Universidad de Guanajuato

Division of Natural and Exact Sciences

Biology department

"Bioinformatics study of elements potentially involved in response to light in *Metarhizium*"

THESIS

TO OBTAIN THE DEGREE OF

MASTER OF SCIENCES

Presents

BSc. Lourdes Lavinia Lavín López

Advisor: PhD. Juan Carlos Torres Guzmán

Co-advisor: PhD. Gloria Angélica González Hernández

UNIVERSIDAD DE GUANAJUATO

Guanajuato, Gto., 01 de octubre de 2021

Dr. Agustín Ramón Uribe Ramírez
Director
División de Ciencias Naturales y Exactas
Universidad de Guanajuato
Campus Guanajuato
Presente

Estimado Dr. Agustín Uribe Ramírez:

Por medio de la presente, hacemos constar que el trabajo de tesis titulado: **"Bioinformatics study of elements potentially involved in response to light in *Metarhizium*"**, fue realizado por la candidata a Maestro en Ciencias; la BSc. Lourdes Lavinia Lavín López, para obtener el grado de Maestro en Ciencias (Biología).

Este trabajo fue desarrollado en el Laboratorio de Genética Molecular de Hongos del Departamento de Biología, de la División de Ciencias Naturales y Exactas, bajo nuestra dirección.

Hacemos constar que el trabajo es original, se realizó con la calidad y el rigor científico requerido; el trabajo está concluido, el manuscrito de tesis ha sido revisado por sus directores de tesis y los miembros de su comité tutorial; por lo que los abajo firmantes, directores de tesis, autorizamos continuar los trámites correspondientes para la obtención del grado de Maestro en Ciencias (Biología).

Sin otro particular aprovechamos la ocasión para enviarle un muy cordial saludo

Atentamente,

Dra. Gloria Angélica González Hernández — Dr. Juan Carlos Torres Guzmán

Directores de tesis

BIO UG

Laboratorio de Genética Molecular de Hongos

Laboratorio de Genética de Molecular de Hongos. Departamento de Biología. División de Ciencias Naturales y Exactas.
Campus Guanajuato
52 (473) 732 00 06 Ext. 8160
www.labgenmolugto.com
Noria Alta s/n | Col. Noria Alta,
Guanajuato Gto. México | CP. 36050

Acknowledgments

First, I thank to my advisors Dr. Juan Carlos Torres Guzmán and Dra. Gloria Angélica González Hernández for allowing me to work in their lab and their infinite patience, initiative and their constant support and the knowledge that they've given me over the years.

I also want to thank my colleagues in the Laboratorio de Genética Molecular de Hongos for their advice, willingness, and encourage during this stage.

In the same way, I thank the Biology Department at the University of Guanajuato for providing me with the necessary resources and tools to carry out this thesis, as well as the scholarship granted by CONACYT (2019-000037-02NACF-27398) that financed this specific project.

Finally, I thank my biological family and my chosen family who have motivated and supported me to follow my dreams. Without your constant encouragement and affection, I wouldn't have achieved this.

Thank you very much to all.

Dedication

I dedicate this work to my parents for allowing me to explore and fall in love with science. They continue to inspire me every day.

And finally, in dedication to women who seek to venture into science, mainly in data science and programming; hopefully in the future we will be more.

General index

Abbreviations

aa	Amino acids
AM	Arbuscular mycorrhizae
BLR1	Blue light regulator 1
BLR2	Blue light regulator 2
BLRC	Blue light response Complex
ccgs	Clock-controlled genes
CD	Conserved Domain
CK	Casein Kinase
CREs	Cis-Regulatory Elements
ELREs	Early Light Response Elements
ENV1	Envoy
EUM1	Envoy Upstream Motif 1
FCD	FRQ-CK1a Interaction Domain
FRH	Frequency-interacting RNA Helicase
FRQ	Frequency protein
Kbp	Kilobase pair
LCB	Locally collinear blocks
LLREs	Late Light Response Elements
LOV	Light and Oxygen Voltage
LRES	Light Response Elements
NLS	Nuclear Localization Signals
nt	Nucleotides
PAS	Per-Arnt-Sim
vvd	Vivid gene
WC1	White Collar 1 protein
WC2	White Collar 2 protein
WCC	White Collar Complex

Figure index

Table index

Abstract

Light is one of the constant abiotic factors on the planet and practically all living organisms present some type of response to the light stimulus. An important aspect after light reception is the gene modulation; this mechanism involves stimulating or inhibiting the expression of certain genes that, possibly, have the purpose of adaptation to different environmental conditions. In fungi, light response has a very important effect on the genes involved in circadian regulation; in this process, the products of the *frq* and *frh* genes regulate the activity of the White-Collar Complex (WCC) in a time-dependent manner.

This phenomenon is widely studied in saprophytic organisms, however; the study of the molecular components and the mechanisms involved in chronobiology and photobiology in entomopathogenic fungi is relatively new. In *Metarhizium* there are reports of phenotypic effects caused by the light stimulus, including growth in constant light or photoperiod; but there is no complete description of the mechanism that leads to these photo-dependent phenotypic characteristics.

Based on the broad distribution of light response genes in filamentous fungi, we performed a bioinformatic analysis on the sub genomic context of said genes using the software Mauve and RStudio. We found that the light response genes appear to be constant in the genomes analyzed demonstrating a micro-syntenic conservation among the *Metarhizium* genus, except for the *wc2* gene. To better understand the possible transcriptional regulation, we analyzed the promoters of the light response genes searching for putative *cis*-regulatory elements, finding conserved motifs in all the *Metarhiuzm* species analyzed; this approach allowed us to identify other possible genes for response to light and circadian control within the genome of this genus.

Finally, we propose different light-dependent mechanisms, in addition to its putative molecular components and the genes involved in these characteristic phenotypes.

1. Introduction

1.1 Fungal photobiology

Light seems to be one of the constant variables in life on earth; every single organism has a response to the photonic stimulus, or it is, at least, photosensitive. Trophic chains are dependent on primary producers: plants and other photosynthetic organisms. Thus, life, as we know, is dependent on the electromagnetic radiation given by our star, the Sun (Hader, 2013). The fungi kingdom is no exception; in these organisms, light controls a plethora of morphologic and physiologic responses; they can sense from the near-UV spectrum to the far-red spectrum, including visible light. Organisms belonging to the fungi clade contain at least 11 photoreceptors in their genome (Yu & Fisher, 2019); these molecules can detect different light stimuli and generate responses inside the cell according to the type of signal that is received. The response generated can lead to changes in the genome, stimulating or inhibiting the expression of determined genes; it is believed that this effect might help the fungi adapt to different abiotic factors; light can act as a signal on environmental conditions (Corrochano, 2019).

Light can also be harmful, UV radiation represents a significant source of stress to the fungal cell; for instance, it can damage the DNA. In that sense, to avoid this damage, cells can produce pigments that might act as a protective barrier that functions absorbing photons or scanning for reactive oxygen species (ROS) generated by light exposure (De Fabo *et al*., 1976).

Several physiological processes have been linked to the light response; for instance, some are correlated to the virulence, sexual development, resistance to harmful conditions, among others. All these processes are related to the exposition to light and even to the type of light that the fungal cell growths (Yu & Fisher, 2019).

There are constant elements in the light response of fungal organisms, mainly photoreceptor proteins. Depends on the type of photoreceptor protein, the subcellular localization varies; in example, blue-light photoreceptors are mainly found in the cytoplasm and nucleus; and green-light photoreceptors are transmembrane proteins (Yu & Fisher, 2019).

The process of the general light response in fungi occurs when a small molecule, a chromophore, absorbs light. This photonic interaction leads to a physicochemical and structural change that favors the formation of a photo adduct with the photoreceptor protein. It is well known that these proteins are conserved in the fungi kingdom (Fuller *et al.,* 2015). Once the photo adduct is formed, a signaling cascade occurs; proteins involved in this cascade act at different levels; for example, photolyases activate and repair damage in the DNA caused by UV light, thanks to the interaction with a chromophore. Other cascades involve changes in genetic expression (Yu & Fisher, 2019). In this case, blue light is the principal stimulus that generates these genetic regulation. The photoreceptor proteins involved in the detection of the blue light stimulus have a dual role; acting also as a transcription factor, that binds to specific sequences called light response elements (LRE's) found in the promoter region of specific genes (Fuller *et al*., 2015).

1.2 Biological models

Regarding fungal photobiology, *Neurospora crassa* has been the star model, where the White-Collar Complex (WCC) function was first described and extensively studied; this complex regulates the genetic expression after the fungi is exposed to light (Fuller *et al*., 2015). WCC is formed by two proteins: White collar-1 (WC1) and White Collar-2 (WC2). WC1 has a dual role: functions as a photoreceptor for blue light; where flavin-adenine dinucleotide (FAD) acts as a chromophore; FAD absorbs light and interacts with WC1 via a specific PAS domain in the protein called light and oxygen voltage (LOV) forming the photo adduct by a cysteine residue (Cys-122) that it's located within the LOV domain. WC1 function as a transcription factor in a similar

way as GATA transcription factors. WC1 contains a zinc finger domain that binds to the DNA; it also has two PAS domains (Per-Arnt-Sim, PER: Circadian rhythm protein, ARNT: Ah-nuclear receptor translocator protein and SIM: Single-minded protein, that regulates the development of the nervous system in *Drosophila sp*.) that are involved in different cellular responses. WC2 has a zinc finger, but does not present a LOV domain; thus, it doesn't participate in light detection. The role of WC2 is to allow the union of WCC to the DNA via the zinc finger (Corrochano, 2007). WC1 isn't the only protein involved in blue-light sensing; there is a small protein that contains a LOV domain that allows the interaction with the FAD chromophore, VIVID (*vvd*). This protein was first described in *N. crassa* and ENVOY (*env1*) is the apparent orthologue in *Trichoderma reesei* (syn. *Hypocrea jecorina*) (Corrochano, 2019). VIVID acts as a regulator of the WCC, generating a negative feedback loop. The WCC strongly induces the expression of *vvd*; accumulation of this protein (VIVID) occurs in the cytoplasm, generating a blockage in the LOV domain of WC1, therefore, inhibiting the formation of the WCC and blocks its function as a transcription factor (Fuller *et al*., 2015).

In *Trichoderma reesei,* light response studies date to the early 1950's where it was first described that this fungus, in a nutrient-rich medium in the dark, can grow as mycelium indefinitely and brief exposition to light induces conidia production (Lilly & Barnett, 1951). This discovery gave rise to different studies on *Trichoderma* light response and its effects and components, finding a system homologous to the White-Collar Complex, commanded by the photoreceptor proteins named Blue Light Regulators. Blue light regulator-1 (BLR1) is homologous to WC1; it also contains two PAS domains, a LOV domain and a zinc finger domain that allows its function as a transcription factor. Blue light regulator-2 (BLR2) has a single PAS domain and a single zinc finger domain, like WC2.

As already mentioned, the apparent ortholog of *vvd* in *T. reesei* is the *env1* gene; however, the function of this protein differs with respect to VIVID*;* instead of acting as an inhibitor of the complex, ENVOY works at different levels, including the regulation of genes that participate in various pathways, from sexual development,

G-protein signaling and other processes independent of blue light regulators (Casas & Herrera, 2013). The effect of ENVOY includes the regulation of a nine gene cluster and genes involved in various metabolic processes (Tisch & Schmoll, 2010). Despite having the same structure, these small proteins' functions differ enormously, which raises the question of the amplitude of the effect of light in living beings, specifically in fungi.

1.3 Circadian rhythms

Organisms are time dependent, being so that almost all organisms have periodical changes in their physiology, metabolism, and behavior resembling the solar cycle. It is proposed that these periodical processes allow the organisms to adapt to different environmental stimuli, thus maximizing their fitness (Andreani *et al.*, 2015). In a nutshell, this trait allows organisms to generate a series of responses in anticipation of short- or long-term environmental changes.

To consider a process as a circadian rhythm, it must meet these characteristics:

i) The periodicity of the biological processes (physiological, metabolic, or behavioral) must resemble the 24-hour solar cycle.

ii) External changes do not affect circadian rhythm; patterns are maintained despite environmental characteristics; that is, it is a self-maintained cycle.

iii) External changes do not interrupt the cycle, but the cycle can adjust to specific changes: light and temperature.

iv) The periodicity is maintained in a wide range of temperatures; this effect is called "temperature compensation."

The phenotypic changes caused by these periodical patterns are generated by the phasic expression of several genes. This process involves a cell-intrinsic transcriptional/translational feedback loop (TTFL). The main characters in this phenomenon are a plethora of transcription factors that promote the expression of genes that performs as inhibitory elements in the loop. Shortly, the circadian rhythms are mediated by the accumulation of the inhibitory elements favoring their translocation to the nucleus, acting on the transcription factors until they are

degraded, releasing the brake on the cycle. The accumulation of the inhibitory elements occurs during the day; thus, the process creates a self-perpetuating gene expression pattern (Andreani *et al*., 2015).

In fungi, the quintessential model for circadian rhythms is *Neurospora crassa* since this phenomenon was described for the first time in this organism. The principal characteristic of this cycle in *N. crassa* is the daily rhythm of conidiation, which occurs approximately every 22-hours, generating a visible pattern and this pattern is maintained in a variety of temperatures, but exposure to blue light resets or abolish the cycle. This biological rhythm is mediated by a protein called frequency (FRQ) that, interestingly, its expression is mediated by the White-Collar Complex (Fuller *et al*., 2015). In a few words, light induces the formation of the WCC, and this induces the expression of FRQ.

