IKANVGHGE


GLRG_10537     ------------------------------------------------------------
CS_23          AAAGLSGLLKAIMAVERAVIPGNPTFLIPNPNID--------------------------
CH063_1191     ------------------------------------------------------------
CS_16          GASGLSSVIKMVLALENSTIPPNINFK---------------------------------
GLRG_06162     GAAGLSSVIKMVLALENSTIPPNINFK---------------------------------
GLRG_09075     GASGLNSLIKAVLALEHKTIPPNI------------------------------------
CH063_1187     GASGLTSLIKAVLALENNTIPPNIKFSTPNPKIPFEDKKLVVPTKPVAWPADRDRRISVN
CH063_1411     GASGLTSLIKAVLALENNTIPPNIKFSTPNPKIPFEDKKLVVPTKPVAWPADRDRRISVN
CS_5           GAAALTSIIKATLMLERETIVPNANFE---------------------------------
GLRG_11893     GAAALTSIIKATLMLERETIVPNANFE---------------------------------
CS_19          GAAGLTSIIKAVLALENRVIPPNI------------------------------------
CH063_0663     GAAGLTSVIKAVLALENRVIPPNI------------------------------------
GLRG_00918     GAAGLTGLIKAVLSLEHGIIPPNIFFD---------------------------------
CH063_0821     GAAGLTGLIKAVLSLEHNV-----------------------------------------
CS_13          GASALTSIIKAVLSLEHRTILPNI------------------------------------
CH063_0053     AASAITSIVKAVLSLETRTIIPNIKFE---------------------------------
CH063_1006     GASGITSLIKSVLALEHRTIPPNIKFD---------------------------------
GLRG_07171     GASGITSLIKSILALEHRTIPPNIKFD---------------------------------
CS_22          GASGITSLIKSILALEHRTIPPNIKFD---------------------------------
CS_14          GSSGLVSLIKAVMSLEHRTIPPNI------------------------------------
CS_26          GASGLTSLIKTVMALEHRTIPPNIKFT---------------------------------
GLRG_11435     GASGLTSLIKTVMALEHRTIPPNIKFT---------------------------------
GLRG_08640     ------------------------------------------------------------
GLRG_10827     AASGLTALIKAVLIVERGIIPPTIGIK---------------------------------
GLRG_11425     AASGLSAVIKAVLVAERGVIPPTRGI----------------------------------
CH063_0013     AASGLSAVIKAILTVERGIIPPTQGVT---------------------------------
GLRG_08482     GASALASIMKVVLSLENGAIAP--------------------------------------
CH063_1130     GASGLTSIIKVVKSMEEGRIAPS-------------------------------------
GLRG_03511     SSSGIAGMLKAIMVLKKNQIPPQLNL----------------------------------
CS_18          ASSGVAGLLKAILILKHGAIPPNIDF----------------------------------
GLRG_05714     ASSGVAGLLKAILILKHGAIPPNIDF----------------------------------
CH063_0922     ------------------------------------------------------------
CH063_1282     GASGLLSLIKATMMLERGVMLPNANFEKLNPAILEASGADRLRVLKEAMPWESSPSPRR
CS_28          AASGMAQVIKTVLVLENGLIPPNTNFE---------------------------------
GLRG_11563     AASGMAQVIKTVMVLENGVIPPNTNFE---------------------------------
C.P.NRPS_6     GCAGLAGVIKVLLAMKHDTIPPNLHLRSLSAKIQPYQHALKVPT-KATPWPTRVSGQHKR
CH063_0617     GCAGLAGVIKVLLAMKHSTIPPNLHLTSLNPKIAALRHVLRVPT-EPTAWPAPAAGQPKR
C.P.NRPS_1     GTAGLAGVIKTTAALQRKTIFPNKHFNKLNPRIAPFYGNLKVPT-VNVPWPKPPNGQPLR
GLRG_09842     GTAGLAGLLKTTAALQRKTIFPNRHFKRLNPRIEPFYKNLKVPT-VEIPWPKPPENQPLR
GLRG_09715     GAAGLAGLLKASVAIQHGQIPPNLHFNRLNPKIKPFYDHLEVPT-SLQSWPELPKGVPRR
C.P.NRPS_8     GAAGIAGVLKASLALKHAAVPPNLLFERPNPRIEPLMANLRVPT-SLEPWPPTEAGEPRR
GLRG_10367     GAAGIAGVLKASLALKHAAVPPNLLFERPNPRIEPLMANLRVPT-SLEPWPPTEAGEPRR
CS/1           GTAGIASILKASLALQNKQIPPNLWFNNVNPKILPFCKNLRVPTINPKEWAEPLNSQPRR
CH063_0489     GTAGLAGLLKGCLALKHAEIPPNL------------------------------------
```

```
CH063_0392   GTAGIAGILKASLAVQHGEIPPNLLFDKLSPAVAPFYTNLEIAT-EVK------------
GLRG_07434   STAGLAGILKVVQAMQHGFVPPNMLLENLNPNVLPFYRHLQIPQ-ELTPWPQPPPGQSRR
CS/20        GTAGLAALIKASLAMKARRIPPNLHLDRLNPAVEPFYKDLEVPT-KLMDWPQPEPGQPLR
CH063_0306   GTAGLAALIKASLAMKARRIPPNLHLNRVNPAVEPFYKDLEVPT-RLREWPQLEPGQPLR
CH030B       GTAGLAAVVKATEALKAATLPPNRLFNRLSPKVEPFYKNLEVVT-SARPWPKIPANGMRR
CS_30        GTAGLAAVIKASRAVQAGILPPNR------------------------------------
CS_29        GTAGLAGVLKASLAVQHGAIPPNQHFE---------------------------------
CH063_0199   GTAGIAGLLKACLAVKHGVIPPNL------------------------------------
C.P.NRPS_2   GCAGMAGLLKITLAMQHKIIPPNLHFHKLNPSVAPWYRNLQICT-RPQEWPRVAPGHPLR
CH063_0288   GAAGLAGLLKVVHAMINRSIPPNLHLE---------------------------------
CS_4         GAAGLAGLLKVIHSMINKSIPPNLHLE---------------------------------
GLRG_08212   GAAGLAGLLKVVHSMINNSIPPNLHLE---------------------------------
CS_15        GAAGVAGLIKGVLTTEKGRIPPNINFE---------------------------------
CH063_1445   GAAGVAGLIKGVLTTEKGRIPPNINFE---------------------------------
GLRG_02583   GAAGVSGLIKGVLTTEKGRIPPNINFE---------------------------------
CS_32        AASGIAGVIKAVMMLERGFLLPN-------------------------------------
GLRG_03360   AASGIAGVIKAAMMLERGFLLPN-------------------------------------
CH063_1222   AASGIVGVIKAAMMLERGFLLPN-------------------------------------
CH063_0134   ASSGIVGVIKAIMMLKHGVIPPSLNYEQANPNIDMNSLGVQVPTSTREWPKDMPRRISVN
CS_21        AASGLASIVKTALALEKGVIPPSINFERPNPKIPLEDWRLRLVRKLEPWAAAVRRASIN
GLRG_09267   AASGLASIVKTVLALERGVIPPSINFE---------------------------------
CH0190       AASGLAGIIKVALSLEKGLMPPS-------------------------------------
CH063_0873   AASGLASVIKVVCALEAGLVPPNANFE---------------------------------
CS_9         GNSGLASFIKGVLCVERGIIPANAWFE---------------------------------
CH063_1490   ------------------------------------------------------------
GLRG_01860   GAAGVAAVIKGVLTLESGVIPGNIWFE---------------------------------
CS_25        GAAGVAAVIKGVLTLESGVIPGNIWFE---------------------------------
CS_34        PAAGIAGLIKGVLALERGLIPPNIHLTLSTA-----------------------------
GLRG_11770   GASGVAALIKSILALEKGYIPPQ-------------------------------------
CH0399       CAAGIAALIKGVLALEKGVIPPNIHFT---------------------------------
CS_11        SAAGVAGLIKGILAMEHGFIPPNIHFD---------------------------------
GLRG_11840   SAAGVAGLIKGILAMEHGFIPPNIHFD---------------------------------
CH063_0781   PAAGVAGFIKGVLSLEHGMIPPNVHFH---------------------------------
CH063_1357   ------------------------------------------------------------
GLRG_10317   GAAGVAGIIKAAFAVERGLIPQNLWFERGEGVAAIVMKPLSQALADGDSIECVIRETGLN
CH063_1471   GGSGLVQVIKAVLMLEQGKIPPSLYYEKPNPRIPMDDWNLRVPTELLPWPAAGLRRISIN
CH063_0506   GASGLAGIIKTVLVLEHGIIPPICNFE---------------------------------
GLRG_11835   GASGIAAILKTVLILENGVIPPNANFE---------------------------------
CS_3         GASGLAGLLKTVLTLERGVIPPIADFE---------------------------------
GLRG_11878   GASGLAGLLKTVLALERGVIPPIADFE---------------------------------
CH063_0051   GCAGLAGLIKTVLCLEAGVIPPVAGLR---------------------------------
CH0135       GCSGLAGVLKSILCLEKGVLIPTAGIE---------------------------------
CS_43        GASGVVSLIKVLLMMHENCIPPQAS-----------------------------------
GLRG_09268   GASGVISLIKVLLMMQESRIPPQASF----------------------------------
GLRG_11857   CTAGVLSLIKVILMMCNGRIPPQASF----------------------------------
CH063_0213   CASGVVSAVKILLMLQKGMIPPQVNLK---------------------------------
CH063_0477   TAAGVASLLKVLAMLRHRAIPP--------------------------------------
CH063_0313   AASGAAGLIKTILMLRHEQIPPQASF----------------------------------
CS_44        AASGAAGLIKAILMMRHGKIPPQASYR---------------------------------
GLRG_11575   AASGAAGLIKAILMMRHGKIPPQASYRSLNPRIPALEPDGMEIPQVLTDWGPAERIACIN
CH063_1347   AASGVAGVIKTLLMMQHQAIPKQANFFRLNRTIKVLPGDADDVVVPESTLAWTTTTTTER
C.PKS        GASGVAGLIKTILMMQNRAIPRQVHF----------------------------------
CS_45        GASGVTGIIRVLLMMQHRTITPQASFR---------------------------------
GLRG_11798   GASGVAGIIRVLLMMQHRTITPQASFR---------------------------------
GLRG_11778   AGAGAAAVIKAIMMFERNIM----------------------------------------
CH0398       AGSGAAALVKSIMMLQANIIPPQAGL----------------------------------
CS_37        GAAGMASLMKCIKMFETNIVPPQAG-----------------------------------
GLRG_11836   GAAGMASLMKCIKMFETNTIPPQAG-----------------------------------
CH063_1255   ASSGVASVIKAILMFQKQILPPQVGMPHALNKQFPPLDDKKIRILSKAAPFRSASGKPRR
CS_40        AASGIASLIKGVFMFRHQIMPPQVG-----------------------------------
CS_38        AAAGIASFIKVLIMYQKGAVPPQIGVAKLNPTLPPDLEERNIGLNWEYSEWSRSNRAAGR
CS_36        AAAGIASLIKVLLMYRHGTIPRHIGI----------------------------------
```

```
GLRG_08632    AAAGITSLLKVLLSFQNNEIPPHVGIK---------------------------------
CH063_0351    AASGVTALTKVLMMMQKNAIPPHVGIK---------------------------------
CS_35         AASGVTALTKVLMMMQKNAIPPHVGIKKEINKTFPKDLSERNVNIAFHLTPFKRRDGKPR
GLRG_04203    AASGVTALTKVLMMMQKNAIPPHVGIKKEINKTFPKDLSERNVNIAFHLTPFKRRDGKPR
GLRG_11956    AASGVTALIKVLLMMQKNEIPPHCGIK---------------------------------
CH063_1223    AASGVTSIAKVLMMMKHSIIPPHIGVKTEFNRRLPNLAQRNARIAMVPTAWPRPPHGGSR
GLRG_08620    AAAGVTSLAKALMMMKHSTLPPHAGIK---------------------------------
CH063_0250    AAAGVTSLAKVLMMMKHSTIPPHIGIK---------------------------------
CS_39         AAAGITSIIKALLMFQHGKVPPHVGIK---------------------------------
GLRG_03507    AGGNTCMVI---------------------------------------------------
GLRG_03509    AAAGVTSVIKALMMFRENAIPPHAGIKTRINSKFPPLEKANIHIARELKPFMTDKEPRRI


GLRG_10537    ------------------------------------------------------------
CS_23         ------------------------------------------------------------
CH063_1191    ------------------------------------------------------------
CS_16         ------------------------------------------------------------
GLRG_06162    ------------------------------------------------------------
GLRG_09075    ------------------------------------------------------------
CH063_1187    SFGIGGSNAHVILEA---------------------------------------------
CH063_1411    SFGIGGSNAHVILEA---------------------------------------------
CS_5          ------------------------------------------------------------
GLRG_11893    ------------------------------------------------------------
CS_19         ------------------------------------------------------------
CH063_0663    ------------------------------------------------------------
GLRG_00918    ------------------------------------------------------------
CH063_0821    ------------------------------------------------------------
CS_13         ------------------------------------------------------------
CH063_0053    ------------------------------------------------------------
CH063_1006    ------------------------------------------------------------
GLRG_07171    ------------------------------------------------------------
CS_22         ------------------------------------------------------------
CS_14         ------------------------------------------------------------
CS_26         ------------------------------------------------------------
GLRG_11435    ------------------------------------------------------------
GLRG_08640    ------------------------------------------------------------
GLRG_10827    ------------------------------------------------------------
GLRG_11425    ------------------------------------------------------------
CH063_0013    ------------------------------------------------------------
GLRG_08482    ------------------------------------------------------------
CH063_1130    ------------------------------------------------------------
GLRG_03511    ------------------------------------------------------------
CS_18         ------------------------------------------------------------
GLRG_05714    ------------------------------------------------------------
CH063_0922    ------------------------------------------------------------
CH063_1282    VCVTNYGLGGANSAILLEQ-----------------------------------------
CS_28         ------------------------------------------------------------
GLRG_11563    ------------------------------------------------------------
C.P.NRPS_6    ASVN------------SFGFGGTNAHVILEAY----------------------------
CH063_0617    ASVN------------SFGFGGTNAHVILES-----------------------------
C.P.NRPS_1    ASVN------------SFGFGGTNSHVILES-----------------------------
GLRG_09842    ASVN------------SFGFGGTNAHVILES-----------------------------
GLRG_09715    ASIN------------SFGFGGTNAHAIIES-----------------------------
C.P.NRPS_8    ASVNKVFPRVFANDDDSFGFGGTNAHAILES-----------------------------
GLRG_10367    ASVN------------NFGFGGTNAHAILES-----------------------------
CS/1          TSVN------------SFGFGGANAHAILEE-----------------------------
CH063_0489    ------------------------------------------------------------
CH063_0392    ------------------------------------------------------------
GLRG_07434    ASVN------------SFGFGGTNAHIILES-----------------------------
CS/20         ASVN------------SFGFGGANAHAIIES-----------------------------
CH063_0306    ASVN------------SFGFGGANAHAIIES-----------------------------
CH030B        VSVN------------SFGFGGANCHAILES-----------------------------
```