FRQ forms a complex with a second protein, FRH (an RNA-helicase that interacts with FRQ); this complex inhibits the formation of WCC; therefore, it modulates itself. For the inhibition of WCC to occur, FRQ must be phosphorylated in the FCD domains (FRQ-CK1a interaction domains) through interaction with kinases, mainly casein-kinase 1 (CK1); this phosphorylation activates FRQ favoring the formation of the complex with FRH. Once formed, it interacts with WCC recruiting kinases that, in turn, phosphorylate WCC preventing the complex from binding to DNA and thus, its role as a transcription factor does not occur (Liu *et al*., 2019).

Neurospora presents two FRQ isoforms: a large one (l-frq, NCBI accession code: AAA57121) and a small one (s-frq; ESA42015); their prevalence and expression is determined by thermal signals (Tong *et al*., 2020).

The transcription of the *frq* gene is mediated by the union of the WCC to the Clock Box, a specific sequence in the promotor region of *frq* in absence of light (Dunlap & Loros, 2006). In the presence of light, WCC recognizes the proximal regulatory element (pLRE), generating an increase in the expression of *frq*. The protein-encoding transcripts appear by late night and early morning. Several chromatin-remodeling events have been proved to control the proper expression of *frq* in both

sites. The co-repressor like RCO-1, is described as core-clock component, necessary for the rhythmicity expression of *frq* controlling the chromatin status (Larrondo & Canessa, 2018). The gene *frq* is one of the most complex loci known in microorganisms. First, a long non-coding antisense *frq* transcript ("*qrf*") is expressed rhythmically at low levels, it reaches a peak phase opposite to that of *frq* (Dunlap & Loros, 2006). Sense *frq* RNAs are generated from two start sites and the promoter can be regulated by light and time of the day. In this process, the first intron can be spliced using one 5' donor site with the same 3' acceptor, and the second intron is alternatively spliced in a thermo-dependent way; this results in six identifiable transcripts, and their abundance reflects the environmental conditions (Diernfellner *et al*., 2005). The second intron, resulting of the alternative splicing process, contains the AUG codon of the ORF; this thermo-influenced splicing generates a mix of the long and short FRQ proteins. However, no different functions for the isoforms have been described (Dunlap & Loros, 2006).

III.a Circadian rhythms in entomopathogenic fungi.

Fungal-insect interactions depend on different adaptations that favor the entomopathogenic organism. Rhythmic air-spore dispersion is an important one since it secures the permanence of the fungi in the environment. In *Entomophtora spp.,* spore levels increase in the last hours of the night, it has been proposed that this response could be modulated by light. Conidia of *Entomophaga maimaiga* appear to be preferentially discharged from dead moth larvae between 2:00 and 8:00 h; it has been suggested that this fungus manipulates the larvae to die in the afternoon so that the sporulation happens at night; this process is mediated by light cues (Larrondo & Canessa, 2018).

In the case of the house fly, the fungus *Entomophaga muscae*, manipulates the fly behavior: in the last hours of life, the fly climbs up and leaves its wings in an up position, providing space for the spores to be launched from the fly's abdomen. Interestingly, the time of the day at which flies stop moving and dies is not random; it occurs at sunset. This behavior is regular in the fungi and the fly only under

photoperiod cycles; flies maintained in dark conditions die at random hours. Posterior transcriptomic analysis in *E. muscae* confirmed the expression of a homolog of WC1; although this fungus does not have a *frq*-like gene, the existence of a functional circadian clock is not dismissed (Larrondo & Canessa, 2018).

Field studies of the interactions between *Ophicordyceps unilateralis* and the ants *Camponotus leonardi;* showed locomotor activity restricted around noon; ants bite plant leaves and then it's killed by the fungus; this behavior is conserved among ant populations. It has been inferred that such specific behavior might be controlled by the fungal internal clock. Genome analysis in the species *Ophicordyceps kimflemingiae* confirmed the presence of *frq, wc1* and *wc2* homologs. Analysis in gene expression of the infected ants *Camponotus castaneous* revealed a different number of oscillating transcripts when the ants were infected by *O. kimflemingiae*; in the photoperiod, over 300 genes showed oscillatory expression, and in dark conditions, only 154 genes presented this oscillation (Larrondo & Canessa, 2018).

In *Beauveria bassiana*, recent reports (Tong *et al.,* 2020) have confirmed the presence of two *frq* genes and two FRQ proteins: FRQ1 and FRQ2. The genome of this entomopathogenic fungus also contains a homolog for the *frh* gene, two putative clock proteins (WCs), the *vvd* gene, and the red/far-red photoreceptor (PHY). FRQ1 (964 aa) contains a large FRQ domain and FRQ2 has two FRQ domains. The analysis performed revealed that the genes of these two FRQ proteins are similarly transcribed, and the transcripts decreased simultaneously with shortening on daylights. The expression of *frq1* and *frq2* appears independent and not the isoform products of a single gene, like the *N. crassa l*-FRQ and *s*-FRQ. The sub-cellular localization of both proteins is time and light dependent, with a difference in the nuclear and cytoplasmatic accumulations. The deletion of both genes generated a delay in conidiation and this conidiation defects were worsened with shortening on daylight lengths (Tong *et al*., 2020). Construction of double deletion mutants failed; therefore, *frq1* or *frq2* is essential for cell viability when the other is absent. The authors propose the importance of these proteins for conidiation in different

photoperiods and maximization of conidial yield for survival and dispersion of *B. bassiana* in host habitats. The other clock-component (*frh*) was non-essential for the viability of *B. bassiana*, but Δf*rh* generated differential repression on four photoreceptor genes; transcript levels of *wc1* and *wc2* were the ones that presented a higher reduction in their expression (Tong *et al*., 2020).

There is no evidence of homologs of *frq* in *Aspergillus*, but *Aspergillus flavus* exhibits rhythms in sclerotia formation, this characteristic is depending on the media composition and persists in dark conditions (Larrondo & Canessa, 2018).

III.b Circadian rhythms in fungi and their interaction with plants

In the mycorrhizal fungi that grow in underground conditions, were light stimulus is restricted, there is strong evidence of the presence of homologs of WCC and actively expressed in *Rhizoglomus irregularis* and the expression of the homolog of *frq* is increased after light pulse (Lee *et al*. 2018). The phytopathogenic fungus *Cercospora kikuchii* displays a circadian rhythm in the hyphal myelinization in photoperiod growth, this characteristic persists under dark conditions (Daub and Ehrenshaft, 2000).

Arbuscular mycorrhizae (AM) are obligatory plant root symbionts that depend on the photosynthesis by host plants for the source of carbon compounds. In exchange the mycorrhizae absorb various nutrients in the soil. The fungi hyphae act as ecological niche for various bacteria and other fungi, forming the backbone of the rhizosphere (Simon *et al.,* 1993). Fascinatingly, AM have conserved clock components in their genome (*frq* and WCC*)*; the conservation of *frq* is notable since there is no reported function for this gene outside its role in circadian rhythm; transcription of this gene in AM was found to be induced by exposure to blue light. Nevertheless, because AM fungi live underground, light and temperature are unlikely to be essential cues unless the organism is close to the soil´s surface. It has been proposed that mycorrhizal fungi may have undergone a selective selection process where carbon compounds secreted by plant roots may act as the main signals that

regulates the circadian clock (Lee *et al.,* 2019). In support of this hypothesis, Hernandez & Allen (2003) described that the hyphal growth pattern in AM has a peak between 12:00 and 18:00 h, which curiously, coincides with the highest photosynthetic activity in plants. It has been inferred that the conservation of this circadian clock component in AM fungi may have an important functional role, allowing interactions with the clock of the host plant, resulting in the coordination of the symbionts´ metabolism (Lee *et al*., 2019).

1.4 *Metarhizium*

The ascomycete and entomopathogenic fungi *Metarhizium* have a cosmopolitan distribution. It has an economic value since it performs as a broad-spectrum bioinsecticide in the agricultural sector. This fungus is not the exception to the process of light response, and various physiological and morphological effects have been shown to have a direct relationship with lighting conditions (Roberts *et al.,* 2004).

The *Metarhizium robertsii* genome has homologous genes to the ones involved in the light response in *N. crassa*, initially identified by Gao *et al*., (2011) in the strain ARSEF 23. Starting with the presence in the genome of White Collar-1 (MAA_04453), while White-Collar 2 is annotated as a cutinase protein palindrome-binding protein (MAA_07440) as well as an ortholog for *vvd* annotated as PAS domain-containing protein (MAA_02250). The successful isolation of this ortholog gene (*cie1*) and its promoter region has been achieved by Aguilar-Gordillo (2010), determining that there are components in the promoter that are specific for light response elements (LRE's), a PAS domain, and the cysteine residue (Cys122) that allows interaction with the FAD chromophore. Gasca (2016) performed the deletion of this gene in the *M. robertsii* CARO4 strain. A hyper-sporulation phenotype was shown in said strain and identified that its cytoplasmic subcellular location is similar to VIVID and ENVOY.

These findings explain the fungu's capacity to sense light and why the fungus presents a plethora of polymorphisms in colony phenotypes determined by light,

exhibiting various strategies that the organism follows. "*Sleeper*" refers to the phenotype where the conidial production is more significant, and the fungus remains in the soil dormant until it forms an interaction with an insect or plant. "*Creeper*" is predominant by the production of mycelium, where the hyphae navigate in the soil looking for an insect or plant to interact. Finally, a "*mixed strategy*" where there is not a noticeable preference. These phenotypes are determinate by the absence of light (Angelone *et al.,* 2018). Oliveira *et al*. (2017) confirmed that conidia production, germination rate, and virulence are strongly related to the type of light in which this fungus is grown. There is an enhancing effect on conidial germination in the presence of white or blue light in contrast to growth in darkness. Similarly, there is an increase in conidia production and virulence (Oliveira *et al.,* 2017).

The effects of the light stimulus do not only apply to phenotypic characteristics. It has been determined that exposing mycelium to visible light increases the tolerance to ultraviolet radiation in *M. acridum*. This outcome is explained by the over-expression of the photolyase coding genes, which participate in the repair of UV-damaged DNA; these enzymes use the energy provided by the light to break the bond formed by two pyrimidines in the DNA caused by the UV radiation (Brancini *et al.,* 2018). Genetic modulations with phenotypic consequences are not the only effects that exposure to light generates. Brancini *et al*., (2019) realized that this outcome interferes with changes in the mRNA and protein abundance; surprisingly, one of the downregulated proteins are subunits of the translation initiation factor 3 and other ribosomal proteins. The authors infer that this result might be caused because the fungus perceives light as a stress factor; therefore, there is an apparent reduction in the translation activity.

Considering that the fungus can grow underground, outside the visible light spectrum, prolonged exposure to light is considered a stressor and gives rise to the genomic, transcriptomic, and proteomic repercussions observed.

2. Justification

The study of fungal photobiology and chronobiology has been restricted to other filamentous fungi. Although there are extensive studies on the phenotypic response generated by the light stimulus, there is no complete description in entomopathogenic fungi that involves the role of FRQ in the regulation of this biological pathway.

The response to light and the circadian cycles are highly related, thus, *mrfrq* gene's role is of utmost importance since it is the link in both processes. Therefore, determine the possible function and regulation of the *frq* gene in *Metarhizium robertsii* might bring us closer to understanding these pathways.

Contributing to the description of the elements of these routes may allow us to understand the diversity of phenotypic traits described in *Metarhizium.*

3. Hypothesis

Circadian regulation similar to the process described in *N. crassa* may be occurring within *Metarhzium,* coupling to the described phenotypes that are dependent on the light stimulus.

4. Objectives

4.1 Main Objective

Determine the putative elements involve in light response and its link with circadian cycle of *Metarhizium* through bioinformatic approach.

4.2 Specific objectives

a) Analyze the homologous elements to those found in *N. crassa* and *T. reesei* involved in light response in different species of *Metarhizium.*

b) Design and implement a computational algorithm for comparative genomic analysis using sub-genomic regions where genes for response to light and circadian regulation can be found.

c) Identification of putative in *cis* regulatory elements for genes involved in these processes.

d) Identify possible clock-controlled and light-responsive genes that might be dependent on the light stimulus.

5. Materials and methods

5.1 Data collection and homology search

A search was performed in the NCBI database (https://www.ncbi.nlm.nih.gov) to identify homologous proteins involved in the light response and circadian rhythms described in *N. crassa.* Once all the elements were identified, a sequence alignment was executed using BLASTP 2.11.0+ (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

N. crassa frq gene, OR74A, ID 3876095.

Table 1. BLASTP results for the FRQ-like proteins

Species & strain	Description	E value	Percentage Identity	Accession number
***M. brunneum* ARSEF 3297**	Frequency clock protein	0.0	47.45%	XP_014547790.1
***M. robertsii* ARSEF23**	Frequency clock protein-like protein	0.0	47.78%	XP_007820960.2
***M. acridum* CQMa 102**	Frequency clock protein	0.0	46.76%	XP_007808256.1
***M. anisopliae* ARSEF 549**	Frequency clock protein	0.0	47.44%	KID60831.1
***M. rileyi* RCEF 4871**	Frequency clock protein	0.0	48.99%	OAA45406.1
***M. guizohuense* ARSEF 977**	Frequency clock protein	0.0	47.87%	KID88336.1
***M. majus* ARSEF 297**	Frequency clock protein	0.0	47.87%	KID98159.1
***M. album* ARSEF 1941**	Frequency clock protein	0.0	48.52%	KHN97182.1
***T. reesei* QM6a**	Frequency clock protein	0.0	48.58%	XP_006965164.1
***B. bassiana* ARSEF 2860**	Frequency clock protein	0.0	45.33%	XP_008594847.1

N. crassa wc1 gene, OR74A, ID 3875924.

Table 2. BLASTP results for WC1-like proteins.

Species & strain	Description	E value	Percentage Identity	Accession
***M. brunneum* ARSEF 3297**	White collar 1	0.0	55.03%	XP_014547561.1
***M. robertsii* ARSEF23**	White collar 1	0.0	55.47%	XP_007820740.2
***M. acridum* CQMa 102**	White collar 1	0.0	57.07%	XP_007808025.1
***M anisopli ae* ARSEF 549**	White collar 1	0.0	54.95%	KID62337.1
***M. rileyi* RCEF 4871**	White collar 1	0.0	75.87	OAA49498.1
***M. guizohuense* ARSEF 977**	White collar 1	0.0	55.58%	KID92568.1
***M. majus* ARSEF 297**	White collar 1	0.0	59.01%	KIE01538.1
***M. album* ARSEF 1941**	White collar 1	0.0	57.35%	KHO01137.1
***T. reesei* QM6a**	Blue light regulator	0.0	55.20%	XP_006965752.1
***B. bassiana* ARSEF 2860**	White collar 1	0.0	59.53%	XP_008603590.1

N. crassa wc2 gene, OR74A 3879968.

Table 3. BLASTP results for WC2-like proteins

Species & strain	Description	E value	Percentage Identity	Accession
***M. brunneum* ARSEF 3297**	Cutinase protein palindrome-binding protein	1e-175	71.88%	XP_014541775.1
***M. robertsii* ARSEF23**	Cutinase protein palindrome-binding protein	1e-174	71.59%	XP_007823629.2
***M. acridum* CQMa 102**	Cutinase protein palindrome-binding protein	2e-151	60.68%	XP_007815884.1
***M. anisopliae* ARSEF 549**	Cutinase protein palindrome-binding protein	7e-176	72.16%	KID70129.1
***M. rileyi* RCEF 4871**	Cutinase protein palindrome-binding protein	6e-171	59.18%	OAA48551.1
***M. guizohuense* ARSEF 977**	Cutinase protein palindrome-binding protein	6e-175	71.59%	KID89572.1
M. *majus* ARSEF 297	GATA-type sexual development transcription factor NsdD	1e-10	49.06%	KID99887.1
***M. album* ARSEF 1941**	Cutinase protein palindrome-binding protein	3e-173	71.47%	KHN98899.1
***T. reesei* QM6a**	Blue light regulator 2	6e-142	49.31%	XP_006966680.1
***B. bassiana* ARSEF 2860**	Cutinase protein palindrome-binding protein	2e-171	68.68%	XP_008594722.1

N. crassa vvd gene, OR74A ID 3873728.

Table 4. BLASTP results for VIVID-like proteins

Species & strain	Description	E value	Percentage Identity	Accession
***M. brunneum* ARSEF 3297**	Element involved in conidiation	2e-47	41.58%	XP_014547267.1
***M. robertsii* ARSEF23**	PAS domain containing protein	2e-51	43.46%	XP_007818439.1
***M. acridum* CQMa 102**	Cellulose signaling associated protein ENVOY	8e-47	48.20%	XP_007809797.1
***M. anisopliae* ARSEF 549**	Element involved in conidiation	2e-40	45.71%	KID67646.1
***M. rileyi* RCEF 4871**	Cellulose signaling associated protein ENVOY	5e-53	50.0%	OAA51832.1
***M. guizhouense* ARSEF 977**	Element involved in conidiation	1e-50	43.98%	KID92866.1
***M. majus* ARSEF 297**	Element involved in conidiation	1e-51	43.98%	KIE03548.1
***M. album* ARSEF 1941**	Cellulose signaling associated protein ENVOY	8e-40	47.48%	KHO01717.1
***T. reesei* QM6a**	Glycoside hydrolase family 15, cellulose signaling associated protein ENVOY	1e-45	38.34%	XP_006968888.1
***B. bassiana* ARSEF 2860**	Vivid PAS protein VVD	2e-41	36.51	XP_008596195.1

N. crassa frh gene, OR74A, ID 3872445.

Table 5. BLASTP results for FRH-like proteins

Species & strain	Description	E value	Percentage Identity	Accession
***M. brunneum* ARSEF 3297**	ATP-dependent RNA helicase DOB1	0.0	68.68%	XP_014543126.1
***M. robertsii* ARSEF23**	ATP-dependent RNA helicase DOB1	0.0	69.51%	XP_007824773.1
***M. acridum* CQMa 102**	ATP-dependent RNA helicase DOB1	0.0	69.16%	XP_007812146.1
***M. anisopliae* ARSEF 549**	ATP-dependent RNA helicase DOB1	0.0	68.77%	KFG79672.1
***M. rileyi* RCEF 4871**	ATP-dependent RNA helicase DOB1	0.0	69.05%	OAA37629.1
***M. guizhouense* ARSEF 977**	ATP-dependent RNA helicase DOB1	0.0	68.24%	KID83565.1
***M. majus* ARSEF 297**	ATP-dependent RNA helicase DOB1	0.0	68.24%	KID96355.1
***M. album* ARSEF 1941**	ATP-dependent RNA helicase DOB1	0.0	74.26%	KHN95494.1
T. *reesei* QM6a	Nuclear exosomal RNA helicase	0.0	70.18%	XP_006966062.1
***B. bassiana* ARSEF 2860**	DSHCT domain-containing protein	0.0	69.83%	XP_008595815.1

Once the homologous proteins were identified in the different species and strains used in this project, a selection was carried out to obtain the nucleotide sequences of each protein to perform a multiple-sequence alignment using the CLUSTALW algorithm in the MEGA software. Once the alignment was obtained, a phylogeny was generated, using the Maximum likelihood estimation for the genes *frq, vvd, frh, wc1* and *wc2.*

5.2 Sub-genomic context of the *frq* gene.

To better understand the localization of the *frq* gene in the filamentous fungi genomes, a search was carried out, looking for the immediate adjacent genes of the target gene (Figure 1), the orientation of the gene in the different genomes and the length of the intergenic regions, upstream and downstream (Figure 2). The query was performed in the NCBI database.

Figure 1. Representation of the location of the adjacent genes of *frq*.

Figure 2. Localization of the upstream and downstream regions of the *frq* gene.

5.3 Putative functional domains in FRQ and FRH.

5.3a Putative functional domains in FRQ

FCD domain identification

FCD domains are essential for the phosphorylation process in FRQ, these domains allow the interaction with CK1, the main kinase that phosphorylates FRQ. Liu *et al*., (2019) characterized these interaction domains (α helix containing domains), FCD1 (aa319-326) and FCD2 (aa488-495) in *N. crassa*.

Since there's no actual reports on the characteristics of FRQ in *Metarhizium,* a search of these domains was performed. The aminoacidic sequences of FRQ were collected form NCBI for each species (including FRQ1, FRQ2 in *B. bassiana* and the two isoforms FRQ-l, FRQ-s in *N. crassa*). A multiple-sequence alignment was carried out using the tool Clustal Omega in the database EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The results were obtained and visualized using the Jalview software (Waterhouse *et al*., 2009).

Nuclear localization signal (NLS)

FRQ's activity depends on its transport to the nucleus which will inactivate the WCC. Therefore, a search for nuclear localization signals was performed using FRQ-l in *N. crassa* as a template. To do this, the NLS Mapper (http://nls-mapper.iab.keio.ac.jp) program was used. The program searches the amino acid sequences of proteins for regions that can be recognized by the αß importin pathway. Specifically, this software mainly looks for the canonical sequences for the NLS, being the putative consensus sequence: (K/R) (K/R) X10-12 (K/R)3/5 where X represents any amino acid and K/R 3/5 represents at least three of either lysine or arginine out of five consecutive amino acids. In this way, NLS mapper calculates the scores within the amino acid sequence, a high score is indicative of a strong NLS activity (8,9, or 10 is indicative of an exclusive nuclear location, 7, 8 indicate a partial nuclear location, 3, 4, or 5 indicate both nuclear and cytoplasmic locations and finally 2 or 1 indicate a cytoplasmic location exclusively).

Conserved domains in FRQ

To determine the conserved domains in the FRQ proteins a CD (conserved domain) search was carried out at NCBI (https://www.ncbi.nlm.nih.gov/Structure). Using the conserved domains database (CDD), this tool compares the amino acid sequences of different proteins and matches them to the already identified domains that can be found in protein structures. Only LAMs (LA motifs) where found in the aminoacidic sequences of FRQ in *N. crassa (*both small and large isoforms) and in *T. reesei* but no LAM was found in *M. robertsii*. This domain allows the interaction with RNA and typically co-occurs with an RNA-recognition motif (Lu *et al*., 2020).

FFC domain

FFC (FRQ-FRH complex) Guo *et al*., (2010) determined that a specific region (773 to 783 aa) in the FRQ sequence is necessary for its interaction with the FRH protein, resulting in the formation of the FFC complex. In FRQ-l sequence in this region is PDDHFVMLVTTR.

5.3b Putative functional domains in FRH

No information is available for the function and characteristics of FRH homologous proteins in *Metarhizium*, therefore, a bioinformatic search for putative functional domains in the identified FRH-like protein in *M. robertsii* ARSEF 23 was performed. Since the homologous proteins had already been identified (this work), the identification of the relevant domains for FRH was performed by a multiple alignment of the amino acid sequences by Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/); the results were visualized using the software Jalview. Subsequently, the conserved domains were identified with the CDD tool (https://www.ncbi.nlm.nih.gov/Structure) for RNA helicases belonging to the species of *N. crassa*, *B.bassiana*, *T. reesei* and *M. robertsii*.

The Walker A box allows the interaction of the helicase with ATP; since these helicases are ATP dependent, this domain is crucial for its function. The DEAD box domain is involved in the RNA metabolism. Lauinger *et al*. (2014) determined that

these domains are essential for a stable FFC domain and FRQ's subsequent phosphorylation. The C-terminal helicase domain has a role in RNA degradation, processing, and splicing pathways. KOW Mtr4 is an inserted domain in Mtr4 domain, cooperates with the eukaryotic nuclear exosome in RNA processing (Lu *et al.,* 2020).

5.4 Synteny analysis

Genome alignment (Mauve) and synteny visualization.

To develop the synteny analysis, a genome or sub-genomic region alignment must be done. To achieve this goal, the software Mauve (http://darlinglab.org/mauve/mauve.html) was employed to generate genome-wide analysis using the progressive Mauve and the algorithms that the software comprises. This software identifies conserved regions that appear to be internally free from genome rearrangement, called locally collinear blocks (LCBs). Mauve then selects anchors in the genome and then applies a progressive MUSCLE global alignment to each LCB (Figure 3). Mauve provides a guide tree calculated by the sub-genomic region given by the data (Darling *et al*., 2010). An alignment using 50 kbp sub-genomic regions of the 11 species was performed, obtaining these regions of the Ensembl Fungi Database (http://fungi.ensembl.org), centering the regions with the target gene of each species. Using the dataset obtained by the alignment, an analysis was performed using the software R and the package genoPlotR (Guy *et al*., 2010). Data is given by two objects that can be read by the package: `dna_seg` and `comparisons`. Using specialized commands for the archives generated by Mauve (`.backbone` and `.guide_tree`), the data was retrieved, and different plots were generated.

Using the guide tree generated by comparing the sub-genomic regions (50 kbp) paired with the length and comparison of the homologous regions in said data, new plots were generated to illustrate the phylogenetic relationship evidenced by synteny.

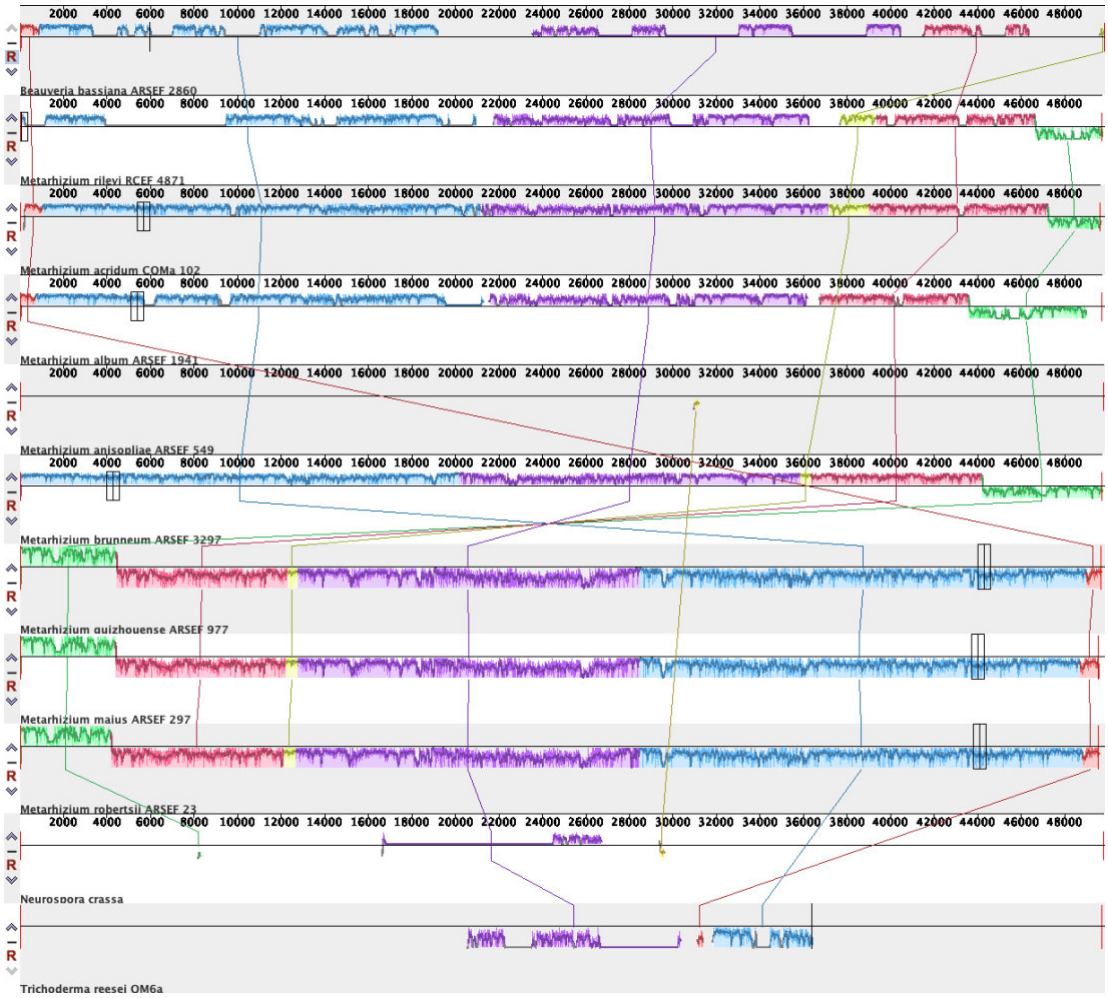


Figure 3. Alignment of sub-genomic regions (50kbp) of 11filamentous fungi.