```
CS_30         ------------------------------------------------------------
CS_29         ------------------------------------------------------------
CH063_0199    ------------------------------------------------------------
C.P.NRPS_2    ASIN------------GFGSGGTNVHVIVES-----------------------------
CH063_0288    ------------------------------------------------------------
CS_4          ------------------------------------------------------------
GLRG_08212    ------------------------------------------------------------
CS_15         ------------------------------------------------------------
CH063_1445    ------------------------------------------------------------
GLRG_02583    ------------------------------------------------------------
CS_32         ------------------------------------------------------------
GLRG_03360    ------------------------------------------------------------
CH063_1222    ------------------------------------------------------------
CH063_0134    NYGYGGTNGHVIIVAICSIGL---------------------------------------
CS_21         NFGYGGANSHIIMEE---------------------------------------------
GLRG_09267    ------------------------------------------------------------
CH0190        ------------------------------------------------------------
CH063_0873    ------------------------------------------------------------
CS_9          ------------------------------------------------------------
CH063_1490    ------------------------------------------------------------
GLRG_01860    ------------------------------------------------------------
CS_25         ------------------------------------------------------------
CS_34         ------------------------------------------------------------
GLRG_11770    ------------------------------------------------------------
CH0399        ------------------------------------------------------------
CS_11         ------------------------------------------------------------
GLRG_11840    ------------------------------------------------------------
CH063_0781    ------------------------------------------------------------
CH063_1357    ------------------------------------------------------------
GLRG_10317    QDGKTPGITMPSSSAQAALIRSTYARAGLDLSKPSDRPQFFEAHGTGTPAGDPIEAAAIR
CH063_1471    SFGYGGTNAHCILDD---------------------------------------------
CH063_0506    ------------------------------------------------------------
GLRG_11835    ------------------------------------------------------------
CS_3          ------------------------------------------------------------
GLRG_11878    ------------------------------------------------------------
CH063_0051    ------------------------------------------------------------
CH0135        ------------------------------------------------------------
CS_43         ------------------------------------------------------------
GLRG_09268    ------------------------------------------------------------
GLRG_11857    ------------------------------------------------------------
CH063_0213    ------------------------------------------------------------
CH063_0477    ------------------------------------------------------------
CH063_0313    ------------------------------------------------------------
CS_44         ------------------------------------------------------------
GLRG_11575    NYGAAGSNAAIMIRQ---------------------------------------------
CH063_1347    RASLVNNYGAAGSNAAILVRQ---------------------------------------
C.PKS         ------------------------------------------------------------
CS_45         ------------------------------------------------------------
GLRG_11798    ------------------------------------------------------------
GLRG_11778    ------------------------------------------------------------
CH0398        ------------------------------------------------------------
CS_37         ------------------------------------------------------------
GLRG_11836    ------------------------------------------------------------
CH063_1255    VLINNFDAAGGNTALVLED-----------------------------------------
CS_40         ------------------------------------------------------------
CS_38         LAVVNSFGAHGGNTTVLL------------------------------------------
CS_36         ------------------------------------------------------------
GLRG_08632    ------------------------------------------------------------
CH063_0351    ------------------------------------------------------------
CS_35         RVFVNNFSAAGGNTGLLLED----------------------------------------
GLRG_04203    RIFVNNFSAAGGNTGLLLED----------------------------------------
GLRG_11956    ------------------------------------------------------------
```

```
CH063_1223      RVLLNNFSAAGGNTSIVMED----------------------------------------
GLRG_08620      ------------------------------------------------------------
CH063_0250      ------------------------------------------------------------
CS_39           ------------------------------------------------------------
GLRG_03507      ------------------------------------------------------------
GLRG_03509      MIGNFGATGGNTCMVIEE------------------------------------------


GLRG_10537      -------------------------------------------------------
CS_23           -------------------------------------------------------
CH063_1191      -------------------------------------------------------
CS_16           -------------------------------------------------------
GLRG_06162      -------------------------------------------------------
GLRG_09075      -------------------------------------------------------
CH063_1187      -------------------------------------------------------
CH063_1411      -------------------------------------------------------
CS_5            -------------------------------------------------------
GLRG_11893      -------------------------------------------------------
CS_19           -------------------------------------------------------
CH063_0663      -------------------------------------------------------
GLRG_00918      -------------------------------------------------------
CH063_0821      -------------------------------------------------------
CS_13           -------------------------------------------------------
CH063_0053      -------------------------------------------------------
CH063_1006      -------------------------------------------------------
GLRG_07171      -------------------------------------------------------
CS_22           -------------------------------------------------------
CS_14           -------------------------------------------------------
CS_26           -------------------------------------------------------
GLRG_11435      -------------------------------------------------------
GLRG_08640      -------------------------------------------------------
GLRG_10827      -------------------------------------------------------
GLRG_11425      -------------------------------------------------------
CH063_0013      -------------------------------------------------------
GLRG_08482      -------------------------------------------------------
CH063_1130      -------------------------------------------------------
GLRG_03511      -------------------------------------------------------
CS_18           -------------------------------------------------------
GLRG_05714      -------------------------------------------------------
CH063_0922      -------------------------------------------------------
CH063_1282      -------------------------------------------------------
CS_28           -------------------------------------------------------
GLRG_11563      -------------------------------------------------------
C.P.NRPS_6      -------------------------------------------------------
CH063_0617      -------------------------------------------------------
C.P.NRPS_1      -------------------------------------------------------
GLRG_09842      -------------------------------------------------------
GLRG_09715      -------------------------------------------------------
C.P.NRPS_8      -------------------------------------------------------
GLRG_10367      -------------------------------------------------------
CS/1            -------------------------------------------------------
CH063_0489      -------------------------------------------------------
CH063_0392      -------------------------------------------------------
GLRG_07434      -------------------------------------------------------
CS/20           -------------------------------------------------------
CH063_0306      -------------------------------------------------------
CH030B          -------------------------------------------------------
CS_30           -------------------------------------------------------
CS_29           -------------------------------------------------------
CH063_0199      -------------------------------------------------------
C.P.NRPS_2      -------------------------------------------------------
CH063_0288      -------------------------------------------------------
```

```
CS_4          -------------------------------------------------------
GLRG_08212    -------------------------------------------------------
CS_15         -------------------------------------------------------
CH063_1445    -------------------------------------------------------
GLRG_02583    -------------------------------------------------------
CS_32         -------------------------------------------------------
GLRG_03360    -------------------------------------------------------
CH063_1222    -------------------------------------------------------
CH063_0134    -------------------------------------------------------
CS_21         -------------------------------------------------------
GLRG_09267    -------------------------------------------------------
CH0190        -------------------------------------------------------
CH063_0873    -------------------------------------------------------
CS_9          -------------------------------------------------------
CH063_1490    -------------------------------------------------------
GLRG_01860    -------------------------------------------------------
CS_25         -------------------------------------------------------
CS_34         -------------------------------------------------------
GLRG_11770    -------------------------------------------------------
CH0399        -------------------------------------------------------
CS_11         -------------------------------------------------------
GLRG_11840    -------------------------------------------------------
CH063_0781    -------------------------------------------------------
CH063_1357    -------------------------------------------------------
GLRG_10317    EAFFGADSHFVPRGPDDTLYVGSIKTVIGHTEGTAGLAAVVKATEALKAATLPPN
CH063_1471    -------------------------------------------------------
CH063_0506    -------------------------------------------------------
GLRG_11835    -------------------------------------------------------
CS_3          -------------------------------------------------------
GLRG_11878    -------------------------------------------------------
CH063_0051    -------------------------------------------------------
CH0135        -------------------------------------------------------
CS_43         -------------------------------------------------------
GLRG_09268    -------------------------------------------------------
GLRG_11857    -------------------------------------------------------
CH063_0213    -------------------------------------------------------
CH063_0477    -------------------------------------------------------
CH063_0313    -------------------------------------------------------
CS_44         -------------------------------------------------------
GLRG_11575    -------------------------------------------------------
CH063_1347    -------------------------------------------------------
C.PKS         -------------------------------------------------------
CS_45         -------------------------------------------------------
GLRG_11798    -------------------------------------------------------
GLRG_11778    -------------------------------------------------------
CH0398        -------------------------------------------------------
CS_37         -------------------------------------------------------
GLRG_11836    -------------------------------------------------------
CH063_1255    -------------------------------------------------------
CS_40         -------------------------------------------------------
CS_38         -------------------------------------------------------
CS_36         -------------------------------------------------------
GLRG_08632    -------------------------------------------------------
CH063_0351    -------------------------------------------------------
CS_35         -------------------------------------------------------
GLRG_04203    -------------------------------------------------------
GLRG_11956    -------------------------------------------------------
CH063_1223    -------------------------------------------------------
GLRG_08620    -------------------------------------------------------
CH063_0250    -------------------------------------------------------
CS_39         -------------------------------------------------------
GLRG_03507    -------------------------------------------------------
```

```
GLRG_03509      ------------------------------------------------------------
```

Appendix D - Alignments used to create the rooted phylogenetic trees in chapter 5

PART 1: ITS sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
CgSl1        CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGTAGGGTCCGCC
JG34.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG42.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG7.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS1.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
CBS_129661
CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
E031         CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
FS13.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
FS34.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
FS37.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
FS5.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
FS8.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
GS1.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG102.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG18.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG2.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG75.00      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG88.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
JG9.001      CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
S3.001_NR    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS100.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS107.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS124.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS159.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS171.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS204.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS240.001    CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS26.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS29.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS41.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS45.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS63.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
SS78.001     CTCCGGATCCCGATGCGAGACGTTAGTTACTACGCAAAGGAGGCTCCGGGAGGGTCCGCC
             ************************************************* **********

CgSl1        ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG34.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG42.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG7.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS1.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
CBS_129661   ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
E031         ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
FS13.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
FS34.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
FS37.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
FS5.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
FS8.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
GS1.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG102.001    ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG18.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG2.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG75.00      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG88.001     ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
JG9.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
```

```
S3.001_NR      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS100.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS107.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS124.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS159.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS171.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS204.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS240.001      ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS26.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS29.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS41.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS45.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS63.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
SS78.001       ACTACTTTTGAGGGCCTACGTCAACCGTAGTGCCCCAACACCAAGCGAAGCTTGAGGGTT
               ************************************************************

CgSl1          GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG34.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG42.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG7.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS1.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
CBS_129661
GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
E031           GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
FS13.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
FS34.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
FS37.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
FS5.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
FS8.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
GS1.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG102.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG18.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG2.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG75.00        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG88.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
JG9.001        GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
S3.001_NR      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS100.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS107.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS124.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS159.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS171.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS204.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS240.001      GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS26.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS29.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS41.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS45.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS63.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
SS78.001       GAAATGACGCTCGAACAGGCATGCCCGCCAGAATGCTGGCGGGCGCAATGTGCGTTCAAA
               ************************************************************

CgSl1          GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG34.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG42.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG7.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS1.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
CBS_129661     GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
E031           GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
FS13.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
FS34.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
FS37.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
```

```
FS5.001         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
FS8.001         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
GS1.001         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG102.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG18.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG2.001         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG75.00         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG88.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
JG9.001         GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
S3.001_NR       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS100.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS107.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS124.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS159.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS171.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS204.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS240.001       GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS26.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS29.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS41.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS45.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS63.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
SS78.001        GATTCGATGATTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTT
                ************************************************************

CgSl1           CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG34.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG42.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG7.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS1.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
CBS_129661      CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
E031            CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
FS13.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
FS34.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
FS37.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
FS5.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
FS8.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
GS1.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG102.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG18.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG2.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG75.00         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG88.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
JG9.001         CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
S3.001_NR       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS100.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS107.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS124.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS159.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS171.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS204.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS240.001       CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS26.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS29.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS41.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS45.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS63.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
SS78.001        CATCGATGCCAGAACCAAGAGATCCGTTGTTAAAAGTTTTAATTATTTGCTTGTGCCACT
                ************************************************************

CgSl1           CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG34.001        CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
```

```
JG42.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG7.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS1.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
CBS_129661    CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
E031          CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
FS13.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
FS34.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
FS37.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
FS5.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
FS8.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
GS1.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG102.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG18.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG2.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG75.00       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG88.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
JG9.001       CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
S3.001_NR     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS100.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS107.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS124.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS159.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS171.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS204.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS240.001     CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS26.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS29.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS41.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS45.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS63.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
SS78.001      CAGAAGAAACGTCGTTAAATCAGAGTTTGGTTATCCTCCGGCGGGCGCCGCGACCCCTCC
              ************************************************************

CgSl1         CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG34.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGAGGGACCCCCTAACCCGCCGAAGC
JG42.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGAGGGACCCCCTAACCCGCCGAAGC
JG7.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGAGGGACCCCCTAACCCGCCGAAGC
SS1.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGAGGGACCCCCTAACCCGCCGAAGC
CBS_129661
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
E031          CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
FS13.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
FS34.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
FS37.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
FS5.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
FS8.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
GS1.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG102.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG18.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG2.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG75.00       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG88.001      CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
JG9.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
S3.001_NR
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS100.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS107.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS124.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
```

```
SS159.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS171.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS204.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS240.001
CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS26.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS29.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS41.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS45.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS63.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
SS78.001       CGGAGGAGGAGGGCGCGGCCGGGAGGGCGTCCCCGGGGGGACCCCCTAACCCGCCGAAGC
          ************************************ ***********************

CgSl1          AACAGTTAGGTATGTTCA
JG34.001       AACAGTTAGGTATGTTCA
JG42.001       AACAGTTAGGTATGTTCA
JG7.001        AACAGTTAGGTATGTTCA
SS1.001        AACAGTTAGGTATGTTCA
CBS_129661     AACAGTTAGGTATGTTCA
E031           AACAGTTAGGTATGTTCA
FS13.001       AACAGTTAGGTATGTTCA
FS34.001       AACAGTTAGGTATGTTCA
FS37.001       AACAGTTAGGTATGTTCA
FS5.001        AACAGTTAGGTATGTTCA
FS8.001        AACAGTTAGGTATGTTCA
GS1.001        AACAGTTAGGTATGTTCA
JG102.001      AACAGTTAGGTATGTTCA
JG18.001       AACAGTTAGGTATGTTCA
JG2.001        AACAGTTAGGTATGTTCA
JG75.00        AACAGTTAGGTATGTTCA
JG88.001       AACAGTTAGGTATGTTCA
JG9.001        AACAGTTAGGTATGTTCA
S3.001_NR      AACAGTTAGGTATGTTCA
SS100.001      AACAGTTAGGTATGTTCA
SS107.001      AACAGTTAGGTATGTTCA
SS124.001      AACAGTTAGGTATGTTCA
SS159.001      AACAGTTAGGTATGTTCA
SS171.001      AACAGTTAGGTATGTTCA
SS204.001      AACAGTTAGGTATGTTCA
SS240.001      AACAGTTAGGTATGTTCA
SS26.001       AACAGTTAGGTATGTTCA
SS29.001       AACAGTTAGGTATGTTCA
SS41.001       AACAGTTAGGTATGTTCA
SS45.001       AACAGTTAGGTATGTTCA
SS63.001       AACAGTTAGGTATGTTCA
SS78.001       AACAGTTAGGTATGTTCA
          ******************
```

PART 2: Concatenated sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
CBS129661      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
S3.001_CBS     TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
CgSl1          TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
GS1.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS63.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
E031           TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS13.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS34.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS37.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS5.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG88.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS100.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS107.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS124.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS159.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS171.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS204.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS240.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS26.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS29.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS41.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS45.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS78.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS8.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG34.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG7.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS1.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG42.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG9.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG75.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG102.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG18.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG2.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
               ************************************************************

CBS129661      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
S3.001_CBS     ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
CgSl1          ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
GS1.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS63.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
E031           ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS13.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS34.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS37.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS5.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG88.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS100.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS107.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS124.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS159.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS171.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS204.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS240.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS26.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS29.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS41.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS45.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS78.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
```

```
FS8.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG34.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG7.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS1.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG42.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG9.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG75.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG102.001     ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG18.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG2.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
              ************************************************************

CBS129661     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAA-----------------
S3.001_CBS    GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
CgSl1         GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
GS1.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS63.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
E031          GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS13.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS34.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS37.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS5.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG88.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS100.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS107.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS124.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS159.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS171.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS204.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS240.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS26.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS29.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS41.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS45.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS78.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS8.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG34.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG7.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS1.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG42.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG9.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG75.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG102.001     GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG18.001      GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG2.001       GTCCCCGGCTGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
              *******************************************

CBS129661     -----------------------------------AGGATACTCTGGTGTCGCAATATAC
S3.001_CBS    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
CgSl1         CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
GS1.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS63.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
E031          CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS13.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS34.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS37.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS5.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG88.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS100.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS107.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS124.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS159.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
```

```
SS171.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS204.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS240.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS26.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS29.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS41.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS45.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS78.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS8.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG34.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG7.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS1.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG42.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG9.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG75.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG102.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG18.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG2.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
                                 ***************** *******

CBS129661     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
S3.001_CBS    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
CgSl1         ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
GS1.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS63.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
E031          ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS13.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS34.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS37.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS5.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG88.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS100.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS107.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS124.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS159.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS171.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS204.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS240.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS26.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS29.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS41.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS45.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS78.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS8.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG34.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG7.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS1.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG42.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG9.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG75.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG102.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG18.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG2.001       ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
              ************************************************************

CBS129661     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
S3.001_CBS    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
CgSl1         GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
GS1.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS63.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
E031          GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS13.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
```

```
FS34.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS37.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS5.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG88.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS100.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS107.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS124.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS159.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS171.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS204.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS240.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS26.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS29.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS41.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS45.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS78.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS8.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG34.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG7.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS1.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG42.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG9.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG75.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG102.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG18.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG2.001       GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
              ************************************************************

CBS129661     CCGCGACCTGAGCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
S3.001_CBS    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
CgSl1         CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
GS1.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS63.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
E031          CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS13.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS34.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS37.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS5.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG88.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS100.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS107.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS124.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS159.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS171.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS204.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS240.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS26.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS29.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS41.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS45.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS78.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS8.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG34.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG7.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS1.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG42.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG9.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG75.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG102.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG18.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG2.001       CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
              *********** ************************************************
```

```
CBS129661    TGCGTAGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
S3.001_CBS   TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
CgSl1        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
GS1.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS63.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
E031         TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS13.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS34.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS37.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS5.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG88.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS100.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS107.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS124.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS159.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS171.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS204.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS240.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS26.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS29.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS41.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS45.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS78.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS8.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG34.001     TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG7.001      TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS1.001      TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG42.001     TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG9.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG75.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG102.001    TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG18.001     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG2.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
             ***** ***** ************************************************

CBS129661    GACTCCTCCCGCGACGATTTCCGGCTCGGGTATCTCAACGCTCTCGACGCGAGGGTGCGG
S3.001_CBS   GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
CgSl1        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
GS1.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS63.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
E031         GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS13.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS34.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS37.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS5.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG88.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS100.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS107.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS124.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS159.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS171.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS204.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS240.001    GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS26.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS29.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS41.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS45.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS78.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS8.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG34.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG7.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
```

```
SS1.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG42.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG9.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG75.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG102.001     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG18.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG2.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
              ******************************** ***************************

CBS129661     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
S3.001_CBS    AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
CgSl1         AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
GS1.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS63.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
E031          AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS13.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS34.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS37.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS5.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
JG88.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS100.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS107.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS124.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS159.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS171.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS204.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS240.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS26.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS29.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS41.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS45.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS78.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS8.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
JG34.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG7.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
SS1.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG42.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG9.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG75.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG102.001     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG18.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG2.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
              ***********************************************************

CBS129661
GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGATATGTCGGTAGAGGAG
S3.001_CBS
GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
CgSl1         GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
GS1.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS63.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
E031          GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS13.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS34.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS37.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS5.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG88.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS100.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS107.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS124.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS159.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS171.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
```

```
SS204.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS240.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS26.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS29.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS41.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS45.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS78.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS8.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG34.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG7.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS1.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG42.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG9.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG75.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG102.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG18.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG2.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
               ******************************************* ****************

CBS129661      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCT
S3.001_CBS     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
CgSl1          TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
GS1.001        TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS63.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
E031           TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS13.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS34.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS37.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS5.001        TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG88.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS100.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS107.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS124.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS159.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS171.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS204.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS240.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS26.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS29.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS41.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS45.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS78.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS8.001        TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG34.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG7.001        TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS1.001        TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG42.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG9.001        TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG75.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG102.001      TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG18.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG2.001        TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
               ********** *************************************************

CBS129661
GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
S3.001_CBS
GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
CgSl1          GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
GS1.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS63.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
E031           GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
```

```
FS13.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS34.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS37.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS5.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG88.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS100.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS107.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS124.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS159.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS171.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS204.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS240.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS26.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS29.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS41.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS45.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS78.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS8.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG34.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG7.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS1.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG42.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG9.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG75.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG102.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG18.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG2.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
              ************************************************************

CBS129661
CCCACAACCGCGGTTCCACAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
S3.001_CBS
CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
CgSl1         CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
GS1.001       CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS63.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
E031          CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
FS13.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
FS34.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
FS37.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
FS5.001       CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
JG88.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS100.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS107.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS124.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS159.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS171.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS204.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS240.001     CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS26.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS29.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS41.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS45.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
SS78.001      CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
FS8.001       CCCACAACCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGG
JG34.001      CCCACAACCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG7.001       CCCACAACCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
SS1.001       CCCACAACCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG42.001      CCCACAACCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG9.001       CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG75.001      CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG102.001     CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
```

```
JG18.001       CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
JG2.001        CCCACAACCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGG
               ********* ******** ******* ******* *************************

CBS129661      GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
S3.001_CBS     GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
CgSl1          GACAAGAAATGGTAGGTCAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
GS1.001        GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS63.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
E031           GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
FS13.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
FS34.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
FS37.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
FS5.001        GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
JG88.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS100.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS107.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS124.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS159.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS171.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS204.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS240.001      GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS26.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS29.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS41.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS45.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
SS78.001       GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
FS8.001        GA---------------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAAC
JG34.001       GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG7.001        GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
SS1.001        GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG42.001       GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG9.001        GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG75.001       GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG102.001      GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG18.001       GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
JG2.001        GA---------------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAAC
               **               *********************** *******************

CBS129661      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
S3.001_CBS     ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
CgSl1          ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
GS1.001        ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS63.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
E031           ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
FS13.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
FS34.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
FS37.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
FS5.001        ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG88.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS100.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS107.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS124.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS159.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS171.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS204.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS240.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS26.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS29.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS41.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS45.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS78.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
```

```
FS8.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG34.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG7.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
SS1.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG42.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG9.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG75.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG102.001     ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG18.001      ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
JG2.001       ATACCACAGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCG
              ************************************************************

CBS129661     CACCTTGGCGCGACCCTCTTCCCCAAGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
S3.001_CBS    CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
CgSl1         CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
GS1.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS63.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
E031          CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
FS13.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
FS34.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
FS37.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
FS5.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG88.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS100.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS107.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS124.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS159.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS171.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS204.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS240.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS26.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS29.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS41.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS45.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS78.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
FS8.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG34.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG7.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
SS1.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG42.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG9.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG75.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG102.001     CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG18.001      CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
JG2.001       CACCTTGGCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATG
              *********************** * **********************************

CBS129661     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
S3.001_CBS    GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
CgSl1         GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
GS1.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS63.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
E031          GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
FS13.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
FS34.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
FS37.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
FS5.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG88.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS100.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS107.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS124.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS159.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
```

```
SS171.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS204.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS240.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS26.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS29.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS41.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS45.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS78.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
FS8.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG34.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG7.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
SS1.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG42.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG9.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG75.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG102.001     GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG18.001      GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
JG2.001       GGATCAAAGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCT
              ************************************************************

CBS129661     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
S3.001_CBS    GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
CgSl1         GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
GS1.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS63.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
E031          GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
FS13.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
FS34.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
FS37.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
FS5.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG88.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS100.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS107.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS124.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS159.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS171.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS204.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS240.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS26.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS29.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS41.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS45.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS78.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
FS8.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG34.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG7.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
SS1.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG42.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG9.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG75.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG102.001     GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG18.001      GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
JG2.001       GCCAAGGGGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCG
              ************************************************************

CBS129661
CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
S3.001_CBS
CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
CgSl1         CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
GS1.001       CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS63.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
```

```
E031         CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
FS13.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
FS34.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
FS37.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
FS5.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG88.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS100.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS107.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS124.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS159.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS171.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS204.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS240.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS26.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS29.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS41.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS45.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS78.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
FS8.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG34.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG7.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
SS1.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG42.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG9.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG75.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG102.001    CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG18.001     CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
JG2.001      CCTGCGAGAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCT
             ************************************************************

CBS129661    GCTGCGCCTTTGTTCTTTGTGGGCAACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
S3.001_CBS   GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
CgSl1        GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
GS1.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS63.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
E031         GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
FS13.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
FS34.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
FS37.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
FS5.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
JG88.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS100.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS107.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS124.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS159.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS171.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS204.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS240.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS26.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS29.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS41.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS45.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
SS78.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
FS8.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
JG34.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGT
JG7.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGT
SS1.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGT
JG42.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGT
JG9.001      GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
JG75.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
JG102.001    GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
JG18.001     GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
```

```
JG2.001         GCTGTGCCTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGT
                **** ******************* ************************** ********

CBS129661       GGCGGCTTCGAGGACGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
S3.001_CBS      GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
CgSl1           GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
GS1.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS63.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
E031            GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
FS13.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
FS34.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
FS37.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
FS5.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG88.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS100.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS107.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS124.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS159.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS171.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS204.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS240.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS26.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS29.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS41.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS45.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS78.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
FS8.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG34.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG7.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
SS1.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG42.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG9.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG75.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG102.001       GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG18.001        GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
JG2.001         GGCGGCTTCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGAT
                ************* **********************************************

CBS129661       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
S3.001_CBS      AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
CgSl1           AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
GS1.001         AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS63.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
E031            AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
FS13.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
FS34.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
FS37.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
FS5.001         AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG88.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS100.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS107.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS124.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS159.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS171.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS204.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS240.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS26.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS29.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS41.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS45.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS78.001        AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
FS8.001         AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
```

```
JG34.001      AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG7.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
SS1.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG42.001      AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG9.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG75.001      AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG102.001     AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG18.001      AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
JG2.001       AAAGGATCTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTG
              ************************************************************

CBS129661     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
S3.001_CBS    TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
CgSl1         TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
GS1.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS63.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
E031          TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
FS13.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
FS34.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
FS37.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
FS5.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG88.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS100.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS107.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS124.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS159.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS171.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS204.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS240.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS26.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS29.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS41.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS45.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS78.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
FS8.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG34.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG7.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
SS1.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG42.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG9.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG75.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG102.001     TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG18.001      TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
JG2.001       TCGGATGCTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGAT
              ************************************************************

CBS129661     TGCGTAGACGCTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCGGTCC
S3.001_CBS    TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
CgSl1         TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
GS1.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS63.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
E031          TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
FS13.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
FS34.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
FS37.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
FS5.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
JG88.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS100.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS107.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS124.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS159.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS171.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
```

```
SS204.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS240.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS26.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS29.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS41.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS45.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
SS78.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
FS8.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCC
JG34.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG7.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
SS1.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG42.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG9.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG75.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG102.001     TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG18.001      TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
JG2.001       TGCGTAGACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCC
              ********* ****************************** ************** ****

CBS129661     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
S3.001_CBS    TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
CgSl1         TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
GS1.001       TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS63.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
E031          TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
FS13.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
FS34.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
FS37.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
FS5.001       TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG88.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS100.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS107.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS124.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS159.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS171.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS204.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS240.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS26.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS29.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS41.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS45.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS78.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
FS8.001       TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCC-AT
JG34.001      TCTTTCTCTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG7.001       TCTTTCTCTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
SS1.001       TCTTTCTCTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG42.001      TCTTTCTCTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCCAT
JG9.001       TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCCAT
JG75.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG102.001     TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG18.001      TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
JG2.001       TCTTTCTCTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAAT
              ******************* ************************************  **

CBS129661     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
S3.001_CBS    GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
CgSl1         GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
GS1.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS63.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
E031          GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
FS13.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
FS34.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
```

```
FS37.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
FS5.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG88.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS100.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS107.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS124.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS159.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS171.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS204.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS240.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS26.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS29.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS41.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS45.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS78.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
FS8.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG34.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG7.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
SS1.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG42.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG9.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG75.001      GGG--ATTGTC-GCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG102.001     GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG18.001      GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
JG2.001       GGGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGG
              ***  ****** ************************************************

CBS129661     TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
S3.001_CBS    TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
CgSl1         TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
GS1.001       TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS63.001      TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
E031          TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
FS13.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
FS34.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
FS37.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
FS5.001       TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG88.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS100.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS107.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS124.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS159.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS171.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS204.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS240.001     TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS26.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS29.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS41.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS45.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS78.001      TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
FS8.001       TAGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG34.001      TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG7.001       TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
SS1.001       TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG42.001      TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG9.001       TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG75.001      TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG102.001     TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG18.001      TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
JG2.001       TAGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCT
              ***************************** ******************************
```

```
CBS129661    TGGTCTTTCGTCATGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
S3.001_CBS   TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
CgSl1        TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
GS1.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS63.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
E031         TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
FS13.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
FS34.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
FS37.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
FS5.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG88.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS100.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS107.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS124.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS159.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS171.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS204.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS240.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS26.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS29.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS41.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS45.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS78.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
FS8.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG34.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG7.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
SS1.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG42.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG9.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG75.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG102.001    TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG18.001     TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
JG2.001      TGGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAG
             ************ ***********************************************

CBS129661    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
S3.001_CBS   CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
CgSl1        CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
GS1.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS63.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
E031         CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
FS13.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
FS34.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
FS37.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
FS5.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG88.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS100.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS107.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS124.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS159.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS171.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS204.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS240.001    CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS26.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS29.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS41.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS45.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS78.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
FS8.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG34.001     CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG7.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
SS1.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
```

```
JG42.001       CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG9.001        CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG75.001       CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG102.001      CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG18.001       CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
JG2.001        CCGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCG
               ************************************************************

CBS129661
CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
S3.001_CBS
CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
CgSl1          CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
GS1.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS63.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
E031           CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
FS13.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
FS34.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
FS37.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
FS5.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG88.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS100.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS107.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS124.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS159.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS171.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS204.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS240.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS26.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS29.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS41.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS45.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS78.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
FS8.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG34.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG7.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
SS1.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG42.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG9.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG75.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG102.001      CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG18.001       CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
JG2.001        CCCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGG
               ************************************************************

CBS129661      CTCGATGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCGCCGTTG
S3.001_CBS     CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
CgSl1          CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
GS1.001        CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS63.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
E031           CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
FS13.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
FS34.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
FS37.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
FS5.001        CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG88.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS100.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS107.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS124.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS159.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS171.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS204.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
```