(N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941)
Colors represent individual LCBs.

5.5 Promoter characterization

To determine the *cis*- regulatory elements of the light response genes (*wc1, wc2, vvd/env1, frq, frh),* the adjacent genes to *frq* and early and late response genes, a visual inspection was carried out in the upstream region of the putative ORF, identifying specific elements involved in different biological processes.

Table 6. Cis-regulatory elements (CRE's)

Element	Consensus sequence	Function & reference.
TATA Box	TATAAA	Transcription factors and RNA polymerase II bind to the TATA box in specific order; therefore, involved in the transcription process (Kwak and Lis 2013).
Early light response element (ELRE)	GATC	Motif found in genes involved in early (transcription inducted 5 min. after the light stimulus) response to the light stimulus (Chen-Hui *et al*., 2009).
Late light response element (LLRE)	TGA---TCA	Motif found in genes involved in late (transcription inducted 15 min. after the light stimulus) response to the light stimulus (Chen-Hui *et al*., 2009)
ENVOY Upstream binding motif 1 (EUM1)	CTGTGC—CTGTGC	Putative protein binding site, involved in the light-mediated genetic expression (Rosales *et al*., 2006)
Glucose response Element (GRE)	CACGTG	Binding site for transcription factors. Involved in the gene expression of glucose-regulated genes. (Hasegawa *et al*., 1999)
Stress Response Element (STRE)	AGGG	Mediates the transcriptional induction by different stress forms. Binding site for the TF's: MSN2, MSS4 and HSF1 (Francisco Estruch, 2000).
WC1 binding site	(A/T)GATA(A/G)	Binding site of the transcription factor WC1; involved in the light-mediated transcription (Qiyang *et al*., 2006)

5.5a LREs conservation in light response genes.

To determine the presence and conservation of said LREs in the promoters of the light response genes, a search was made using MEME Suite (https://meme-suite.org/meme/tools/meme), this software allows the identification of conserved motifs among sequences and performs a sequence alignment. This analysis generates figures than illustrate the specific localization of the LREs in the promoters.

5.6 Quantitative analysis on consensus sequences in the promoter region.

We performed a compilation to differentiate the number of consensus sequences that were identified by the promoter characterization. The illustration was made using the ggplot2 package in Rstudio.

5.7 Putative clock-controlled (ccgs) genes in *M. robertsii* ARSEF23

To determine the putative clock-controlled genes in the genomes of the *species M. robertsii* strain ARSEF 23; first a bibliographic search was performed searching for the ccgs identified in *N. crassa* (Bell-Pendersen *et al*., 1996) and upregulated genes in *Metarhizium* under the constant light-treatments (Brancini *et al*., 2018).
For the ccgs identified *in N. crassa* a protein-BLAST was performed against the available *M. robertsii* ARSEF 23 genome in the NCBI database. Then, the results were grouped, and a promoter characterization of each gene was performed, searching for putative CREs. For the up-regulated genes identified by Brancini *et al.* (2018) a search was performed in the NCBI database, using the gene ID reported. The promoter characterization was manually performed for each gene.

6. Results

6.1 Homology and phylogenetic analysis.

Several phylogenetic trees were generated to determine the phylogenetic proximity from the different genes involved in response to light. First, the nucleotide sequences of each of them were obtained (*wc1, wc2, vvd*/*env1*, *frq* and *frh*). Subsequently, a multiple sequence alignment was carried out using the ClustalW tool in the Mega package software, the phylogenies were generated from the analysis of maximal likelihood.

The maximum likelihood estimate determines for the probability within the generated trees so that, finally, bootstrap pseudo-replicates are generated; for these analyses 500 pseudo-replicates were chosen for the phylogenetic reconstructions.

Maximum likelihood phylogenetic tree (Figure 4) for the *frq* genes in filamentous fungi, with 0.10 nucleotide substitution in the sequences retrieved for this specific analysis; exemplifies the expected grouping of the *Metarhizium* clade with an apparent closeness to *B. bassiana*, an entomopathogenic fungi.

wc1 presents a higher conservation rate than the tested genes, this can be perceived in the phylogenetic tree (Figure 5) generated where the bootstrap values are high. The branches longitude in the *Metarhizium* clade are small, indicating a small value in nucleotide changes.

In the phylogenetic reconstruction of the *wc2* genes, *M. majus* GATA-type sexual transcription factor is placed far from the *Metarizhium* clade (Figure 6), this is to be expected since in the remaining *Metarhizium* species this gene is described as a cutinase protein palindrome binding protein.

The phylogenetic reconstruction groups the clade *Metarhizium* depending on the identification of genes homologous to *vvd* (Figure 7*)*. In the first group we can observe those species that present this gene identified as element involved in conidiation, (*M. brunneum, M. robertsii, M. anisopliae, M. guizhouense and M.*

majus) while the second group is associated with cellulose signaling associated protein ENVOY (*M. album, M. acridum and M. rileyi*). This separation seems to be maintained in posterior analyzes, even though there is a homology of all these genes with the VIVID protein as described in the identification of homologous elements in *Metarhizium* and BLASTP results (Table 4).

For the *frh* genes, the phylogenetic reconstruction appears to be constant, maintaining the *Metarhizium* clade grouped (Figure 8) and with the lowest rate of nucleotide substitution (0.050) among the analyzed genes.

There is evidence of the elements involved in the light detection and circadian rhythms are conserved in *Metarhizium;* therefore, we can infer these fungi have the capability to react to the light stimulus, even though that *Metarhizium* is found in the rhizosphere.

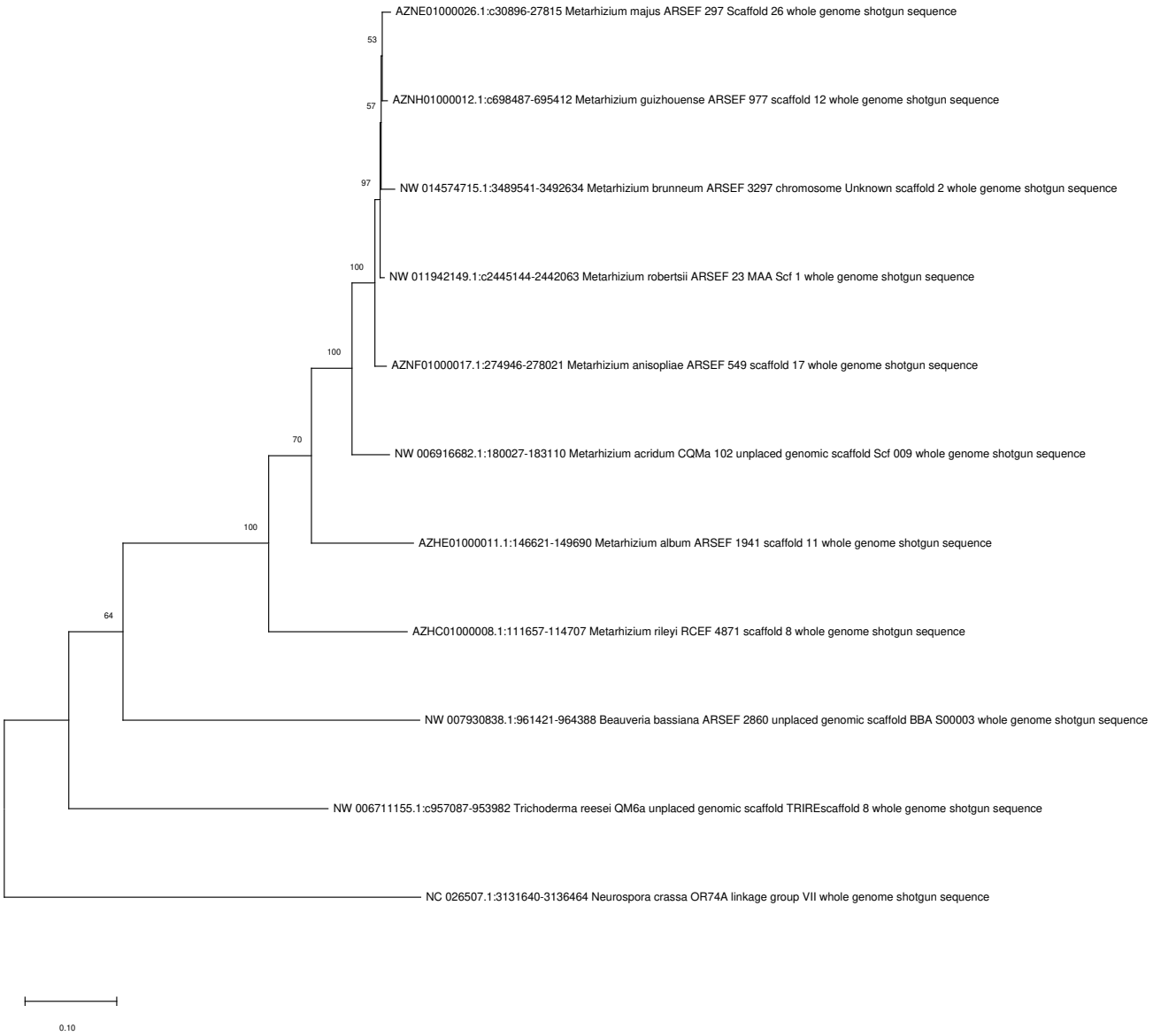


Figure 4. Maximum likelihood phylogenetic tree based on the alignments of the *frq* genes in different fungi species.

The bar scale indicates 0.10 changes per nucleotide. Bootstrap values are shown and were calculated with 500 pseudo-replicates.

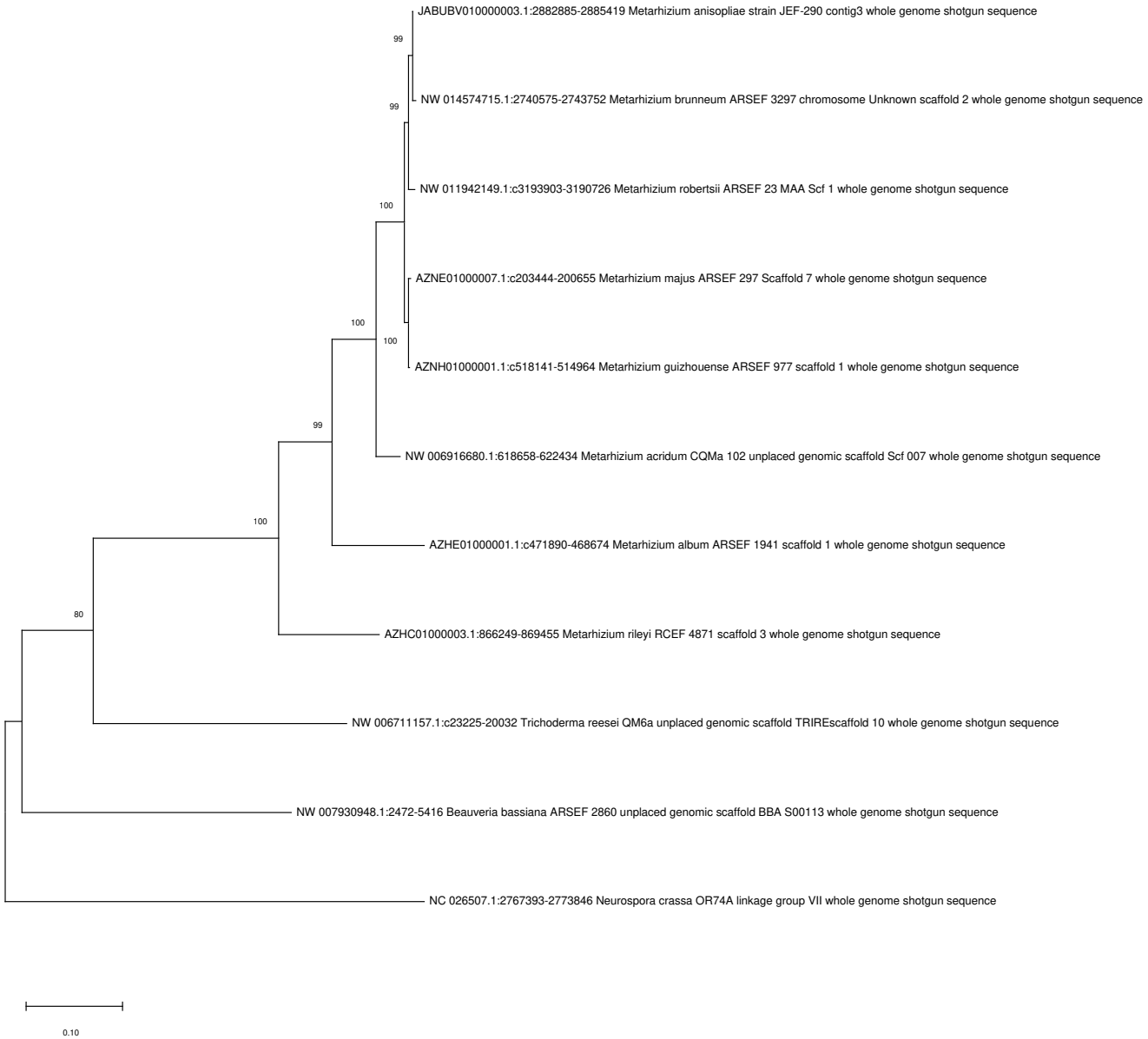


Figure 5. Maximum likelihood phylogenetic tree based on the alignments of the *wc1* genes in different fungi species.