```
SS240.001     CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS26.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS29.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS41.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS45.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS78.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
FS8.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG34.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG7.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
SS1.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG42.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG9.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG75.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG102.001     CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG18.001      CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
JG2.001       CTCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTG
              ***** *********************************************** ** ***

CBS129661     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
S3.001_CBS    AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
CgSl1         AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
GS1.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS63.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
E031          AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
FS13.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
FS34.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
FS37.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
FS5.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG88.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS100.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS107.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS124.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS159.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS171.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS204.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS240.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS26.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS29.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS41.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS45.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS78.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
FS8.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG34.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG7.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
SS1.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG42.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG9.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG75.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG102.001     AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG18.001      AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
JG2.001       AAGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCG
              ************************************************************

CBS129661     ATGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
S3.001_CBS    ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
CgSl1         ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
GS1.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS63.001      ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
E031          ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS13.001      ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS34.001      ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS37.001      ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
```

```
FS5.001         ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG88.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS100.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS107.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS124.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS159.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS171.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS204.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS240.001       ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS26.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS29.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS41.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS45.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS78.001        ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS8.001         ATGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG34.001        ATGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG7.001         ATGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS1.001         ATGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG42.001        ATGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG9.001         ATGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG75.001        ATGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG102.001       ATGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG18.001        ATGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG2.001         ATGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
                ****  ********************************
```

PART 3: APN2 sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
CBS129661      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG34.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG42.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG7.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS1.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG102.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG18.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG2.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG75.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG9.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
CgSl1          TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
E031           TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS13.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS34.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS37.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS5.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
FS8.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
GS1.001        TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
JG88.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS100.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS107.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS124.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS159.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS171.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS204.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS240.001      TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS26.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS29.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS41.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS45.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS63.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
SS78.001       TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
S3.001_CBS     TCGTTTGGTCATCGCTTCTGATCAGCTCGCCAGGCAATGTTCGAAACCTTGGAGGCGGAT
               ************************************************************

CBS129661      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG34.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG42.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG7.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS1.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG102.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG18.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG2.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG75.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG9.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
CgSl1          ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
E031           ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS13.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS34.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS37.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS5.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
FS8.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
GS1.001        ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
JG88.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS100.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS107.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS124.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS159.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
```

```
SS171.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS204.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS240.001      ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS26.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS29.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS41.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS45.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS63.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
SS78.001       ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
S3.001_CBS     ATTGTGGTTATGCAAGAAACCAAGATTCAGAGAAAAGATTTGCGTGATGACATGGTTCTC
               ************************************************************

CBS129661      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAA----------------
JG34.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG42.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG7.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS1.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG102.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG18.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG2.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG75.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG9.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
CgSl1          GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
E031           GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS13.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS34.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS37.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS5.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
FS8.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
GS1.001        GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
JG88.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS100.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS107.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS124.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS159.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS171.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS204.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS240.001      GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS26.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS29.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS41.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS45.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS63.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
SS78.001       GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
S3.001_CBS     GTCCCCGGCTGGGACGTTGTTTTCAGCCTACCAAAACACAAGAAAGGTGAGCCTATCCGA
               ********************************************

CBS129661      -----------------------------------AGGATACTCTGGTGTCGCAATATAC
JG34.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG42.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG7.001        CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS1.001        CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG102.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG18.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG2.001        CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG75.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG9.001        CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
CgSl1          CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
E031           CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS13.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS34.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS37.001       CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
```

```
FS5.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
FS8.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
GS1.001      CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
JG88.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS100.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS107.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS124.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS159.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS171.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS204.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS240.001    CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS26.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS29.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS41.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS45.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS63.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
SS78.001     CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
S3.001_CBS   CTTTGAACGGGTGGGTGTGACCTGACTCACCGGACAGGATACTCTGGTGTCGTAATATAC
                         ***************** *******

CBS129661    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG34.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG42.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG7.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS1.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG102.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG18.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG2.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG75.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG9.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
CgSl1        ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
E031         ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS13.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS34.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS37.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS5.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
FS8.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
GS1.001      ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
JG88.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS100.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS107.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS124.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS159.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS171.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS204.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS240.001    ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS26.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS29.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS41.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS45.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS63.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
SS78.001     ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
S3.001_CBS   ACACGCAACGCGACATGCTGCCCAATCCGCGCCGAGGAAGGCATTACGGGCATTCTCCCG
             ************************************************************

CBS129661    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG34.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG42.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG7.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS1.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG102.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG18.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
```

```
JG2.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG75.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG9.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
CgSl1        GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
E031         GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS13.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS34.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS37.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS5.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
FS8.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
GS1.001      GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
JG88.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS100.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS107.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS124.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS159.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS171.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS204.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS240.001    GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS26.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS29.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS41.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS45.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS63.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
SS78.001     GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
S3.001_CBS   GCCCCCAATTCCTCCACTAGCTTTCGCGACCTGCCCGAGACCCAGCAGATTGGCGGCTAC
             ************************************************************

CBS129661    CCGCGACCTGAGCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG34.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG42.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG7.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS1.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG102.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG18.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG2.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG75.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG9.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
CgSl1        CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
E031         CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS13.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS34.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS37.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS5.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
FS8.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
GS1.001      CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
JG88.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS100.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS107.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS124.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS159.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS171.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS204.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS240.001    CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS26.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS29.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS41.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS45.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS63.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
SS78.001     CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
S3.001_CBS   CCGCGACCTGACCAACTACTTGGTGAAGTGGACGAGACAACGCTTGACTCCGAGGGCCGC
             *********** ************************************************
```

```
CBS129661      TGCGTAGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG34.001       TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG42.001       TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG7.001        TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS1.001        TGCGTGGTGCTTGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG102.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG18.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG2.001        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG75.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG9.001        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
CgSl1          TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
E031           TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS13.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS34.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS37.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS5.001        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
FS8.001        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
GS1.001        TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
JG88.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS100.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS107.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS124.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS159.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS171.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS204.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS240.001      TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS26.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS29.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS41.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS45.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS63.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
SS78.001       TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
S3.001_CBS     TGCGTGGTGCTCGAATTCCCGGCCTTTGTCCTGATCGGAACCTACAGCCCGGCCACGCGC
               ***** ***** ************************************************

CBS129661      GACTCCTCCCGCGACGATTTCCGGCTCGGGTATCTCAACGCTCTCGACGCGAGGGTGCGG
JG34.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG42.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG7.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS1.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG102.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG18.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG2.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG75.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG9.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
CgSl1          GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
E031           GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS13.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS34.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS37.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS5.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
FS8.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
GS1.001        GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
JG88.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS100.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS107.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS124.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS159.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS171.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS204.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS240.001      GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
```

```
SS26.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS29.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS41.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS45.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS63.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
SS78.001       GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
S3.001_CBS     GACTCCTCCCGCGACGATTTCCGGCTCGGGTACCTCAACGCTCTCGACGCGAGGGTGCGG
               ******************************** ***************************

CBS129661      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG34.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG42.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG7.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
SS1.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG102.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG18.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG2.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG75.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
JG9.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
CgSl1          AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
E031           AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS13.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS34.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS37.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS5.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
FS8.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
GS1.001        AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
JG88.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS100.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS107.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS124.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS159.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS171.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS204.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS240.001      AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS26.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS29.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS41.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS45.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS63.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
SS78.001       AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAA
S3.001_CBS     AACCTCGTGGCGGCGGGGAAACAGGTCATCCTAACTGGAGATTTGAATGTGATTCGTGAC
               ***********************************************************

CBS129661
GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGATATGTCGGTAGAGGAG
JG34.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG42.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG7.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS1.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG102.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG18.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG2.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG75.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG9.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
CgSl1          GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
E031           GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS13.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS34.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS37.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS5.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
FS8.001        GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
```

```
GS1.001       GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
JG88.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS100.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS107.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS124.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS159.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS171.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS204.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS240.001     GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS26.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS29.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS41.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS45.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS63.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
SS78.001      GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
S3.001_CBS
GAGAGGGACACATGTAATATCAGGGAGGCACTGCGTAAAGAAGGTATGTCGGTAGAGGAG
              ******************************************* ****************

CBS129661     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCT
JG34.001      TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG42.001      TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG7.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS1.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG102.001     TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG18.001      TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG2.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG75.001      TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG9.001       TGGATGAGGACGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
CgSl1         TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
E031          TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS13.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS34.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS37.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS5.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
FS8.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
GS1.001       TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
JG88.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS100.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS107.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS124.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS159.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS171.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS204.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS240.001     TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS26.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS29.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS41.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS45.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS63.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
SS78.001      TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
S3.001_CBS    TGGATGAGGATGCCAAGTCGGCGGCTGTTCAACCATCTTGTCTTTGATGGACAAGTCGCC
              ********** ************************************************

CBS129661
GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG34.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG42.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG7.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS1.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG102.001     GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG18.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
```

```
JG2.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG75.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG9.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
CgSl1          GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
E031           GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS13.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS34.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS37.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS5.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
FS8.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
GS1.001        GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
JG88.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS100.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS107.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS124.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS159.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS171.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS204.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS240.001      GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS26.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS29.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS41.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS45.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS63.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
SS78.001       GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
S3.001_CBS
GGTGAAAGGGATGAAGGACGAGAGACACCAGTGCTTTACGACCTGACGCGAATATTCCAC
               ************************************************************

CBS129661      CCCACAA
JG34.001       CCCACAA
JG42.001       CCCACAA
JG7.001        CCCACAA
SS1.001        CCCACAA
JG102.001      CCCACAA
JG18.001       CCCACAA
JG2.001        CCCACAA
JG75.001       CCCACAA
JG9.001        CCCACAA
CgSl1          CCCACAA
E031           CCCACAA
FS13.001       CCCACAA
FS34.001       CCCACAA
FS37.001       CCCACAA
FS5.001        CCCACAA
FS8.001        CCCACAA
GS1.001        CCCACAA
JG88.001       CCCACAA
SS100.001      CCCACAA
SS107.001      CCCACAA
SS124.001      CCCACAA
SS159.001      CCCACAA
SS171.001      CCCACAA
SS204.001      CCCACAA
SS240.001      CCCACAA
SS26.001       CCCACAA
SS29.001       CCCACAA
SS41.001       CCCACAA
SS45.001       CCCACAA
SS63.001       CCCACAA
SS78.001       CCCACAA
S3.001_CBS     CCCACAA
```

****

PART 4: MAT1/APN2sequences

```
CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment
CBS129661       CCGCGGTTCCACAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG34.001        CCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG7.001         CCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
SS1.001         CCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG42.001        CCGCGGTTCCAAAGGATCTAAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG9.001         CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
CgSl1           CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGACAAGA
JG102.001       CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG18.001        CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG2.001         CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
JG75.001        CCGCGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
S3.001_CBS      CCACGGTTCCAAAGGATCTGAAAGGGCGTGAACACCCATGACGGCTCCAGAGGGA-----
E031            CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
FS13.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
FS34.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
FS37.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
FS5.001         CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
GS1.001         CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
JG88.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS100.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS107.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS124.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS159.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS171.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS204.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS240.001       CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS26.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS29.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS41.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS45.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS63.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
SS78.001        CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
FS8.001         CCACGGTTCCAAAGGATCTGAAAGGGCTTGAACACCCATGACGGCTCCAGAGGGA-----
                ** ******** ******* ******* ***************************

CBS129661       ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG34.001        ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG7.001         ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
SS1.001         ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG42.001        ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG9.001         ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
CgSl1           AATGGTAGGTCAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG102.001       ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG18.001        ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG2.001         ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
JG75.001        ----------CAAGAAATGGTAGGTGAGAACGGGGGAGAGAGGGGAGAGCAACATACCAC
S3.001_CBS      ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
E031            ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
FS13.001        ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
FS34.001        ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
FS37.001        ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
FS5.001         ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
GS1.001         ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
JG88.001        ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS100.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS107.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS124.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS159.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
```

```
SS171.001      ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS204.001      ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS240.001      ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS26.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS29.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS41.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS45.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS63.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
SS78.001       ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
FS8.001        ----------CAAGAAATGGTAGGTGAGAACGGTGGAGAGAGGGGAGAGCAACATACCAC
                         *********************** **************************

CBS129661      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG34.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG7.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS1.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG42.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG9.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
CgSl1          AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG102.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG18.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG2.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG75.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
S3.001_CBS     AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
E031           AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
FS13.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
FS34.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
FS37.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
FS5.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
GS1.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
JG88.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS100.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS107.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS124.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS159.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS171.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS204.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS240.001      AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS26.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS29.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS41.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS45.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS63.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
SS78.001       AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
FS8.001        AGTTGACACCCGGCTTCTTGGTCTGGAGGATGATGCAGGGCTCCCCATGCTCGCACCTTG
               ************************************************************

CBS129661      GCGCGACCCTCTTCCCCAAGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG34.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG7.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS1.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG42.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG9.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
CgSl1          GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG102.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG18.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG2.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG75.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
S3.001_CBS     GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
E031           GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
FS13.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
FS34.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
```

```
FS37.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
FS5.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
GS1.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
JG88.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS100.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS107.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS124.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS159.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS171.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS204.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS240.001      GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS26.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS29.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS41.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS45.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS63.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
SS78.001       GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
FS8.001        GCGCGACCCTCTTCCCAAGGTTGAACTTGGACCAAGACTCCTTGACCTCGATGGGATCAA
               **************** * *****************************************

CBS129661      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG34.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG7.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS1.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG42.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG9.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
CgSl1          AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG102.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG18.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG2.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG75.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
S3.001_CBS     AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
E031           AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
FS13.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
FS34.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
FS37.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
FS5.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
GS1.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
JG88.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS100.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS107.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS124.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS159.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS171.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS204.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS240.001      AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS26.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS29.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS41.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS45.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS63.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
SS78.001       AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
FS8.001        AGACTCTTTCCGAGGTATCGATTCCAGGCACCGTAGTGGTGGCGGAAGAGCCTGCCAAGG
               ************************************************************

CBS129661      GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG34.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG7.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS1.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG42.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG9.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
CgSl1          GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
```

```
JG102.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG18.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG2.001         GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG75.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
S3.001_CBS      GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
E031            GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
FS13.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
FS34.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
FS37.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
FS5.001         GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
GS1.001         GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
JG88.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS100.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS107.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS124.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS159.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS171.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS204.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS240.001       GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS26.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS29.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS41.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS45.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS63.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
SS78.001        GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
FS8.001         GGGACGGAGAAGGTTCTATTGGCGGCGGAGCGTCTACTGTTGCGCCCGCTGCGCCTGCGA
                ************************************************************

CBS129661
GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGCGC
JG34.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG7.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS1.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG42.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG9.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
CgSl1           GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG102.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG18.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG2.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG75.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
S3.001_CBS
GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
E031            GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
FS13.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
FS34.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
FS37.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
FS5.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
GS1.001         GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
JG88.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS100.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS107.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS124.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS159.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS171.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS204.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS240.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS26.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS29.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS41.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS45.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS63.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
SS78.001        GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGGCCACCGCCTGCGGCTGCTGTGC
```

```
FS8.001       GAGGAGGAGGGGTCTTTGATACCCCCGAGAGAGGTGGCCACCGCCTGCGGCTGCTGTGC
              ********************************************************* **

CBS129661     CTTTGTTCTTTGTGGGCAACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG34.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGTGGCGGCT
JG7.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGTGGCGGCT
SS1.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGTGGCGGCT
JG42.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCCCGACGGGTGGCGGCT
JG9.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
CgSl1         CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG102.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG18.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG2.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG75.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
S3.001_CBS    CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
E031          CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
FS13.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
FS34.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
FS37.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
FS5.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
GS1.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
JG88.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS100.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS107.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS124.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS159.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS171.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS204.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS240.001     CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS26.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS29.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS41.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS45.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS63.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
SS78.001      CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
FS8.001       CTTTGTTCTTTGTGGGCGACGGCTGGACTACCGCCGTATCGGCCTCGACGGGTGGCGGCT
              ***************** ************************** ***************

CBS129661     TCGAGGACGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG34.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG7.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS1.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG42.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG9.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
CgSl1         TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG102.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG18.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG2.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG75.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
S3.001_CBS    TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
E031          TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
FS13.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
FS34.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
FS37.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
FS5.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
GS1.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
JG88.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS100.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS107.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS124.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS159.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS171.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
```

```
SS204.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS240.001     TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS26.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS29.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS41.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS45.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS63.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
SS78.001      TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
FS8.001       TCGAGGGCGGTTGCTTGGGTTTGAAGAAGCCTGTGAGGGTGCTTTGTTTGGATAAAGGAT
              ****** *****************************************************

CBS129661     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG34.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG7.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS1.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG42.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG9.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
CgSl1         CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG102.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG18.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG2.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG75.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
S3.001_CBS    CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
E031          CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
FS13.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
FS34.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
FS37.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
FS5.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
GS1.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
JG88.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS100.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS107.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS124.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS159.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS171.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS204.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS240.001     CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS26.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS29.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS41.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS45.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS63.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
SS78.001      CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
FS8.001       CTGATTCGCTGGCCTTGGTCTTCTTTGAGGGACGTCTTGAGGGCATTGCTGTGTCGGATG
              ************************************************************

CBS129661     CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG34.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG7.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS1.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG42.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG9.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
CgSl1         CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG102.001     CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG18.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG2.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG75.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
S3.001_CBS    CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
E031          CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
FS13.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
FS34.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
FS37.001      CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
```

```
FS5.001         CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
GS1.001         CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
JG88.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS100.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS107.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS124.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS159.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS171.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS204.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS240.001       CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS26.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS29.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS41.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS45.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS63.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
SS78.001        CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
FS8.001         CTCGCTTCAAGGGGGATCTTGTTTGTGAGGTTGGCAACTCTACCGGGCTTGATTGCGTAG
                ************************************************************

CBS129661       ACGCTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCGGTCCTCTTTCT
JG34.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG7.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
SS1.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG42.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG9.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
CgSl1           ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
JG102.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG18.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG2.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
JG75.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
S3.001_CBS      ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTGCTGAGAGAATTCTCCGTCCTCTTTCT
E031            ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
FS13.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
FS34.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
FS37.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
FS5.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
GS1.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
JG88.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS100.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS107.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS124.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS159.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS171.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS204.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS240.001       ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS26.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS29.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS41.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS45.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS63.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
SS78.001        ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
FS8.001         ACACTTGGGATGTTGGTGGCTCCAAAGTTTCTTCCTGAGAGAATTCTCCGTCCTCTTTCT
                ** ****************************** ************* ************

CBS129661       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG34.001        CTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG7.001         CTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS1.001         CTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG42.001        CTTGGGGGGTGGCCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCCATG
JG9.001         CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCCATG
CgSl1           CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG102.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
```

```
JG18.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG2.001        CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG75.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
S3.001_CBS     CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
E031           CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
FS13.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
FS34.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
FS37.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
FS5.001        CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
GS1.001        CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
JG88.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS100.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS107.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS124.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS159.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS171.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS204.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS240.001      CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS26.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS29.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS41.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS45.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS63.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
SS78.001       CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCCAATG
FS8.001        CTTGGGGGGTGGTCAGCTCCTCGGAGACGAGACTCTGCGACTGTGACTCC-ATG
             ************ ************************************* ***
```

PART 5: SOD2 sequences

CLUSTAL FORMAT: MUSCLE (3.8) multiple sequence alignment

```
CBS_129661      GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
E031            GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
FS13.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
FS34.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
FS37.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
FS5.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
FS8.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG88.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
S3001_CBS1      GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS100.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS107.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS124.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS159.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS171.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS204.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS240.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS26.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS29.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS41.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS45.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS78.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
CgSl1           GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
GS1.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS63.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG34.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG42.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG7.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
SS1.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG75.001        GG--ATTGTC-GCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG102.001       GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG18.001        GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG2.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
JG9.001         GGACATTGTCAGCACTTCTTCCCAAAACAGACTCCCACTGGCGTTCACGAACCTGAAGGT
                **  ****** *************************************************

CBS_129661      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
E031            AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
FS13.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
FS34.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
FS37.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
FS5.001         AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
FS8.001         AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG88.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
S3001_CBS1      AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS100.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS107.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS124.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS159.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS171.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS204.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS240.001       AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS26.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS29.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS41.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS45.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS78.001        AGTACGCGTGTTCCCACATGTCGACACCAAAAATGGGCACGCCCTTGGTGACAGGGTCTT
```

```
CgSl1         AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
GS1.001       AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS63.001      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG34.001      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG42.001      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG7.001       AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
SS1.001       AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG75.001      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG102.001     AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG18.001      AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG2.001       AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
JG9.001       AGTACGCGTGTTCCCACATGTCGACACCGAAAATGGGCACGCCCTTGGTGACAGGGTCTT
              **************************** *******************************

CBS_129661    GGTCTTTCGTCATGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
E031          GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
FS13.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
FS34.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
FS37.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
FS5.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
FS8.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG88.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
S3001_CBS1    GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS100.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS107.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS124.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS159.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS171.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS204.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS240.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS26.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS29.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS41.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS45.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS78.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
CgSl1         GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
GS1.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS63.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG34.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG42.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG7.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
SS1.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG75.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG102.001     GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG18.001      GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG2.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
JG9.001       GGTCTTTCGTCGTGATGATGCTGAGGCCCGTTACGTCGTCCTTGACCAGCCATCCCCAGC
              *********** ************************************************

CBS_129661    CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
E031          CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
FS13.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
FS34.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
FS37.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
FS5.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
FS8.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG88.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
S3001_CBS1    CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS100.001     CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS107.001     CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS124.001     CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS159.001     CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
```

```
SS171.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS204.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS240.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS26.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS29.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS41.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS45.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS78.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
CgSl1          CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
GS1.001        CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS63.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG34.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG42.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG7.001        CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
SS1.001        CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG75.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG102.001      CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG18.001       CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG2.001        CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
JG9.001        CGCTGCCGGTGATACCCAGCAGTGTGGCGTTGAAAGCCTGCTTGAACTGGTCGAGCCCGC
               ************************************************************

CBS_129661
CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
E031           CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
FS13.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
FS34.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
FS37.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
FS5.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
FS8.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG88.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
S3001_CBS1
CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS100.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS107.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS124.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS159.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS171.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS204.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS240.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS26.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS29.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS41.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS45.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS78.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
CgSl1          CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
GS1.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS63.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG34.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG42.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG7.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
SS1.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG75.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG102.001      CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG18.001       CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG2.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
JG9.001        CCCAGACGCGGGTGATCTCAGCGACGAGCTTGGGCGCCGCATCGGGCGAGGCATCTGGGC
               ************************************************************

CBS_129661
TCGATGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCGCCGTTGA
E031           TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
```

```
FS13.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
FS34.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
FS37.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
FS5.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
FS8.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG88.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
S3001_CBS1
TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS100.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS107.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS124.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS159.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS171.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS204.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS240.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS26.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS29.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS41.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS45.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS78.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
CgSl1          TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
GS1.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS63.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG34.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG42.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG7.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
SS1.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG75.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG102.001      TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG18.001       TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG2.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
JG9.001        TCGAGGCTGGGGACAGGTTCTCCCAGAAGAGGGAGTGGTTGATGTGGCCGCCTCCATTGA
               **** ************************************************ ** ****

CBS_129661
AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
E031           AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
FS13.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
FS34.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
FS37.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
FS5.001        AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
FS8.001        AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
JG88.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
S3001_CBS1
AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS100.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS107.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS124.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS159.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS171.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS204.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS240.001      AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS26.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS29.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS41.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS45.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS78.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
CgSl1          AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
GS1.001        AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
SS63.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
JG34.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
JG42.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGTTTGCATTGTAAGCCTCGA
```

```
JG7.001         AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
SS1.001         AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
JG75.001        AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
JG102.001       AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
JG18.001        AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
JG2.001         AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
JG9.001         AGTTGAGGGCCGCAAGGACGGCGATGCGGTTTTGGAGGGGGGTTTGCATTGTAAGCCTCGA
                *************************************************************

CBS_129661      TGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
E031            TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS13.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS34.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS37.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS5.001         TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
FS8.001         TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG88.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
S3001_CBS1      TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS100.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS107.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS124.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS159.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS171.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS204.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS240.001       TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS26.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS29.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS41.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS45.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS78.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
CgSl1           TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
GS1.001         TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS63.001        TGGCTTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG34.001        TGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG42.001        TGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG7.001         TGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
SS1.001         TGGCCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG75.001        TGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG102.001       TGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG18.001        TGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG2.001         TGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
JG9.001         TGGTCTTGTTCAGATTTGTAACGTATGCTTGGTGGCT
                ***  ********************************
```

References

1. Abu Assar, A. H., Uptmoor, R., Abdelmula, A. A., Salih, M., Ordon, F., and Friedt, W. 2005. Genetic Variation in Sorghum Germplasm from Sudan, ICRISAT, and USA Assessed by Simple Sequence Repeats (SSRs). Crop Science 45:1636-1644.
2. Aimi, T., Kano, S., Yotsutani, Y., and Morinaga, T. 2002. Telomeric fingerprinting of the white root rot fungus, Rosellinia necatrix: a useful tool for strain identification. FEMS microbiology letters 217:95-101.
3. Ali, M. E. K., and Warren, H. L. 1987. Physiological races of *Colletotrichum graminicola* on Sorghum. Plant DiseasePlant Disease 71 402-404.
4. Anas, Y. T. 2004. Sorghum diversity evaluated by simple sequence repeat (SSR) markers and phenotypic performance. Plant Production Science 7:301–308.
5. Ball, C. R. 1902. Johnson grass: report of investigations made during the season of 1901. U. S. Dep. of Agr. Bur. of Plant Ind. Bull. .
6. Barbanti, L., Grandi, S., Vecchi, A., and Venturi, G. 2006. Sweet and fibre sorghum (*Sorghum bicolor* (L.) Moench), energy crops in the frame of environmental protection from excessive nitrogen loads. European Journal of Agronomy 25:30-39.
7. Baroncelli, R., Amby, D. B., Zapparata, A., Sarrocco, S., Vannacci, G., Le Floch, G., Harrison, R. J., Holub, E., Sukno, S. A., and Sreenivasaprasad, S. 2016. Gene family expansions and contractions are associated with host range in plant pathogens of the genus Colletotrichum. BMC Genomics 17:1.
8. Baroncelli, R., Sanz-Martín, J. M., Rech, G. E., Sukno, S. A., and Thon, M. R. 2014. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. Genome announcements 2:e00540-00514.
9. Basavaraj, G., Rao, P. P., Basu, K., Reddy, C. R., Kumar, A. A., Rao, P. S., and Reddy, B. 2013. Assessing viability of bio-ethanol production from sweet sorghum in India. Energy policy 56:501-508.
10. Bentley, S., Chater, K., Cerdeno-Tarraga, A.-M., Challis, G., Thomson, N., James, K., Harris, D., Quail, M., Kieser, H., and Harper, D. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). Nature 417:141-147.
11. Birch, A. 1967. Biosynthesis of polyketides and related compounds. Science 156:202-206.
12. Bitzer, M. J. 1997. Production of sweet sorghum for syrup in Kentucky. Page 4. Cooperative Extension Service, College of Agriculture, University of Kentucky, Lexington, KY.
13. Bitzer, M. J. 2009. Sweet Sorghum for Syrup. Page 3. Cooperative Extension Service, College of Agriculture, University of Kentucky, Lexington, KY.
14. Boora, K. S., Frederiksen, R., and Magill, C. 1998. DNA-based markers for a recessive gene conferring anthracnose resistance in sorghum. Crop Science 38:1708-1709.
15. Borrell, A. K., Hammer, G. L., and Douglas, A. C. L. 2000. Does Maintaining Green Leaf Area in Sorghum Improve Yield under Drought? I. Leaf Growth and Senescence. Crop Science 40:1026-1037.

16. Brakhage, A. A., and Schroeckh, V. 2011. Fungal secondary metabolites–strategies to activate silent gene clusters. Fungal Genetics and Biology 48:15-22.
17. Brefort, T., Tanaka, S., Neidig, N., Doehlemann, G., Vincon, V., and Kahmann, R. 2014. Characterization of the largest effector gene cluster of *Ustilago maydis*. PLoS Pathog 10:e1003866.
18. Broadhead, D. M., Coleman, O. H., and Freeman, K. C. 1970. Dale, a new variety of sweet sorghum for sirup production. Mississippi Farm Research 33.
19. Browning, M., Rowley, L. V., Zeng, P., Chandlee, J. M., and Jackson, N. 1999. Morphological, pathogenic, and genetic comparisons of Colletotrichum graminicola isolates from Poaceae. Plant Disease 83:286-292.
20. Campbell, C. L., and Madden, L. V. 1990. Introduction to plant disease epidemiology. John Wiley & Sons.
21. Cantu, D., Segovia, V., MacLean, D., Bayles, R., Chen, X., Kamoun, S., Dubcovsky, J., Saunders, D. G., and Uauy, C. 2013. Genome analyses of the wheat yellow (stripe) rust pathogen *Puccinia striiformis* f. sp. *tritici* reveal polymorphic and haustorial expressed secreted proteins as candidate effectors. BMC genomics 14:1.
22. Cardwell, K., Kirkpatrick, T., Dale, J., and Hollier, C. 1987. Reported regional outbreaks of anthracnose of sorghum in a tristate area in 1985-86. Sorghum Newsletter 30:89-90.
23. Carlson, D. R., and Burnside, O. C. 1981. Use of the Recirculating Sprayer to Control Tall Weed Escapes in Crops. Weed Science 29:174-179.
24. Casela, C., and Frederiksen, R. 1993. Variability in the sorghum anthracnose fungus Colletotrichum graminicola in Brazil and USA. Pages 310-310 in: Durability of Disease Resistance. Springer.
25. Casela, C. R., Ferreira, A. S., and Santos, F. G. 1998. Associação de virulência de Colletotrichum graminicola à resistência genética em sorgo. Fitopatologia Brasileira 23:143-146.
26. Casela, C. R., Ferreira, A. S., Zeller, K. A., and Levy, M. 1995. Pathotype variation in the sorghum anthracnose fungus: A phylogenetic perspective for resistance breeding. Disease Analysis through Genetics and Molecular Biology: Interdisciplinary Bridges to Improved Sorghum and Millet Crops. JF Leslie and RA Frederiksen, eds. Iowa State University Press, Ames:257-276.
27. Casela, C. R., Santos, F., and Ferreira, A. S. 2000. Associacoes de patogenicidade e diversidade fenotipica de *Colletotrichum graminicola,* agente causal da antracnose em sorgo. Fitopatol. bras 25.
28. Chala, A., Alemu, T., Prom, L., and Tronsmo, A. 2010. Effect of host genotypes and weather variables on the severity and temporal dynamics of sorghum anthracnose in Ethiopia. Plant Pathology Journal (Faisalabad) 9:39-46.
29. Chala, A., Brurberg, M., and Tronsmo, A. 2010. Incidence and severity of sorghum anthracnose in Ethiopia. Plant Pathology Journal (Faisalabad) 9:23-30.
30. Chala, A., Tronsmo, A., and Brurberg, M. 2011. Genetic differentiation and gene flow in *Colletotrichum sublineolum* in Ethiopia, the centre of origin and diversity of sorghum, as revealed by AFLP analysis. Plant Pathology 60:474-482.
31. Chowdhury, S. C. 1936. A disease of *Zea mays* caused by *Colletotrichum graminicola* [Ces.] Wils. . Indian Journal of Agricultural Science 6:p. 833-843.