The bar scale indicates 0.10 changes per nucleotide. Bootstrap values are shown and were calculated with 500 pseudo-replicates.

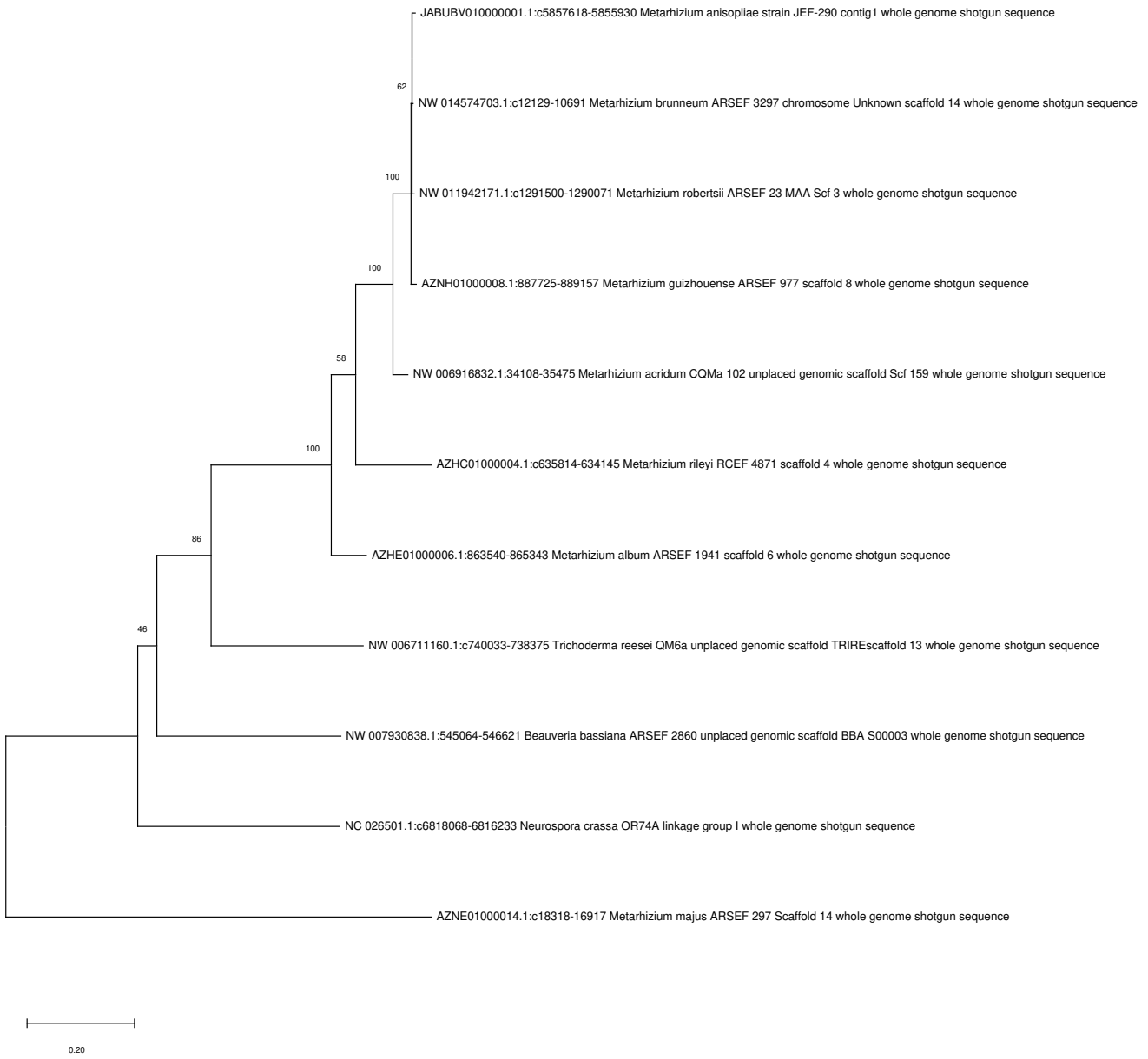


Figure 6. Maximum likelihood phylogenetic tree based on the alignments of the *wc2* genes in different fungi species.

The bar scale indicates 0.20 changes per nucleotide. Bootstrap values are shown and were calculated with 500 pseudo-replicates.

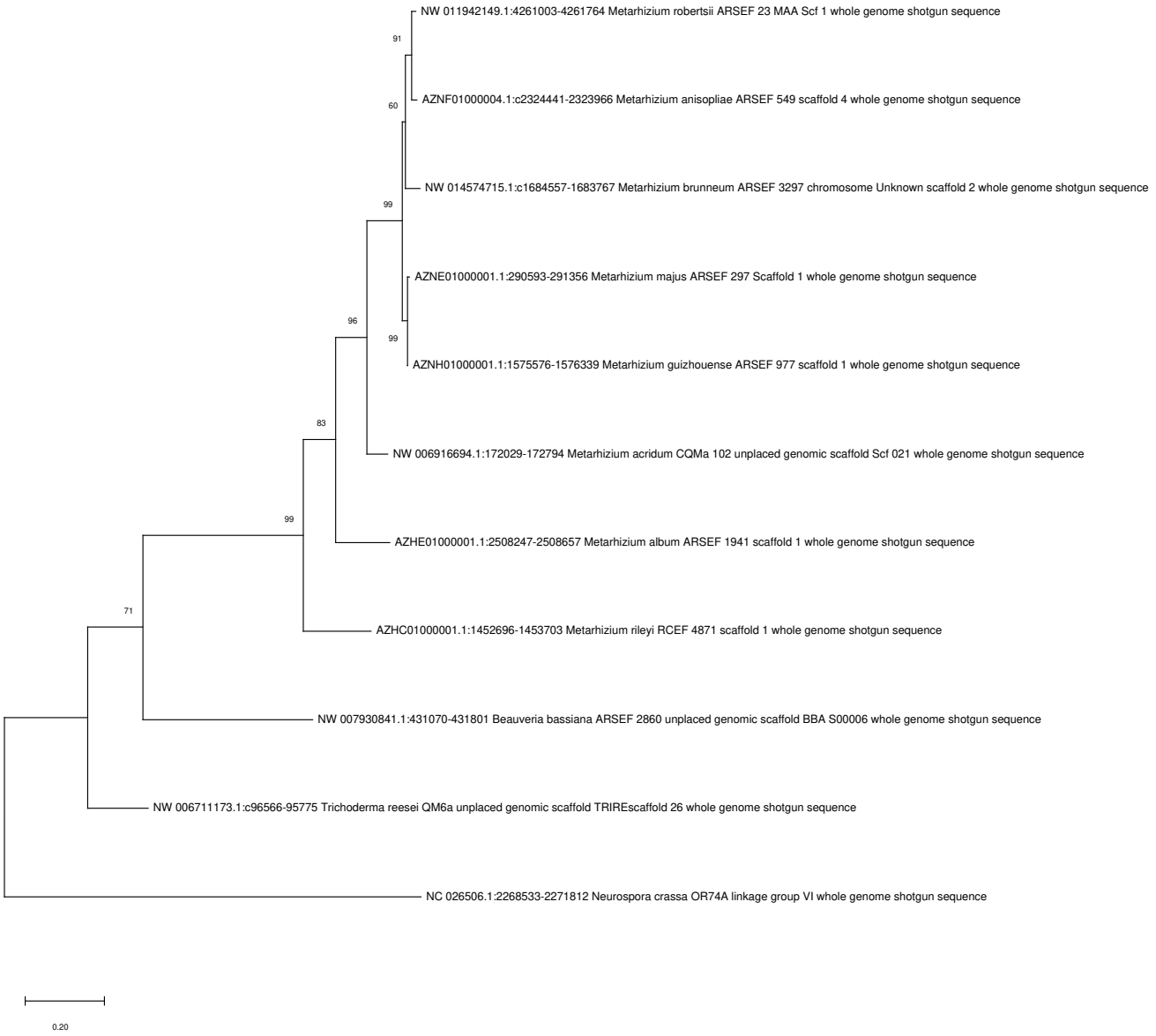


Figure 7.Maximum likelihood phylogenetic tree based on the alignments of the *vvd/env1* genes in different fungi species

The bar scale indicates 0.20 changes per nucleotide. Bootstrap values are shown and were calculated with 500 pseudo-replicates.

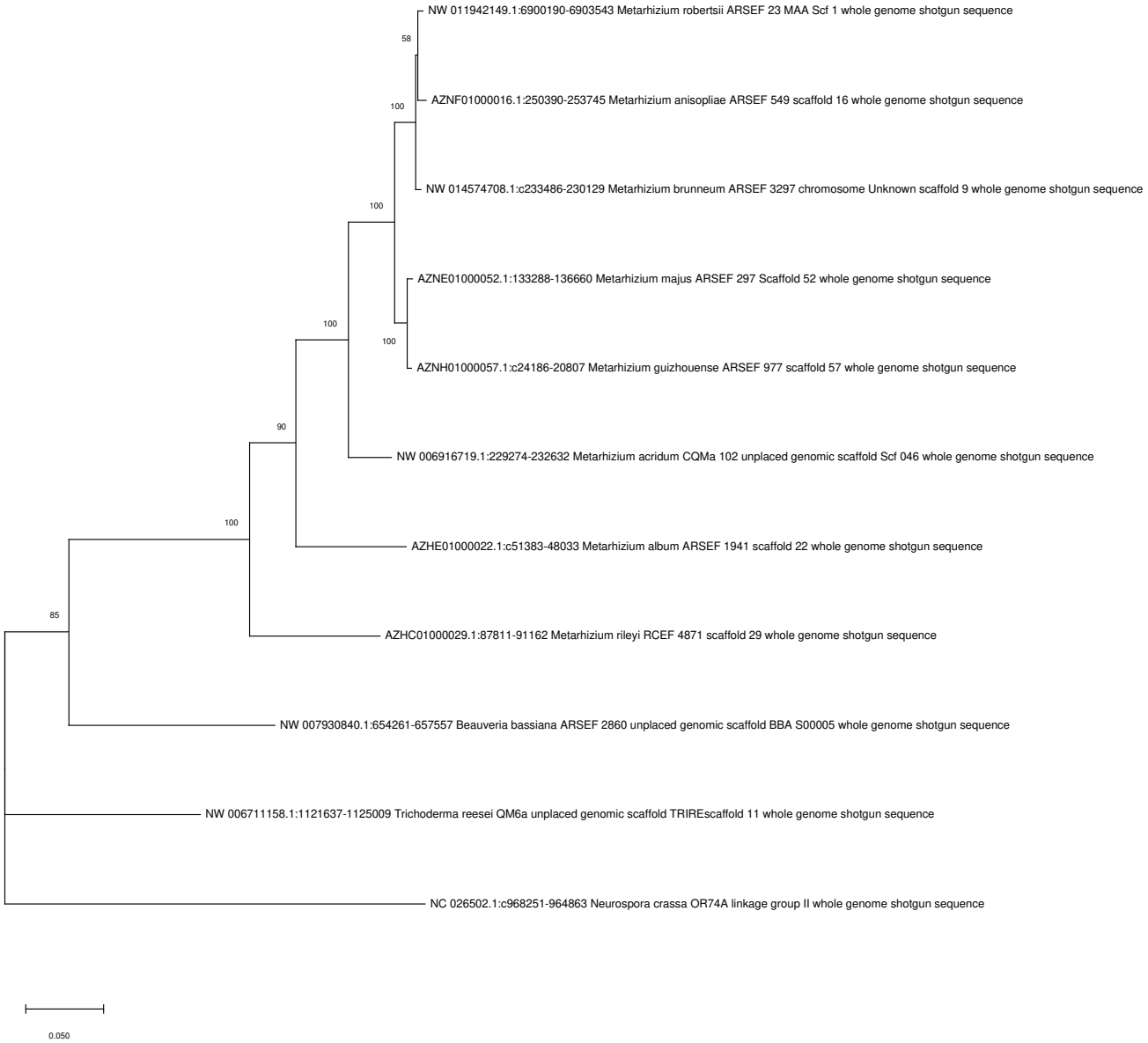


Figure 8. Maximum likelihood phylogenetic tree based on the alignments of the *frh* genes in different fungi species.

The bar scale indicates 0.050 changes per nucleotide. Bootstrap values are shown and were calculated with 500 pseudo-replicates.

6.2 The *frq* gene sub-genomic context.

A search was made in the NCBI (https://www.ncbi.nlm.nih.gov) and Fungi Ensembl (http://fungi.ensembl.org/index.html) databases to determine the *frq* sub-genomic context.