32. Chum, H. L., Warner, E., Seabra, J. E., and Macedo, I. C. 2014. A comparison of commercial ethanol production systems from Brazilian sugarcane and US corn. Biofuels, bioproducts and biorefining 8:205-223.
33. Chuma, I., Isobe, C., Hotta, Y., Ibaragi, K., Futamata, N., Kusaba, M., Yoshida, K., Terauchi, R., Fujita, Y., and Nakayashiki, H. 2011. Multiple translocation of the *AVR-Pita* effector gene among chromosomes of the rice blast fungus *Magnaporthe oryzae* and related species. PLoS Pathog 7:e1002147.
34. Clegg, M. 1982. Effect of soybean on yield and nitrogen response of subsequent sorghum crops in eastern Nebraska. Field Crops Research 5:233-239.
35. Coleman, O. H., and Stokes, I. 1954. The inheritance of resistance to stalk red rot in sorghum. Agronomy Journal 46:61-63.
36. Condon, B. J., Leng, Y., Wu, D., Bushley, K. E., Ohm, R. A., Otillar, R., Martin, J., Schackwitz, W., Grimwood, J., and MohdZainudin, N. 2013. Comparative genome structure, secondary metabolite, and effector coding capacity across Cochliobolus pathogens. PLoS Genet 9:e1003233.
37. Corredor, D. Y., Salazar, J. M., Hohn, K. L., Bean, S., Bean, B., and Wang, D. 2009. Evaluation and Characterization of Forage Sorghum as Feedstock for Fermentable Sugar Production. Appl Biochem Biotechnol 158:164-179.
38. Crouch, J., O'Connell, R., Gan, P., Buiate, E., Torres, M. F., Beirn, L., Shirasu, K., and Vaillancourt, L. 2014. The genomics of Colletotrichum. Pages 69-102 in: Genomics of Plant-Associated Fungi: Monocot Pathogens. Springer.
39. Crouch, J. A., Clarke, B. B., and Hillman, B. I. 2006. Unraveling evolutionary relationships among the divergent lineages of Colletotrichum causing anthracnose disease in turfgrass and corn. Phytopathology 96:46-60.
40. Crouch, J. A., Clarke, B. B., White, J. F., and Hillman, B. I. 2009. Systematic analysis of the falcate-spored graminicolous Colletotrichum and a description of six new species from warm-season grasses. Mycologia 101:717-732.
41. Crouch, J. A., and Tomaso-Peterson, M. 2012. Anthracnose disease of centipedegrass turf caused by Colletotrichum eremochloae, a new fungal species closely related to Colletotrichum sublineola. Mycologia 104:1085-1096.
42. Crouch, J. A., Tredway, L. P., Clarke, B. B., and Hillman, B. I. 2009. Phylogenetic and population genetic divergence correspond with habitat for the pathogen Colletotrichum cereale and allied taxa across diverse grass communities. Molecular ecology 18:123-135.
43. Cuevas, H. E., Prom, L. K., and Erpelding, J. E. 2015. Tapping the US sweet sorghum collection to identify biofuel germplasm. Sugar Tech 17:428-438.
44. da Costa, R., Cota, L., da Silva, D., Parreira, D., Casela, C., Landau, E., and Figueiredo, J. 2014. Races of Colletotrichum graminicola pathogenic to maize in Brazil. Crop Protection 56:44-49.
45. Dale, J. 1963. Corn anthracnose. Plant Dis. Rep 47:245-249.
46. Day, J., Duncan, R., Raymer, P., Lovell, G., Thompson, D., Garrett, H., and Zummo, N. 1995. Top 76-6: a new sweet sorghum variety for sirup production. Research Report-Georgia Agricultural Experiment Stations.
47. Dayakar Rao, B., Ratnavathi, C., Karthikeyan, K., Biswas, P., Rao, S., Vijay Kumar, B., and Seetharama, N. 2004. Sweet sorghum cane for bio-fuel production: A

SWOT analysis in Indian context. National Research Centre for Sorghum, Rajendranagar, Hyderabad, AP 500:030.

48. de Jonge, R., Bolton, M. D., and Thomma, B. P. 2011. How filamentous pathogens co-opt plants: the ins and outs of fungal effectors. Current opinion in plant biology 14:400-406.

49. De Oliveira, M. E. D., Vaughan, B. E., and Rykiel Jr, E. J. 2005. Ethanol as fuel: energy, carbon dioxide balances, and ecological footprint. BioScience 55:593-602.

50. de Wit, P. J., Van Der Burgt, A., Ökmen, B., Stergiopoulos, I., Abd-Elsalam, K. A., Aerts, A. L., Bahkali, A. H., Beenen, H. G., Chettri, P., and Cox, M. P. 2012. The genomes of the fungal plant pathogens *Cladosporium fulvum* and *Dothistroma septosporum* reveal adaptation to different hosts and lifestyles but also signatures of common ancestry. PLoS Genet 8:e1003088.

51. Dean, R., Van Kan, J. A., Pretorius, Z. A., Hammond-Kosack, K. E., Di Pietro, A., Spanu, P. D., Rudd, J. J., Dickman, M., Kahmann, R., and Ellis, J. 2012. The Top 10 fungal pathogens in molecular plant pathology. Molecular plant pathology 13:414-430.

52. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., and Lescot, M. 2008. Phylogeny. fr: robust phylogenetic analysis for the non-specialist. Nucleic acids research 36:W465-W469.

53. Djamei, A., Schipper, K., Rabe, F., Ghosh, A., Vincon, V., Kahnt, J., Osorio, S., Tohge, T., Fernie, A. R., and Feussner, I. 2011. Metabolic priming by a secreted fungal effector. Nature 478:395-398.

54. Dodd, J. L. 1980. The photosynthetic stress translocation balance concept of sorghum stalk rot. Pages 300-305 in: Proceedings of International Workshop of Sorghum Diseases.

55. Dodd, J. L. 1980. The role of plant stresses in development of corn stalk rots. Plant Disease 64:533.

56. Doggett, H. 1988. Sorghum. 2nd ed. John Wiley & Sons, New York.

57. Doggett, H. 1988. Sorghum. Harlow, Essex, England New York: Longman Scientific & Technical. Wiley 18:512.

58. Dong, S., Stam, R., Cano, L. M., Song, J., Sklenar, J., Yoshida, K., Bozkurt, T. O., Oliva, R., Liu, Z., and Tian, M. 2014. Effector specialization in a lineage of the Irish potato famine pathogen. Science 343:552-555.

59. Donofrio, N. M., and Raman, V. 2012. Roles and delivery mechanisms of fungal effectors during infection development: common threads and new directions. Current opinion in microbiology 15:692-698.

60. Dou, D., Kale, S. D., Wang, X., Jiang, R. H., Bruce, N. A., Arredondo, F. D., Zhang, X., and Tyler, B. M. 2008. RXLR-mediated entry of *Phytophthora sojae* effector Avr1b into soybean cells does not require pathogen-encoded machinery. The Plant Cell 20:1930-1947.

61. Du, M., Schardl, C. L., Nuckles, E. M., and Vaillancourt, L. J. 2005. Using mating-type gene sequences for improved phylogenetic resolution of Collectotrichum species complexes. Mycologia 97:641-658.

62. Ellis, J. G., Rafiqi, M., Gan, P., Chakrabarti, A., and Dodds, P. N. 2009. Recent progress in discovery and functional analysis of effector proteins of fungal and oomycete plant pathogens. Current opinion in plant biology 12:399-405.

63. Erickson, J. E., Woodard, K. R., and Sollenberger, L. E. 2012. Optimizing sweet sorghum production for biofuel in the southeastern USA through nitrogen fertilization and top removal. BioEnergy Research 5:86-94.

64. Erpelding, J. E., and Prom, L. K. 2004. Evaluation of Malian sorghum germplasm for resistance against anthracnose. Plant Pathol. J 3:65-71.

65. Farman, M. L. 2011. Targeted cloning of fungal telomeres. Pages 11-31 in: Fungal Genomics. Springer.

66. FARMAN, M. L., and KIM, Y. S. 2005. Telomere hypervariability in Magnaporthe oryzae. Molecular plant pathology 6:287-298.

67. Ferreira, A., and Warren, H. 1982. Resistance of sorghum to Colletotrichum graminicola. Plant Disease 66:773-775.

68. Ferreira, A. S., and Warren, H. L. 1982. Resistance of sorghum to *Colletotrichum sublineolum*. Plant Disease 66:773–775.

69. Figueiredo, J. E. F., Depaoli, H. C., da Silva Coelho, V., Casela, C., da Silva Ferreira, A., Guimarães, C., Gomes, E., and Bressan, W. 2006. Genetic diversity among Colletotrichum sublineolum pathotypes isolated from sorghum (*Sorghum bicolor*). Revista Brasileira de Milho e Sorgo 5:304-318.

70. Forgey, W., Blanco, M., and Loegering, W. 1978. Differences in pathological capabilities and host specificity of *Colletotrichum graminicola* on *Zea mays* [maize]. Plant Disease Reporter (USA).

71. Frederiksen, R. A., and Odvody, G. N. 2000. Compendium of sorghum diseases. American Phytopathological Society (APS Press).

72. Gale, L. R. 2008. A Population Genetic Approach to Variation in Colletotrichum graminicola, the Causal. Sorghum and millets diseases:191.

73. Gale, L. R. 2002. A population genetic approach to variation in *Colletotrichum graminicola*, the causal agent of soghum anthracnose. Pages 191-195 in: Sorghum and Millets Diseases. J. F. Leslie, ed. Iowa State Press, Ames, Iowa.

74. Gan, P., Narusaka, M., Kumakura, N., Tsushima, A., Takano, Y., Narusaka, Y., and Shirasu, K. 2016. Genus-wide comparative genome analyses of Colletotrichum species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. Genome biology and evolution 8:1467-1481.

75. Ganapathy, K. N., Gomashe, S. S., Rakshit, S., Prabhakar, B., Ambekar, S. S., Ghorade, R. B., Biradar, B. D., Saxena, U., and Patil, J. V. 2012. Genetic diversity revealed utility of SSR markers in classifying parental lines and elite genotypes of sorghum (*Sorghum bicolor* L. Moench). Australian Journal of Crop Science 6:1486-1493.

76. Geng, S., Hills, F., Johnson, S., and Sah, R. 1989. Potential Yields and On-Farm Ethanol Production Cost of Corn, Sweet Sorghum, Fodderbeet, and Sugarbeet. Journal of Agronomy and Crop Science 162:21-29.

77. Grassi, G., Tondi, G., and Helm, P. 2004. Small-sized commercial bioenergy technologies as an instrument of rural development. and Agriculture:277.

78. Guthrie, P., Magill, C., Frederiksen, R., and Odvody, G. 1992. Random amplified polymorphic DNA markers: a system for identifying and differentiating isolates of Colletotrichum graminicola. Phytopathology 82:832-835.
79. Hagan, A. 2015. Yield Parameters and Disease Activity of Biomass, Forage, and Sweet Sorghum Varieties in Alabama in 2013. PP-763.
80. Hagan, A., Bowen, K., Pegues, M., and Jones, J. 2014. Nitrogen Rate and Variety Impact Diseases and Yield of Sorghum for Biofuel. Agronomy Journal 106:1205-1211.
81. Hallam, A., Anderson, I. C., and Buxton, D. R. 2001. Comparative economic analysis of perennial, annual, and intercrops for biomass production. Biomass and Bioenergy 21:407-424.
82. Harnmerschmidt, R., and Nicholson, R. 1977. Resistance of Maize to Anthraenose: Effect of Light Intensity on Lesion Development. development 67:247-250.
83. Harris, H. B., Johnson, B. J., Dobson, J. W., and Luttrell, E. S. 1964. Evaluation of Anthracnose on Grain Sorghum. Crop Science 4:460-462.
84. Hill, R. J. G. S., and Forer, L. B. 1981. Mile-a-minute, *Polygonum perfoliatum* L. (Polygonaceae), a new potential orchard and nursery weed. . Regulatory Horticulture 7:25-28.
85. Holou, R. A. Y., and Stevens, G. 2012. Juice, sugar, and bagasse response of sweet sorghum (*Sorghum bicolor* (L.) Moench cv. M81E) to N fertilization and soil type. GCB Bioenergy 4:302-310.
86. House, L. R. 1985. A Guide to Sorghum Breeding. Second Edition ed. International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India.
87. Hunter, E., and Anderson, I. 1997. Sweet sorghum. Horticultural reviews 21:73-104.
88. Hyde, K., Cai, L., Cannon, P., Crouch, J., Crous, P., Damm, U., Goodwin, P., Chen, H., Johnston, P., and Jones, E. 2009. Colletotrichum—names in current use. Fungal Diversity 39:147-182.
89. Ibraheem, F., Gaffoor, I., and Chopra, S. 2010. Flavonoid phytoalexin-dependent resistance to anthracnose leaf blight requires a functional yellow seed1 in *Sorghum bicolor*. Genetics 184:915-926.
90. Ito, K., Tanaka, T., Hatta, R., Yamamoto, M., Akimitsu, K., and Tsuge, T. 2004. Dissection of the host range of the fungal plant pathogen Alternaria alternata by modification of secondary metabolism. Molecular microbiology 52:399-411.
91. Jaccard, P. 1908. Nouvelles recherches sur la distribution florale.
92. Jamil, F., and Nicholson, R. 1987. Susceptibility of corn to isolates of Colletotrichum graminicola pathogenic to other grasses. Plant Disease 71:809-810.
93. Jones, J. D., and Dangl, J. L. 2006. The plant immune system. Nature 444:323-329.
94. Jothi, R., Zotenko, E., Tasneem, A., and Przytycka, T. M. 2006. COCO-CL: hierarchical clustering of homology relations based on evolutionary correlations. Bioinformatics 22:779-788.
95. Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. Phytopathology 44:41.
96. Kang, S., Sweigard, J. A., and Valent, B. 1995. The PWL host specificity gene family in the blast fungus *Magnaporthe grisea*. MPMI-Molecular Plant Microbe Interactions 8:939-948.

97. Kankanala, P., Czymmek, K., and Valent, B. 2007. Roles for rice membrane dynamics and plasmodesmata during biotrophic invasion by the blast fungus. The Plant Cell Online 19:706-724.
98. Karimi, A., and Saeidi, H. 2016. Genetic diversity of Sorghum halepense (L.) Pers. in Iran as revealed by IRAP markers. Plant Genetic Resources 14:132-141.
99. Kemen, E., Kemen, A. C., Rafiqi, M., Hempel, U., Mendgen, K., Hahn, M., and Voegele, R. T. 2005. Identification of a protein from rust fungi transferred from haustoria into infected plant cells. Molecular Plant-Microbe Interactions 18:1130-1139.
100. Khang, C. H., Berruyer, R., Giraldo, M. C., Kankanala, P., Park, S.-Y., Czymmek, K., Kang, S., and Valent, B. 2010. Translocation of *Magnaporthe oryzae* effectors into rice cells and their subsequent cell-to-cell movement. The Plant Cell 22:1388-1403.
101. Khosla, C., Gokhale, R. S., Jacobsen, J. R., and Cane, D. E. 1999. Tolerance and specificity of polyketide synthases. Annual review of biochemistry 68:219-253.
102. Kim, M., and Day, D. F. 2011. Composition of sugar cane, energy cane, and sweet sorghum suitable for ethanol production at Louisiana sugar mills. Journal of industrial microbiology & biotechnology 38:803-807.
103. Kroken, S., Glass, N. L., Taylor, J. W., Yoder, O., and Turgeon, B. G. 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proceedings of the National Academy of Sciences 100:15670-15675.
104. Le Beau, F., and Coleman, O. 1950. The inheritance of resistance in Sorghum to leaf anthracnose. Agronomy Journal 42:33-34.
105. LeBeau, F. 1946. The eradicant action of a fungicide on the Colletotrichum-Lilii in lily bulbs Pages 391-393 American Phytopathological Society 3340 pilot knob road, St. Paul, MN 55121
106. LeBeau, F. J., Stokes, I., and Coleman, O. H. 1951. Anthracnose and red rot of sorghum. US Government Printing Office.
107. Lee, H. A., Kim, S. Y., Oh, S. K., Yeom, S. I., Kim, S. B., Kim, M. S., Kamoun, S., and Choi, D. 2014. Multiple recognition of RXLR effectors is associated with nonhost resistance of pepper against *Phytophthora infestans*. New Phytologist 203:926-938.
108. Lee, S.-L., Floss, H. G., and Heinstein, P. 1976. Purification and properties of dimethylallylpyrophosphate: Tryptophan dimethylallyl transferase, the first enzyme of ergot alkaloid biosynthesis in *Claviceps*. sp. SD 58. Archives of Biochemistry and Biophysics 177:84-94.
109. Levis, C., Giraud, T., Dutertre, M., Fortini, D., and Brygoo, Y. 1997. Telomeric DNA of Botrytis cinerea: a useful tool for strain identification. FEMS Microbiology Letters 157:267-272.
110. Li, B.-Z., Balan, V., Yuan, Y.-J., and Dale, B. E. 2010. Process optimization to convert forage and sweet sorghum bagasse to ethanol based on ammonia fiber expansion (AFEX) pretreatment. Bioresource Technology 101:1285-1292.
111. Li, L. S. C. J., and Roos, D. S. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research 13:2178-2189.

112. LO, S.-C. C., DE VERDIER, K., and NICHOLSON, R. L. 1999. Accumulation of 3-deoxyanthocyanidin phytoalexins and resistance to Colletotrichum sublineolum in sorghum. Physiological and Molecular Plant Pathology 55:263-273.
113. Martin, J. F., and Liras, P. 1989. Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. Annual Reviews in Microbiology 43:173-206.
114. Matsumura, K., and Tosa, Y. 1995. The rye mildew fungus carries avirulence genes corresponding to wheat genes for resistance to races of the wheat mildew fungus. Phytopathology 85:753-756.
115. McAlpine, J. B., Bachmann, B. O., Piraee, M., Tremblay, S., Alarco, A.-M., Zazopoulos, E., and Farnet, C. M. 2005. Microbial genomics as a guide to drug discovery and structural elucidation: ECO-02301, a novel antifungal agent, as an example. Journal of Natural Products 68:493-496.
116. McWhorter, C. G. 1981. Control of Johnsongrass Ecotypes. Weed Science 19:229-233.
117. Mims, C., and Vaillancourt, L. 2002. Ultrastructural characterization of infection and colonization of maize leaves by *Colletotrichum graminicola*, and by a *C. graminicola* pathogenicity mutant. Phytopathology 92:803-812.
118. Mitchell, J. K., Yerkes, C. N., Racine, S. R., and Lewis, E. H. 2008. The interaction of two potential fungal bioherbicides and a sub-lethal rate of glyphosate for the control of shattercane. Biological Control 46:391-399.
119. Moreno-Hagelsieb, G., and Latimer, K. 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. Bioinformatics 24:319-324.
120. Morrell, P., Williams-Coplin, T., Lattu, A., Bowers, J., Chandler, J., and Paterson, A. 2005. Crop-to-weed introgression has impacted allelic composition of johnsongrass populations with and without recent exposure to cultivated sorghum. Molecular Ecology 14:2143-2154.
121. Mortimore, C., and Ward, G. 1964. Root and stalk rot of corn in southwestern Ontario: III. Sugar levels as a measure of plant vigor and resistance. Canadian Journal of Plant Science 44:451-457.
122. Mosquera, G., Giraldo, M. C., Khang, C. H., Coughlan, S., and Valent, B. 2009. Interaction transcriptome analysis identifies *Magnaporthe oryzae* BAS1-4 as biotrophy-associated secreted proteins in rice blast disease. The Plant Cell 21:1273-1290.
123. Muchow, R., and Davis, R. 1988. Effect of nitrogen supply on the comparative productivity of maize and sorghum in a semi-arid tropical environment II. Radiation interception and biomass accumulation. Field Crops Research 18:17-30.
124. Mughogho, L. 1982. Strategies for sorghum disease control.
125. Murakami, J., Tosa, Y., Kataoka, T., Tomita, R., Kawasaki, J., Chuma, I., Sesumi, Y., Kusaba, M., Nakayashiki, H., and Mayama, S. 2000. Analysis of host species specificity of *Magnaporthe grisea* toward wheat using a genetic cross between isolates from wheat and foxtail millet. Phytopathology 90:1060-1067.
126. Negeri, A., Wang, G.-F., Benavente, L., Kibiti, C. M., Chaikam, V., Johal, G., and Balint-Kurti, P. 2013. Characterization of temperature and light effects on the defense response phenotypes associated with the maize Rp1-D21 autoactive resistance gene. BMC plant biology 13:1.

127. Nemri, A., Saunders, D. G., Anderson, C., Upadhyaya, N. M., Win, J., Lawrence, G. J., Jones, D. A., Kamoun, S., Ellis, J. G., and Dodds, P. N. 2014. The genome sequence and effector complement of the flax rust pathogen *Melampsora lini*.
128. Nicholson, R., and Warren, H. 1981. The issue of races of Colletotrichum graminicola pathogenic to corn. Plant Diseases.
129. Nicholson, R. L., Jamil, F. F., Snyder, B. A., Lue, W. L., and Hipskind, J. 1988. Phytoalexin synthesis in the juvenile sorghum leaf. Physiological and Molecular Plant Pathology 33:271-278.
130. O'Connell, R., Herbert, C., Sreenivasaprasad, S., Khatib, M., Esquerré-Tugayé, M.-T., and Dumas, B. 2004. A novel Arabidopsis-Colletotrichum pathosystem for the molecular dissection of plant-fungal interactions. Molecular Plant-Microbe Interactions 17:272-282.
131. O'Connell, R. J., Thon, M. R., Hacquard, S., Amyotte, S. G., Kleemann, J., Torres, M. F., Damm, U., Buiate, E. A., Epstein, L., and Alkan, N. 2012. Lifestyle transitions in plant pathogenic Colletotrichum fungi deciphered by genome and transcriptome analyses. Nature genetics.
132. Paterson, A. H., Schertz, K. F., Lin, Y.-R., Liu, S.-C., and Chang, Y.-L. 1995. The weediness of wild plants: molecular analysis of genes influencing dispersal and persistence of johnsongrass, Sorghum halepense (L.) Pers. Proceedings of the National Academy of Sciences 92:6127-6131.
133. Pederson, G. A. 2010. Status of the U.S. Sweet Sorghum Collection for Biofuel Research. in: International Symposium on Forage, Turfgrass and Biofuel Germplasm Research.
134. Peel, M. D. 1998. Crop Rotations for Increased Productivity. North Dakota State University.
135. Prom, L., Perumal, R., Erattaimuthu, S., Little, C., No, E., Erpelding, J., Rooney, W., Odvody, G., and Magill, C. 2012. Genetic diversity and pathotype determination of Colletotrichum sublineolum isolates causing anthracnose in sorghum. European journal of plant pathology 133:671-685.
136. Prom, L. K., Perumal, R., Erpelding, J., Isakeit, T., Montes-Garcia, N., and Magill, C. 2009. A pictorial technique for mass screening of sorghum germplasm for anthracnose (Colletotrichum sublineolum) resistance. The Open Agriculture Journal 3:20-25.
137. Raffaele, S., Farrer, R. A., Cano, L. M., Studholme, D. J., MacLean, D., Thines, M., Jiang, R. H., Zody, M. C., Kunjeti, S. G., and Donofrio, N. M. 2010. Genome evolution following host jumps in the Irish potato famine pathogen lineage. Science 330:1540-1543.
138. Rafiqi, M., Ellis, J. G., Ludowici, V. A., Hardham, A. R., and Dodds, P. N. 2012. Challenges and progress towards understanding the role of effectors in plant–fungal interactions. Current opinion in plant biology 15:477-482.
139. Ratnavathi, C., Biswas, P., Pallavi, M., Maheswari, M., Kumar, B. S. V., and Seetharama, N. 2004. Alternative uses of sorghum–Methods and feasibility: Indian perspective. in: Alternative uses of Sorghum and Pearl Millets in Asia. Proceedings of the Expert Meeting.
140. Reddy, B. B. S., Ramesh, S., and Reddy, P. S. 2006. Sorghum genetic resources, cytogenetics and improvement. . Pages 309-363 in: Genetic resurces, chromosome

engineering, and crop improvement. R. J. Singh and P. P. Jauhar, eds. Taylor & Francis Group, Boca Raton, FL.

141. Rep, M., Van Der Does, H. C., Meijer, M., Van Wijk, R., Houterman, P. M., Dekker, H. L., De Koster, C. G., and Cornelissen, B. J. 2004. A small, cysteine-rich protein secreted by *Fusarium oxysporum* during colonization of xylem vessels is required for I-3-mediated resistance in tomato. Molecular microbiology 53:1373-1383.

142. Rooney, W. L., Blumenthal, J., Bean, B., and Mullet, J. E. 2007. Designing sorghum as a dedicated bioenergy feedstock. Biofuels, Bioproducts and Biorefining 1:147-157.

143. Rosewich, U., Pettway, R., McDonald, B., Duncan, R., and Frederiksen, R. 1998. Genetic structure and temporal dynamics of a *Colletotrichum graminicola* population in a sorghum disease nursery. Phytopathology 88:1087-1093.

144. Rosewich, U. L. 1996. The population genetics of Colletotrichum graminicola from different ecosystems of sorghum. Texas A&M University.

145. Saballos, A. 2008. Development and utilization of sorghum as a bioenergy crop. Pages 211-248 in: Genetic Improvement of Bioenergy Crops. Springer.

146. Saccardo, P., and Trotter, A. 1913. Supplementum Universale Pars IX. Sylloge Fungorum vol. 22: 1–1612. Padova, Italy.

147. Sánchez-Alonso, P., Valverde, M. E., Paredes-López, O., and Guzmán, P. 1996. Detection of genetic variation in Ustilago maydis strains by probes derived from telomeric sequences. Microbiology 142:2931-2936.

148. Sasaki, T., and Antonio, B. A. 2009. Plant genomics: Sorghum in sequence. Nature 457:547–548.

149. Saunders, D. G., Win, J., Cano, L. M., Szabo, L. J., Kamoun, S., and Raffaele, S. 2012. Using hierarchical clustering of secreted protein families to classify and rank candidate effectors of rust fungi. PLoS One 7:e29847.

150. Schall, R., Nicholson, R., and Warren, H. 1980. Influence of light on maize anthracnose in the greenhouse. Phytopathology 70:1023-1026.

151. Schardl, C. L., Young, C. A., Hesse, U., Amyotte, S. G., Andreeva, K., Calie, P. J., Fleetwood, D. J., Haws, D. C., Moore, N., and Oeser, B. 2013. Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. PLoS Genet 9:e1003323.

152. Schirawski, J., Mannhaupt, G., Münch, K., Brefort, T., Schipper, K., Doehlemann, G., Di Stasio, M., Rössel, N., Mendoza-Mendoza, A., and Pester, D. 2010. Pathogenicity determinants in smut fungi revealed by genome comparison. Science 330:1546-1548.

153. Schulze-Lefert, P., and Panstruga, R. 2011. A molecular evolutionary concept connecting nonhost resistance, pathogen host range, and pathogen speciation. Trends in plant science 16:117-125.

154. Scoles, G., Nga, N., Hau, V., and Tosa, Y. 2009. Identification of genes for resistance to a Digitaria isolate of *Magnaporthe grisea* in common wheat cultivars. Genome 52:801-809.

155. Shapouri, H., Duffield, J. A., and Wang, M. Q. 2002. The energy balance of corn ethanol: an update. United States Department of Agriculture, Economic Research Service.

156. Sharma, R., Upadhyaya, H., Manjunatha, S., Rao, V., and Thakur, R. 2012. Resistance to foliar diseases in a mini-core collection of sorghum germplasm. Plant Disease 96:1629-1633.
157. Sherriff, C., Whelan, M., Arnold, G., and Bailey, J. 1995. rDNA sequence analysis confirms the distinction between Colletotrichum graminicola and C. sublineolum. Mycological Research 99:475-478.
158. Sipos, B., Réczey, J., Somorai, Z., Kádár, Z., Dienes, D., and Réczey, K. 2009. Sweet Sorghum as Feedstock for Ethanol Production: Enzymatic Hydrolysis of Steam-Pretreated Bagasse. Appl Biochem Biotechnol 153:151-162.
159. Slewinski, T. L. 2012. Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. Journal of Experimental Botany.
160. Smith, C. W., and Frederiksen, R. A. 2000. Sorghum: origin, history, technology, and production. NY: John Wiley and Sons, New York.
161. Snyder, B. A., Leite, B., Hipskind, J., Butler, L. G., and Nicholson, R. L. 1991. Accumulation of sorghum phytoalexins induced by Colletotrichum graminicola at the infection site. Physiological and Molecular Plant Pathology 39:463-470.
162. Snyder, B. A., and Nicholson, R. L. 1990. Synthesis of phytoalexins in sorghum as a site-specific response to fungal ingress. Science 248:1637-1639.
163. Spanu, P. D., Abbott, J. C., Amselem, J., Burgis, T. A., Soanes, D. M., Stüber, K., van Themaat, E. V. L., Brown, J. K., Butcher, S. A., and Gurr, S. J. 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330:1543-1546.
164. Stachelhaus, T., Mootz, H. D., and Marahiel, M. A. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & biology 6:493-505.
165. Sutton, B. 1968. The appressoria of Colletotrichum graminicola and C. falcatum. Canadian Journal of Botany 46:873-876.
166. Swanson, A. F., and Laude, H. H. 1934. Varieties of sorghum in Kansas Agricultural experiment station Bulletin 266:50.
167. Swigoňová, Z., Lai, J., Ma, J., Ramakrishna, W., Llaca, V., Bennetzen, J. L., and Messing, J. 2004. Close split of sorghum and maize genome progenitors. Genome research 14:1916-1923.
168. Takabayashi, N., Tosa, Y., Oh, H., and Mayama, S. 2002. A gene-for-gene relationship underlying the species-specific parasitism of Avena/Triticum isolates of *Magnaporthe grisea* on wheat cultivars. Phytopathology 92:1182-1188.
169. Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal genetics and biology 31:21-32.
170. Tenkouano, A., Miller, F. R., Hart, G. E., Frederiksen, R. A., and Nicholson, R. L. 1993. Phytoalexin assay in juvenile sorghum: an aid to breeding for anthracnose resistance. Crop Science 33:243-248.
171. Tesso, T., Perumal, R., Little, C. R., Adeyanju, A., Radwan, G. L., Prom, L. K., and Magill, C. W. 2012. Sorghum pathology and biotechnology-a fungal disease perspective: Part II. Anthracnose, stalk rot, and downy mildew. European Journal of Plant Science and Biotechnology 6:31-44.