The *Metarhizum* clade has a clear sub-genomic conservation, this is since the genes adjacent to *frq* are similar in all the analyzed species of this genus and in *T. reesei* (Tables 7 and 8*)*. A phosphoglycerate mutase is located on the left flank of *frq*, which participates in glycolysis (Galperin *et al*.,1998) and on the right flank a prenyl-cysteine oxidase that participates in protein degradation (Herrera *et al.,* 2018). These adjacent genes are necessary for basic cell functions, and among these we can find *frq*, the possible main circadian regulator in *Metarhizum* and other fungi.

Table 7. Summary on the sub-genomic context of the *frq*-like genes in filamentous fungi.

a) Gene identification and comparison of the adjacent genes of *frq* (green: similitude between species, yellow & orange: non-similar genes adjacent to *frq*).

b) Orientation on the gene *frq* in different species.

Species & strain	Left flank	FRQ	Right flank	Orientation of FRQ gene
Neurospora crassa OR749A	NCU_02267 Mitochondrial protein FMP25	NCU_02265 Frequency protein	NCU_02264 Prefoldin subunit 3	→
Trichoderma reesei QM6a	TRIREDRAFT_77656 Phosphoglycerate mutase	TRIREDRAFT_121670 Frequency protein	TRIREDRAFT_47814 Uncharacterized protein (pblast= 97.45% similarity with Prenyl oxidase *T. reesei*)	←
Metarhizium robertsii ARSEF23	MAA_04672 Phosphoglycerate mutase	MAA_04673 Frequency protein	MAA_04674 Prenyl-cystein oxidase	←
Metarhizium album ARSEF1941	MAM_04778 2-3, biphosphoglycerate-independent phosphoglycerate mutase	MAM_04779 Frequency protein	MAM_04780 Prenyl cysteine oxidase	→
Metarhizium acridum CQMa102	MAC_01915 2,3-phosphoglycerate-independent phosphoglycerate mutase	MAC_01916 Frequency protein	MAC_01917 Prenyl cysteine oxidase	→
Metarhizium anisopliae ARSEF549	MAN_09558 2,3-phosphoglycerate-independent phosphoglycerate mutase	MAN_09579 Frequency protein	MAN_09580 Prenyl cysteine oxidase	←
Metarhizium brunneum ARSEF 3297	MBR_03082 2,3-phosphoglycerate-independent phosphoglycerate mutase	MBR_03083 Frequency protein	MBR_03084 Prenyl cysteine oxidase	→
Metarhizium guizohuense ARSEF 977	MGU_0475 2,3-phosphoglycerate-independent phosphoglycerate mutase	MGU_0746 Frequency protein	MGU_0477 Prenyl cysteine oxidase	←
Metarhizium majus ARSEF297	MAJ_05789 2,3-phosphoglycerate-independent phosphoglycerate mutase	MAJ_05790 Frequency protein	MAJ_05791 Prenyl cysteine oxidase	←
Metarhizium rileyi RCEF 4871	NOR_03194 2,3-phosphoglycerate-independent phosphoglycerate mutase	NOR_03195 Frequency protein	NOR_03196 Prenyl cysteine oxidase	→
Beauveria bassiana ARSEF 2860	BBA_08956 Tubulin beta chain	BBA_08957 Frequency protein	BBA_08958 Casein kinase epsilon 1	→

Table 8. Adjacent genes to *frq*

(Green: similitude between species, yellow & orange: non-similar genes flanking *frq*). Length of the CDS, products and intergenic regions upstream and downstream of the *frq* gene in different species. (nt=nucleotides, aa=amino acids)

Species & strain	Left flank	Upstream region (nt)	FRQ	Downstream region (nt)	Right flank
Neurospora crassa OR749A	NCU_02267 CDS 2344 nt 606 aa	7324	NCU_02265 4825 nt 989 aa	879	NCU_02264 558 nt 255 aa
Trichoderma reesei QM6a	TRIREDRAFT_77656 1062 nt 533 aa	4905	TRIREDRAFT_121670 3045 nt 1014 aa	461	TRIREDRAFT_47814 1671 nt 566 aa
Metarhizium robertsii ARSEF23	MAA_04672 1728 nt 575 aa	3782	MAA_04673 3018 nt 1005 aa	1118	MAA_04674 1701 nt 566 aa
Metarhizium album ARSEF1941	MAM_04778 1593 nt 530 aa	1679	MAM_04779 3006 nt 1001 aa	896	MAM_04780 1740 nt 579 aa
Metarhizium acridum CQMa102	MAC01915 1728 nt 575 aa	3717	MAC01916 3018 nt 1005 aa	1244	MAC01917 1647 nt 548 aa
Metarhizium anisopliae ARSEF549	MAN_09558 1728 nt 575 aa	3762	MAN_09579 3012 nt 1003 nt	1057	MAN_09580 1701 nt 566 aa
Metarhizium brunneum ARSEF 3297	MBR_03082 1728 nt 575 aa	3747	MBR_03083 3021 nt 1066 aa	1077	MBR_03084 1701 nt 566 aa
Metarhizium guizohuense ARSEF 977	MGU_0475 1728nt 575 aa	3719	MGU_0746 3012 nt 1003 nt	1108	MGU_0477 1701 nt 566 aa
Metarhizium majus ARSEF297	MAJ_05789 1494 nt 497 aa	3717	MAJ_05790 3018 nt 1005 aa	1050	MAJ_05791 1701 nt 566 aa
Metarhizium rileyi RCEF 4871	NOR_03194 1596 nt 531 aa	4364	NOR_03195 2988 nt 955 aa	1267	NOR_03196 1704 nt 567 aa
Beauveria bassiana ARSEF 2860	BBA_08956 1344 nt 447 aa	1679	BBA_08957 1752 nt 583 aa	892	BBA_08958 378 nt 125 aa

6.3 Putative functional domains in FRQ and FRH

6.3a Putative functional domains in FRQ

The FCD domains are highly conserved among the analyzed *Metarhizium* species (Figure 9b). Similarities are visible in the aminoacidic sequences between FRQ-I and the putative FRQ proteins in *Metarhizium*; therefore, is possible that a similar interaction with kinases happens in these FCD domains in the *Metarhizium* species. The analyzed sequences demonstrate that various NLS are found in the FRQ-like proteins. High scores are perceptible in all the analyzed species, therefore, a similar import to the nucleus might be happening in the FRQ homologous proteins in *T. reesei* and *M. robertsii* orchestrated by the αß importin pathway (Figure 10).

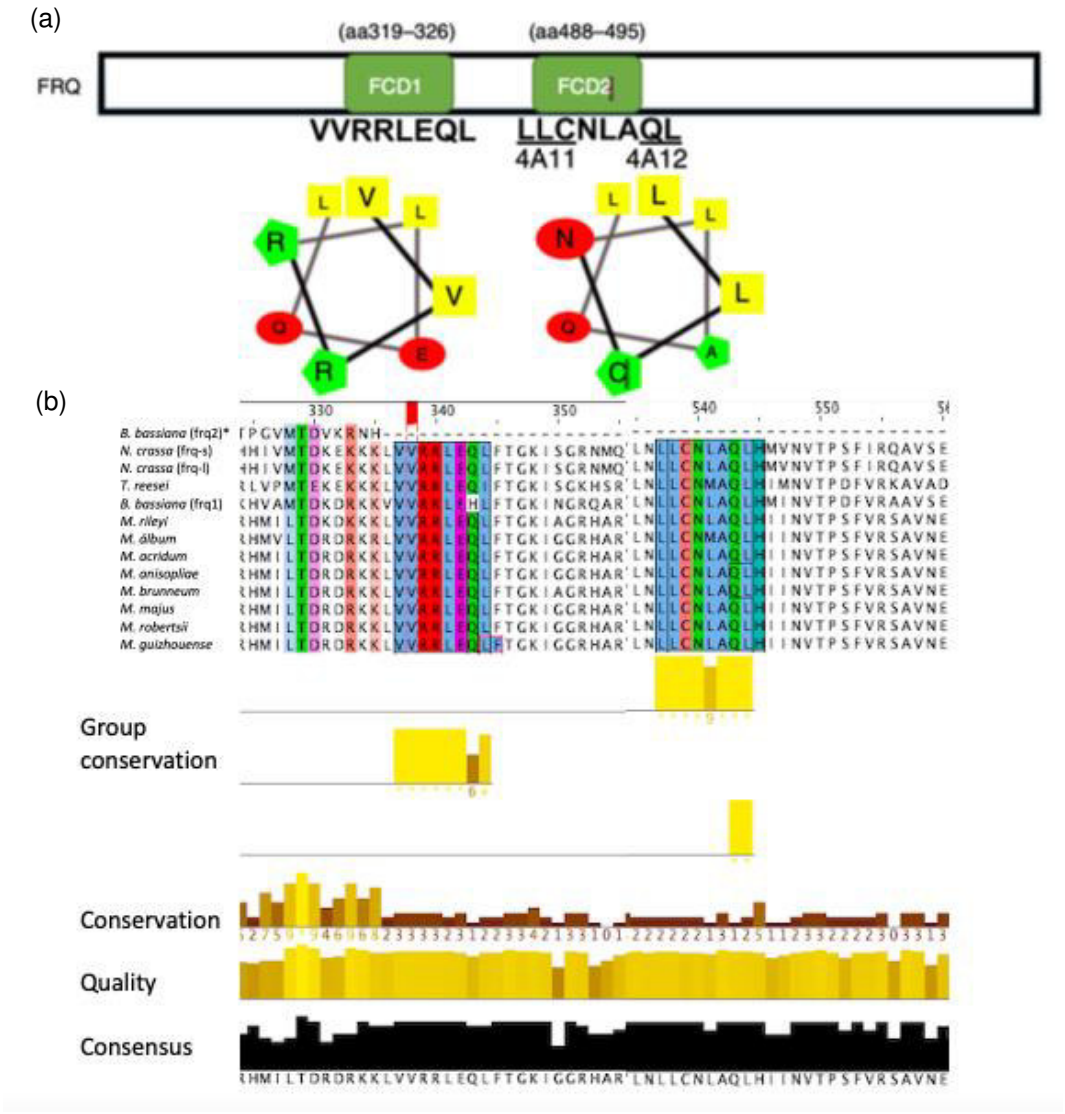


Figure 9.Localization of functional domains in FRQ.

(a) FRQ's FCD domains and aminoacidic sequences identified by Liu et al., (2019) in *N. crassa*
(b) Alignment results visualized using Jalview. The specific aminoacidic sequence for each FCD domain and the group conservation can be detected. FCD1 (336-343) and FCD2 (536-543) appear to be highly conserved in the genome from the analyzed *Metarhizium* species.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

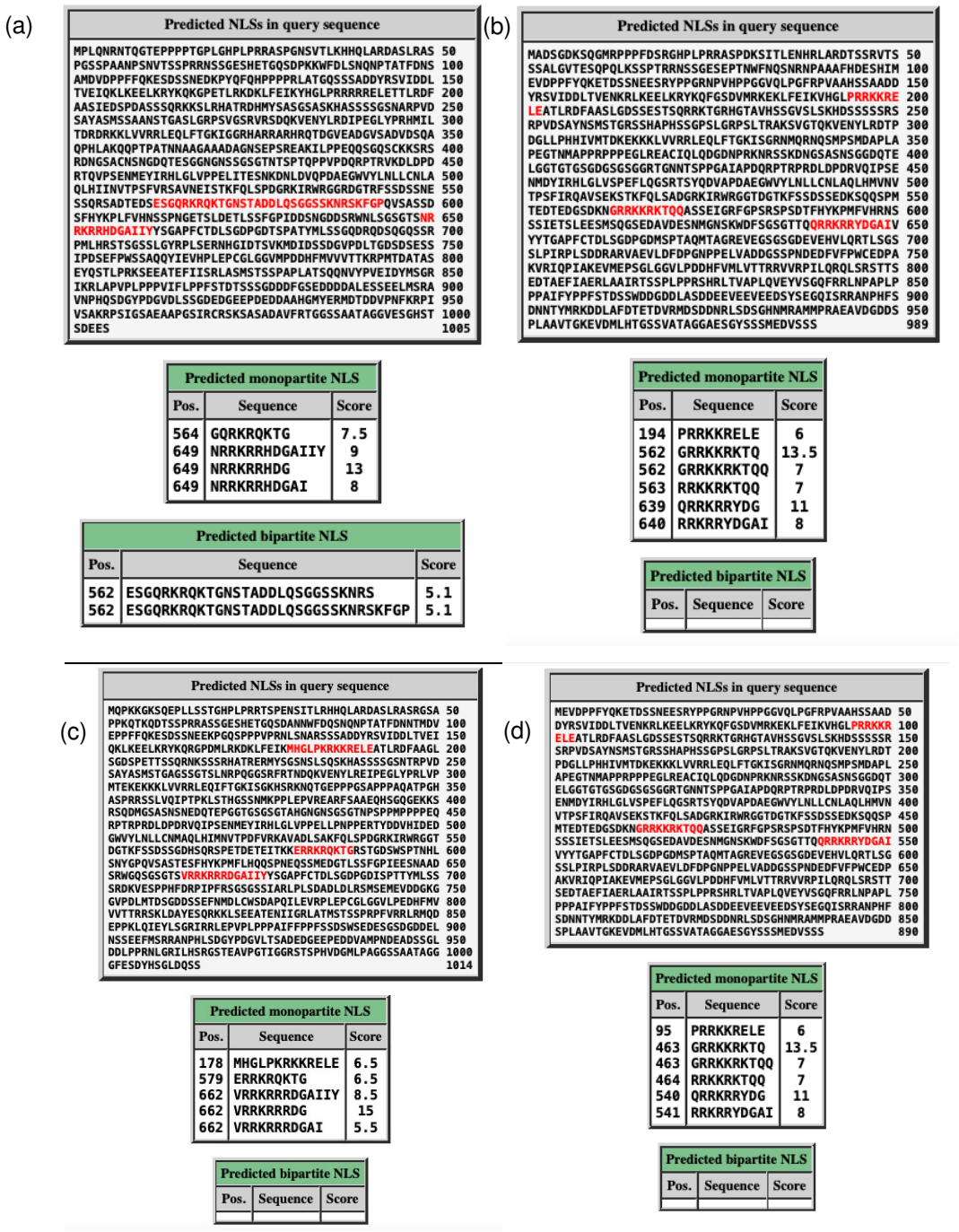


Figure 10. NLS Mapper results.