172. Thakur, R., Mathur, K., Rao, V., Chandra, S., Shivramkrishnan, S., Kannan, S., Hiremath, R., Tailor, H., Kushwaha, U., and Dwivedi, R. 1998. Pathogenic and genetic characterization of six Indian populations of Colletotrichum sublineolum, the causal agent of sorghum anthracnose. Indian Phytopathology 51:338-348.
173. Thomas, M. D., Sissoko, I., and Sacko, M. 1996. Development of leaf anthracnose and its effect on yield and grain weight of sorghum in West Africa. Plant disease. 80 151-153.
174. Thon, M., Nuckles, E., and Vaillancourt, L. 2000. Restriction enzyme-mediated integration used to produce pathogenicity mutants of Colletotrichum graminicola. Molecular Plant-Microbe Interactions 13:1356-1365.
175. Torres, M. F., Cuadros, D. F., and Vaillancourt, L. J. 2014. Evidence for a diffusible factor that induces susceptibility in the Colletotrichum–maize disease interaction. Molecular plant pathology 15:80-93.
176. Torres, M. F., Ghaffari, N., Buiate, E. A., Moore, N., Schwartz, S., Johnson, C. D., and Vaillancourt, L. J. 2016. A *Colletotrichum graminicola* mutant deficient in the establishment of biotrophy reveals early transcriptional events in the maize anthracnose disease interaction. BMC genomics 17:1.
177. Tosa, Y. 1992. A model for the evolution of formae speciales and races. Phytopathology 82:728-730.
178. Tosa, Y., Tamba, H., Tanaka, K., and Mayama, S. 2006. Genetic analysis of host species specificity of *Magnaporthe oryzae* isolates from rice and wheat. Phytopathology 96:480-484.
179. Tuite, J. 1969. Plant pathological methods. Fungi and bacteria. Burgess Publishing Co. Minneapolis MN. .
180. Tyler, B. M. 2009. Entering and breaking: virulence effector proteins of oomycete plant pathogens. Cellular microbiology 11:13-20.
181. Vaillancourt, L. J., and Hanau, R. M. 1992. Genetic and morphological comparisons of *Glomerella* (*Colletotrichum*) isolates from maize and from sorghum. Experimental Mycology 16:219-229.
182. Vale, F., Fernandes Filho, E., and Liberato, J. 2003. QUANT. A software for plant disease severity assessment. Page 105 in: 8th International Congress of Plant Pathology.
183. Valent, A., Bénard, J., Clausse, B., Barrois, M., Valteau-Couanet, D., Terrier-Lacombe, M.-J., Spengler, B., and Bernheim, A. 2001. In vivo elimination of acentric double minutes containing amplified MYCN from neuroblastoma tumor cells through the formation of micronuclei. The American journal of pathology 158:1579-1584.
184. Valerio, H., Resende, M., Weikert-Oliveira, R., and Casela, C. 2005. Virulence and molecular diversity in Colletotrichum graminicola from Brazil. Mycopathologia 159:449-459.
185. van der Does, H. C., and Rep, M. 2007. Virulence genes and the evolution of host specificity in plant-pathogenic fungi. Molecular Plant-Microbe Interactions 20:1175-1182.
186. van der Hoorn, R. A., and Kamoun, S. 2008. From guard to decoy: a new model for perception of plant pathogen effectors. The Plant Cell 20:2009-2017.

187. Venard, C., and Vaillancourt, L. 2007. Penetration and colonization of unwounded maize tissues by the maize anthracnose pathogen Colletotrichum graminicola and the related nonpathogen C. sublineolum. Mycologia 99:368-377.

188. Vermerris, W., Erickson, J., Wright, D., Newman, Y., and Rainbolt, C. 2008. Production of Biofuel Crops in Florida: Sweet Sorghum. University of Florida, Agronomy Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. .

189. Vleeshouwers, V. G., and Oliver, R. P. 2014. Effectors as tools in disease resistance breeding against biotrophic, hemibiotrophic, and necrotrophic plant pathogens. Molecular Plant-Microbe Interactions 27:196-206.

190. Wall, D., Fraser, H., and Hirsh, A. 2003. Detecting putative orthologs. Bioinformatics 19:1710-1711.

191. Wang, F., and Liu, C.-Z. 2009. Development of an Economic Refining Strategy of Sweet Sorghum in the Inner Mongolia Region of China. Energy & Fuels 23:4137-4142.

192. Wang, M., Han, J., Dunn, J. B., Cai, H., and Elgowainy, A. 2012. Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for US use. Environmental Research Letters 7:045905.

193. Warren, H., and Frederiksen, R. 1986. Compendium of sorghum diseases. Compendium of sorghum diseases.

194. Well, K. F. C. 1989. Pathotypes of Colletotrichum graminicola and seed transmission of sorghum anthracnose. Plant Disease:255.

195. Wharton, P., and Julian, A. 1996. A cytological study of compatible and incompatible interactions between *Sorghum bicolor* and Colletotrichum sublineolum. New Phytologist 134:25-34.

196. Wharton, P. S., and Nicholson, R. L. 2000. Temporal synthesis and radiolabelling of the sorghum 3-deoxyanthocyanidin phytoalexins and the anthocyanin, cyanidin 3-dimalonyl glucoside. New Phytologist 145:457-469.

197. Wheals, A. E., Basso, L. C., Alves, D. M., and Amorim, H. V. 1999. Fuel ethanol after 25 years. Trends in biotechnology 17:482-487.

198. Wheeler, H., Politis, D., and Poneleit, C. 1974. Pathogenicity, host range, and distribution of *Colletotrichum graminicola* on corn. Phytopathology 64:293-296.

199. White, D., Yanney, J., and Anderson, B. 1987. Variation in pathogenicity, virulence, and aggressiveness of Colletotrichum graminicola on corn. Phytopathology 77:999-1001.

200. White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications 18:315-322.

201. Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic acids research 18:6531-6535.

202. Williams, L., and Willis, G. 1963. Disease of corn caused by *Colletotrichum graminicolum*. Phytopathology 53:364-365.

203. Win, J., Morgan, W., Bos, J., Krasileva, K. V., Cano, L. M., Chaparro-Garcia, A., Ammar, R., Staskawicz, B. J., and Kamoun, S. 2007. Adaptive evolution has

targeted the C-terminal domain of the RXLR effectors of plant pathogenic oomycetes. The Plant Cell 19:2349-2369.

204. Wortmann, C. S., Liska, A., Ferguson, R. B., Lyon, D. J., Klein, R., and Dweikat, I. 2010. Dryland performance of sweet sorghum and grain crops for biofuel in Nebraska. Agronomy Journal 102:319-326.

205. Xue, M., Yang, J., Li, Z., Hu, S., Yao, N., Dean, R. A., Zhao, W., Shen, M., Zhang, H., and Li, C. 2012. Comparative analysis of the genomes of two field isolates of the rice blast fungus Magnaporthe oryzae. PLoS Genet 8:e1002869.

206. Yamoah, C. F., Clegg, M. D., and Francis, C. A. 1998. Rotation effect on sorghum response to nitrogen fertilizer under different rainfall and temperature environments. Agriculture, ecosystems & environment 68:233-243.

207. Yoshida, K., Saunders, D. G., Mitsuoka, C., Natsume, S., Kosugi, S., Saitoh, H., Inoue, Y., Chuma, I., Tosa, Y., and Cano, L. M. 2016. Host specialization of the blast fungus *Magnaporthe oryzae* is associated with dynamic gain and loss of genes linked to transposable elements. BMC genomics 17:1.

208. Yu, X. 2014. Genomic selection of biomass traits in a global collection of 976 sorghum accessions. in: Plant and Animal Genome XXII Conference Plant and Animal Genome.

209. Zegada-Lizarazu, W., and Monti, A. 2012. Are we ready to cultivate sweet sorghum as a bioenergy feedstock? A review on field management practices. Biomass and Bioenergy 40:1-12.

Vita

Katia Viana Xavier

EDUCATION

1. University of Kentucky (USA) PhD Plant Pathology. 2012 - present
2. Universidade Federal de Lavras (Brazil) M. Sc. Plant Pathology. 2011
3. Universidade Federal de Viçosa (Brazil) B. Sc. Agronomy Engineer. 2009

PROFESSIONAL EXPERIENCE

1. Chromatin Inc. Collaborative Project: "Response of ten parental sorghum lines to 16 isolates of *Colletotrichum*", University of Kentucky, USA. 2015
2. Graduate Research Assistant, Fungal Genetics, University of Kentucky. 2012-present
3. Graduate Research Assistant, Plant Pathology, Universidade Federal de Lavras, Brazil. 2009-11
4. Graduate Research Assistant, Genetics, Universidade Federal de Viçosa, Brazil. 2006
5. Volunteer trainee, Crop protection. 2004-05

ORAL PRESENTATIONS AT CONFERENCES:

1. The potential impact of anthracnose on sweet sorghum production. APS Southern Division meeting. Balm, Florida, February 20–22, 2016.
2. Genetic diversity of Colletotrichum isolates from *Sorghum bicolor* and *S. halepense* in the Southeastern United States. APS Annual meeting. Tampa, Florida, July 3 - August 3, 2016.
3. Controle da Ferrugem do Cafeeiro com Produtos Alternativos. 31° Congreso Brasileiro de Pesquisas Cafeeiras, 2005, Guaraparí, ES.

SCIENTIFIC ACTIVITIES: Reviewer, Plant Disease.

ABSTRACTS AND PROCEEDINGS:

1. **Xavier, K.,** Pfeiffer, T., Parreira, D., & Vaillancourt, L. (2016, April). The potential impact of anthracnose on sweet sorghum production. In PHYTOPATHOLOGY (Vol. 106, No. 2, pp. 16-16). 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA: AMER PHYTOPATHOLOGICAL SOC.

2. **Xavier, K. V.**, M. F. Torres, E. A. Buiate, I. Gaffoor, S. Chopra, and L. J. Vaillancourt. "Comparison of putative secondary metabolite genes and gene clusters of *Colletotrichum graminicola* and *C. sublineolum*." In PHYTOPATHOLOGY, vol. 102, no. 7, pp. 138-139. 3340 PILOT KNOB ROAD, ST PAUL, MN 55121 USA: AMER PHYTOPATHOLOGICAL SOC, 2012.
3. Waquim, A. P.; **Xavier, K. V.**; Sakiyama, N. S.; Caixeta, E. T.; Zambolim, E. M.; Zambolim, L. Identificação de marcadores moleculares ligados aos genes de resistência do cafeeiro à ferrugem. In: XVII SIC, 2007, Viçosa-MG. XVII SIC, 2007.
4. **Xavier, K. V.; Caixeta**, E. T.; Barbosa, J. C.; Zambolim, E. M.; Zambolim, L.; Sakiyama, N. S. Escolha de primers para identificação de marcadores rapd ligados a resistência do cafeeiro à ferrugem. In: Simpósio de Iniciação Científica (XVI SIC), 2007, Vicosa-MG. Simpósio de Iniciação Científica (XVI SIC), 2007.
5. Braga, P. P.; Zambolim, Laércio; Sousa, A. F.; **Xavier, K. V.**; Souza Neto, P. N. Controle Químico da Ferrugem do Cafeeiro (Hemileia vastatrix Berk.et Br) pela Aplicação de Fungicidas Sistêmicos via Solo e Foliar, Associados com a Aplicação de Fungicida Cuprico. In: XIV Simpósio de Iniciação Científica VI Simpós - Mostra Cientifica da Pós-graduação e IV Simpósio de Extensão Universitária, 2006, Viçosa. XIV Simpósio de Iniciação Científica, 2006.
6. **Xavier, K. V.**; Zambolim, Laércio; Sousa, A. F.; Braga, P. P.; Souza Neto, P. N. Controle da Ferrugem do Cafeeiro Hemileia vastratrix, com Produtos Alternativos. In: XV Simpósio de Iniciação Científica, V Mostra Cientifica da Pós-graduação e III Simpósio de Extensão Universitária Pesquisa, Extensão e Desenvolvimento, 2006, Viçosa, MG. XV Simpósio de Iniciação Científica, V Mostra Cientifica da Pós-graduação e III Simpósio de Extensão Universitária Pesquisa, Extensão e Desenvolvimento, 2006.
7. Braga, P. P.; Zambolim, Laércio; Sousa, A. F.; **Xavier, K. V.** O Contole Químico da Ferrugem do Cafeeiro pela Aplicação de Fungicidas Sistêmicos Via Solo e Foliar, Associados com a Aplicação de Fungicidas Cuprico. In: 31° Congreso Brasileiro de Pesquisas Cafeeiras, 2005, Guaraparí, ES. 31° Congreso Brasileiro de Pesquisas Cafeeiras. Guaraparí, ES: Gráfica e Editora Bom Pastos, 2005. p. 353-354.
8. **Xavier, K. V.**; Zambolim, Laércio; Sousa, A. F.; Braga, P. P.; Souza Neto, P. N. Controle da Ferrugem do Cafeeiro com Produtos Alternativos. In: 31° Congreso Brasileiro de Pesquisas Cafeeiras, 2005, Guaraparí, ES. 31° Congreso Brasileiro de Pesquisas Cafeeiras. Guaraparí, ES: Gráfica e Editora Bom Pastor, 2005. p. 388-390.
9. Barbosa, J. C.; Zambolim, Laércio; Duarte, H. S. S.; **Xavier, K. V.**. Controle Químico da Ferrugem do Feijoeiro nos Sistemas de Plantio Direto e Convencional. In: XIV Simpósio de Iniciação Científica, IV Mostra Científica da Pós-graduação e II Simpósio de Extensão Universitária Pesquisa, Extensão e Desenvolvimento, 2004, Viçosa, MG. XIV Simpósio de Iniciação Científica, 2004. p. 130-130.

FIRST AUTHOR PUBLICATIONS IN PROGRESS:

1. A comparative genomic analysis of putative pathogenicity genes in two host-specific *Colletotrichum* sibling species. Journal: BMC Genomics
2. Impact of *Colletotrichum* strains from cultivated and wild *Sorghum* spp. on the sweet sorghum variety Sugar Drip in field, greenhouse, and laboratory assays. Journal: Plant disease
3. Genetic diversity of *Colletotrichum* isolates from *Sorghum bicolor* and *S. halepense* in the Southeastern United States. Journal: Phytopathology
4. Identification of two new species of *Colletotrichum* infecting *Sorghum bicolor* and *S. halepense* in the Southeastern United States. Journal: Phytopathology

RELEVANT EXTRACURRICULAR ACTIVITIES AND FELLOWSHIPS

1. Reviewer, Plant Disease.
2. Conference/ meetings Leadership (2).
3. Administrative Coordinator of NEFIT (Association of Plant Pathology Scholars – UFLA). 2009-11
4. Fellowships CAPES. 2005-2006; 2009-11

Katia Viana Xavier