(a) Predicted NLS in FRQ (*Metarhizium robertsii* ARSEF23).

(b) Predicted NLS in FRQ-l (*Neurospora crassa* OR74A).

(c) Predicted NLS in FRQ (*Trichoderma reesei* QM6a).

(d) Predicted NLS in FRQ-s (*Neurospora* crassa OR74A).

Table 9. Identification of putative functional domains in FRQ-like proteins.

Domain, sequence, and reference.	**FRQ variant 2 *N. crassa* OR74A (Accession XP_011395124.1) 890 aa**	**FRQ variant 1 *N. crassa* OR74A (Accession XP_011395126.1) 989 aa**	**Frequency clock protein *T. reesei* QM6a (Accession XP_006965164.1) 1014 aa**	**Frequency-like clock protein *M. robertsii* ARSEF23 (Accession XP_007820960.2) 1005 aa**
FCD1 VVRRLEQL (Liu *et al.*, 2019)	219-226 aa	319-326 aa	310-317 aa	309-316 aa
FCD2 LLCNLAQL (Liu *et al.*, 2019)	388-395 aa	488-495 aa	507-514 aa	595-502 aa
NLS	463aa	639 aa	662aa	649 aa
LAM	380-427 aa	479-526 aa	496-545 aa	*
FFC PDDHFVMLVTTR (Guo *et al.*, 2010)	672-683 aa	773-782 aa	794-805 aa	780-791 aa

*Domain not found in the protein.

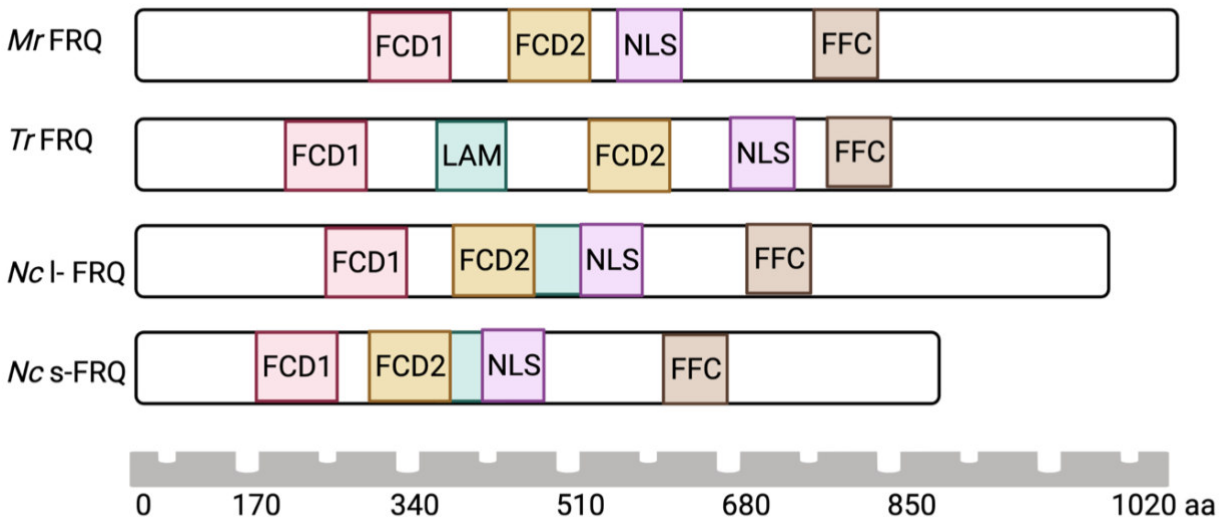


Figure 11. Representation of the putative functional domains in FRQ-like proteins

Mr FRQ (Frequency-like clock protein. *M. robertsii* ARSEF23).

Tr FRQ (Frequency clock protein. *T. reesei* QM6a).

Nc l-FRQ (FRQ variant 1. *N. crassa* OR74A).

Nc s-FRQ (FRQ variant 2 *N. crassa* OR74A)

With the localizations and the presence of the putative functional domains in the proteins analyzed (Table 9), we were able to generate a graphic representation (Figure 11) of these molecules as well as the domains that we found using the illustration tool BioRender (https://biorender.com).

In the specific case of FRQ-like protein in *M. robertsii* ARSEF23, the LAM domains are absent, they are necessary for the protein to interact with RNA, although until now there are no reports of the interaction of FRQ with RNA molecules in *N. crassa*. Despite this, this FRQ-like protein has an NLS value of 13, so we could expect its transport within the nucleus (Figure 11).

6.3.b Putative functional domains in FRH

The alignments made to the protein sequences show the presence of the Walker A and DEAD box domains (Figure 12 b); this is to be expected since these are essential for the functions carried out by an ATP-dependent RNA helicase, as is the case with all identified homologous proteins. However, these domains are also necessary for the stable formation of the FFC, for this; we can infer that, in case of the interaction with the FRQ protein; these domains may be crucial in species where their role in circadian regulation has not yet been reported; that is, in the *Metarhizium* genus.

Despite not having been characterized by their interaction with FRQ, these RNA helicases in *Metarhizium* can function in a similar way to what happens with FRH in *N. crassa* since they have these structural similarities, including the specific domain which allows direct interaction with FRQ (identified as Helicase C domain in the Figure 13).

(a)

(b)

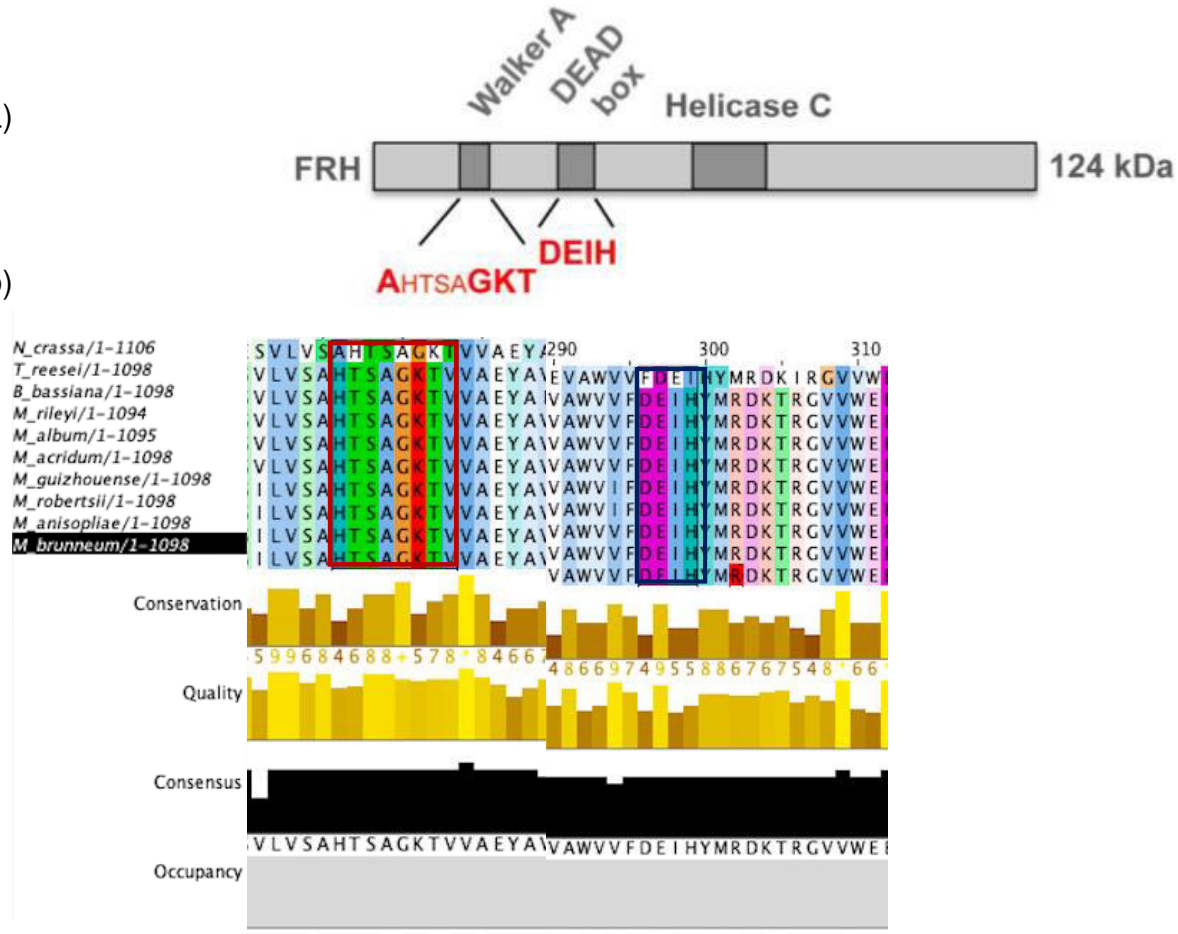


Figure 12. FRH-like domain identification.

Specific localization of the Walker A Box and DEAD Box within FRH-like proteins in *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

(a) Representation of FRH domains in *N. crassa* (Lauinger *et al*., 2014)

(b) Alignment results and Walker A and DEAD box identification in FRH-like proteins in filamentous fungi.

Table 10. Identification of putative functional domains in FRH-like proteins.

Domain, sequence, and reference.	**FRQ interacting RNA helicase *N. crassa* OR74A (XP_956298) 1106 aa**	**DSHCT domain containing protein *B. bassiana ARSEF 2680* (Accession XP_008595815) 1098 aa**	**Nuclear exosomal RNA helicase *T. reesei* QM6a (Accession XP_006966062) 1098 aa**	**ATP-dependent RNA helicase DOB1 *M. robertsii* ARSEF23 (Accession XP_007824773) 1098 aa**
Walker A Box (AHTSAGKT) Lauinger *et al.*, 2014	202-209 aa	186-193 aa	186-193 aa	183-190 aa
DEAD Box (DEIH) (Lauinger *et al.*, 2014)	293-296 aa	277-280 aa	277-280 aa	274-277 aa
C-terminal Helicase domain	353-594 aa	337-576 aa	337-576 aa	334-574 aa
KOW Mtr4	707-841aa	689-833 aa	689-883 aa	686-833 aa

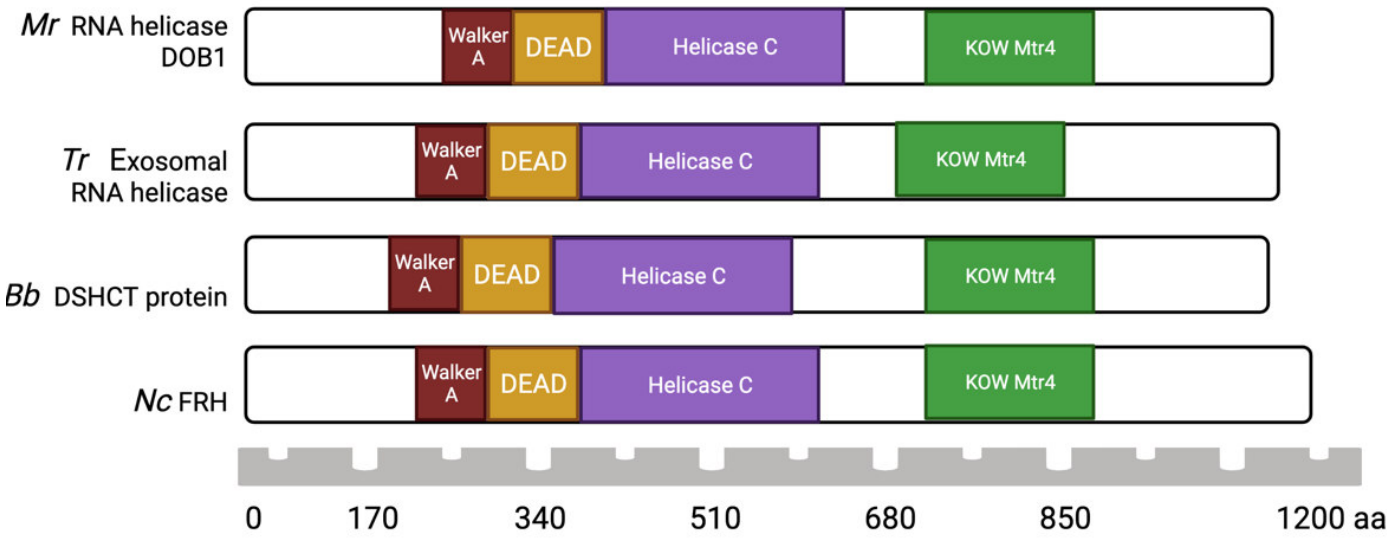


Figure 13. Representation of the putative functional domains in FRH-like proteins

Walker A box a DEAD box are characteristic RNA helicase domains, necessary for the FFC stabilization. Helicase C domain allows the specific interaction with FRQ.

Mr RNA helicase DOB1 (*M. robertsii* ARSEF23).
Tr Exosomal RNA helicase(*T. reesei* QM6a).
Bb DSHCT protein (*B. bassiana* ARSEF 2680).
Nc FRH (*N. crassa* OR74A)

6.4 Synteny analysis on light-response genes.

Synteny Plots (genoPlotR)

To generate the synteny visualization, genoPlotR obtains and compares the data set given by Mauve; the DNA segments are no longer genes but Mauve blocks; therefore, the comparison reflects the correspondences between these blocks.

As a result of these analyzes, a free access programming manual (https://github.com/lavinialavin/Synteny_GenoPlotR) was generated where the specific instructions and commands used in this specific analysis are explained.

wc1 Gene

To perform a micro-synteny analysis, a sub genomic region (50 kbp) was selected, centered in the *wc1* and their homologous in the genomes of the specific species and strains *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941. A progressive alignment was carried out with the Mauve software, where homologous regions were identified in each species. Using the software RStudio and the package GenoPlotR different plots were obtained. For *B. bassiana* only 18 kbp were available from the data base (FungiEnsembl) but the progressive and global alignment were made.

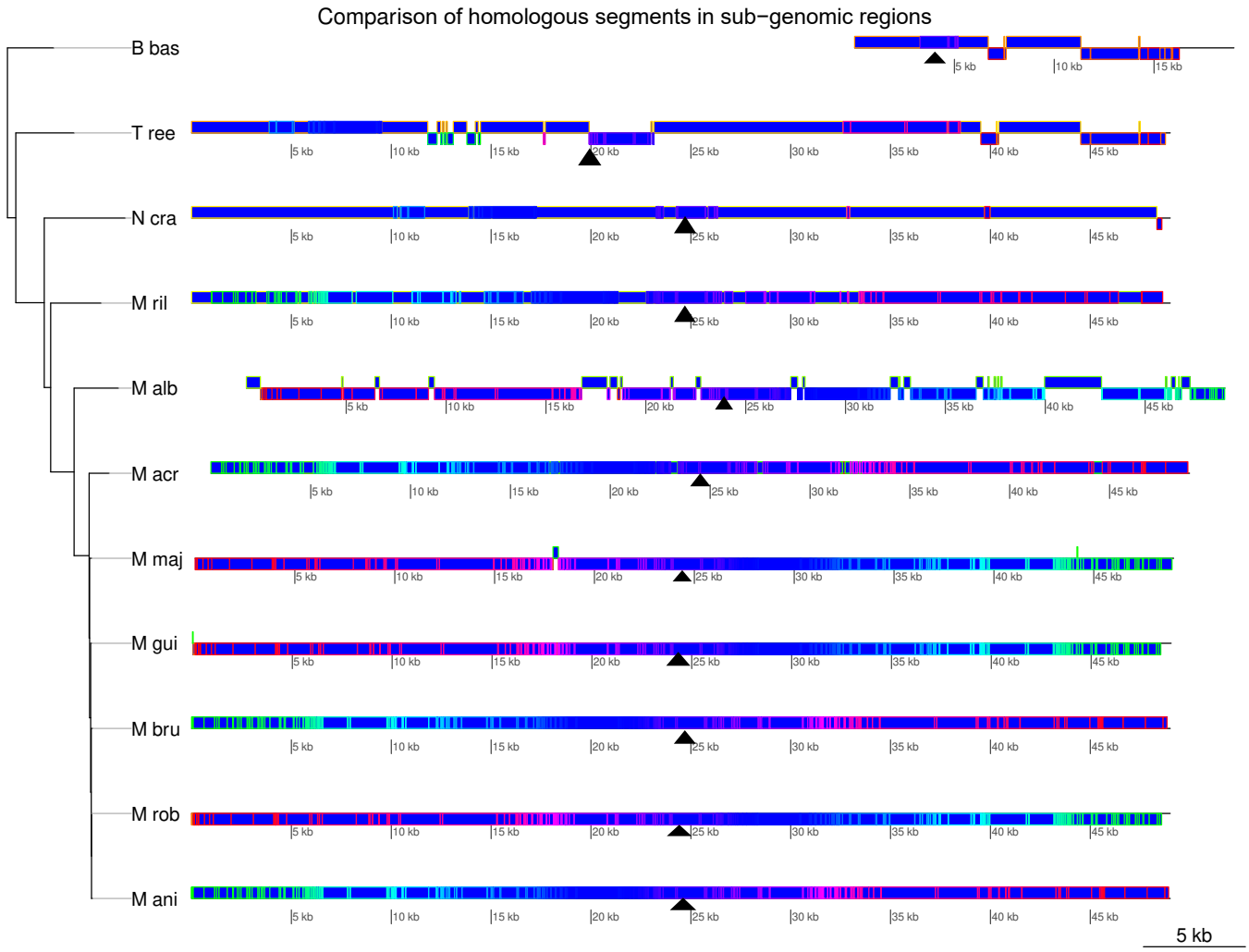


Figure 14. *wc1* gene and block localization in each sub-genomic region.

The comparison between the LCBs are similar, therefore, a color gradient bellow the LCBs can be appreciated. *N. crassa* OR74A, T. *reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941.

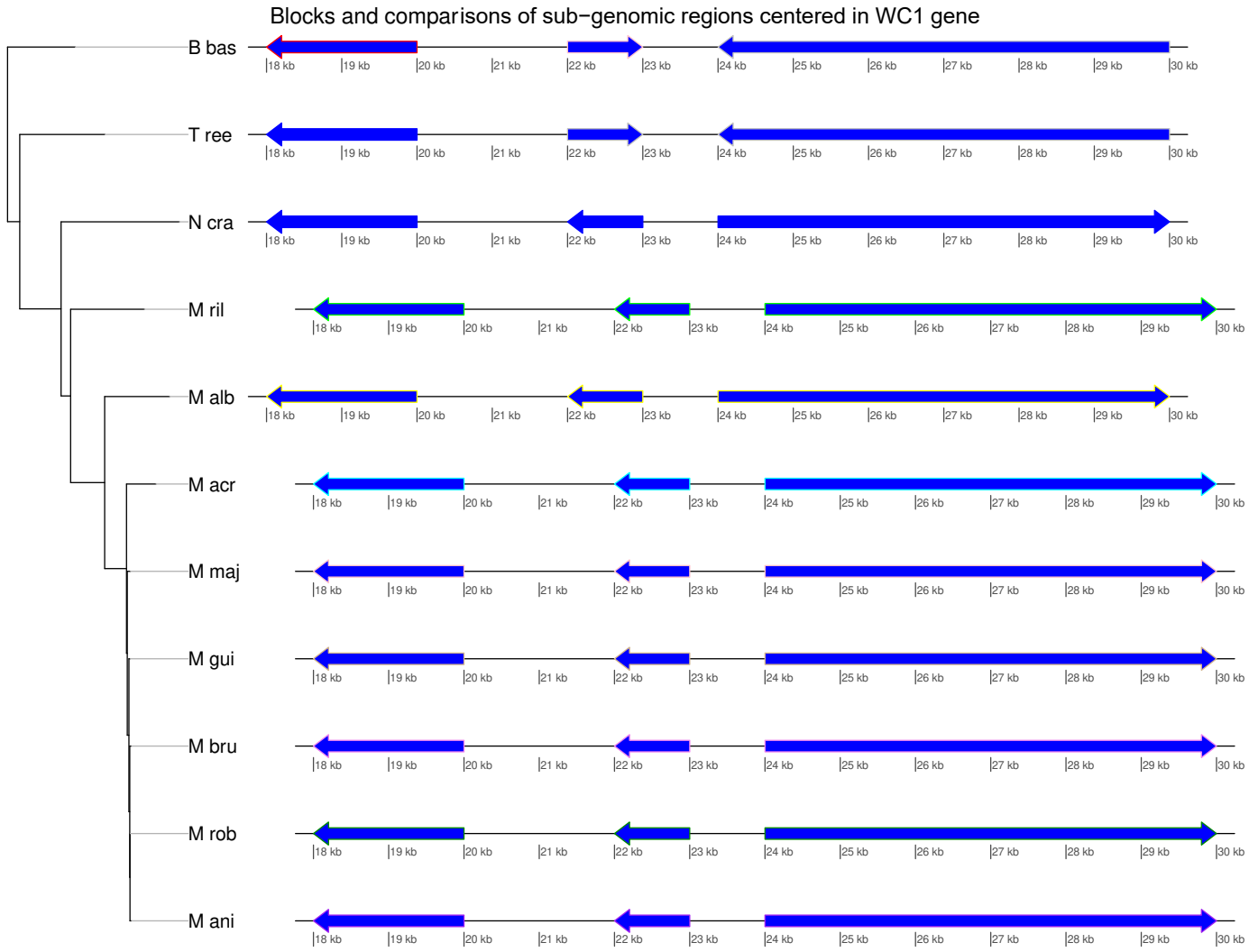


Figure 15. *wc1* block localization.

A phylogeny was obtained as a result from the progressive and global alignments. The biggest arrow corresponds to the block were the *wc1/blr1* gene is located in each sub-genomic region.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

The similarity in the block were the *wc1* gene locates are visible by the comparison analysis (red and blue coloration under the LCBs in Figures 14 and 15); therefore, this sub genomic region is comparable and highly similar in each species except for *T. reesei* since the *blr1* gene was selected for this analysis. The sub-genomic regions compared within the *Metarhizium* genus show a microsynteny pattern determined by the co-linearity characteristic given by this analysis

wc2 Gene

Sub-genomic regions of 50 kbp were selected for this analysis from the genomes of the specific species and strains *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297*, M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297*, M. album* ARSEF 1941. The available contigs vary in size for several species, therefore the *wc2* gene could not be centered within 25 kbp as in previous analyzes, that is; the location of *wc2* varies. Therefore, adjustments were generated within the original code to calculate comparative analyzes in said sub-genomic regions.

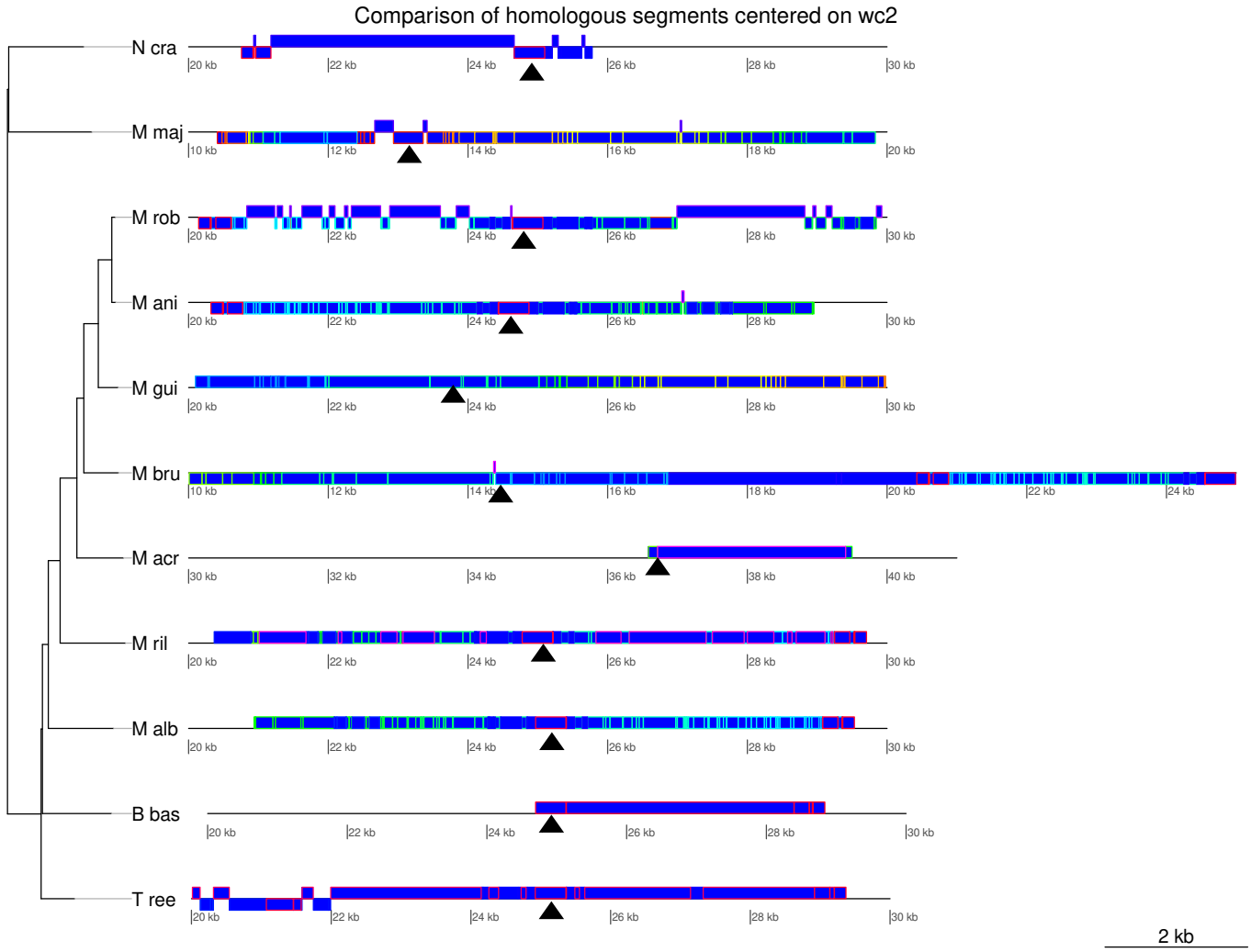
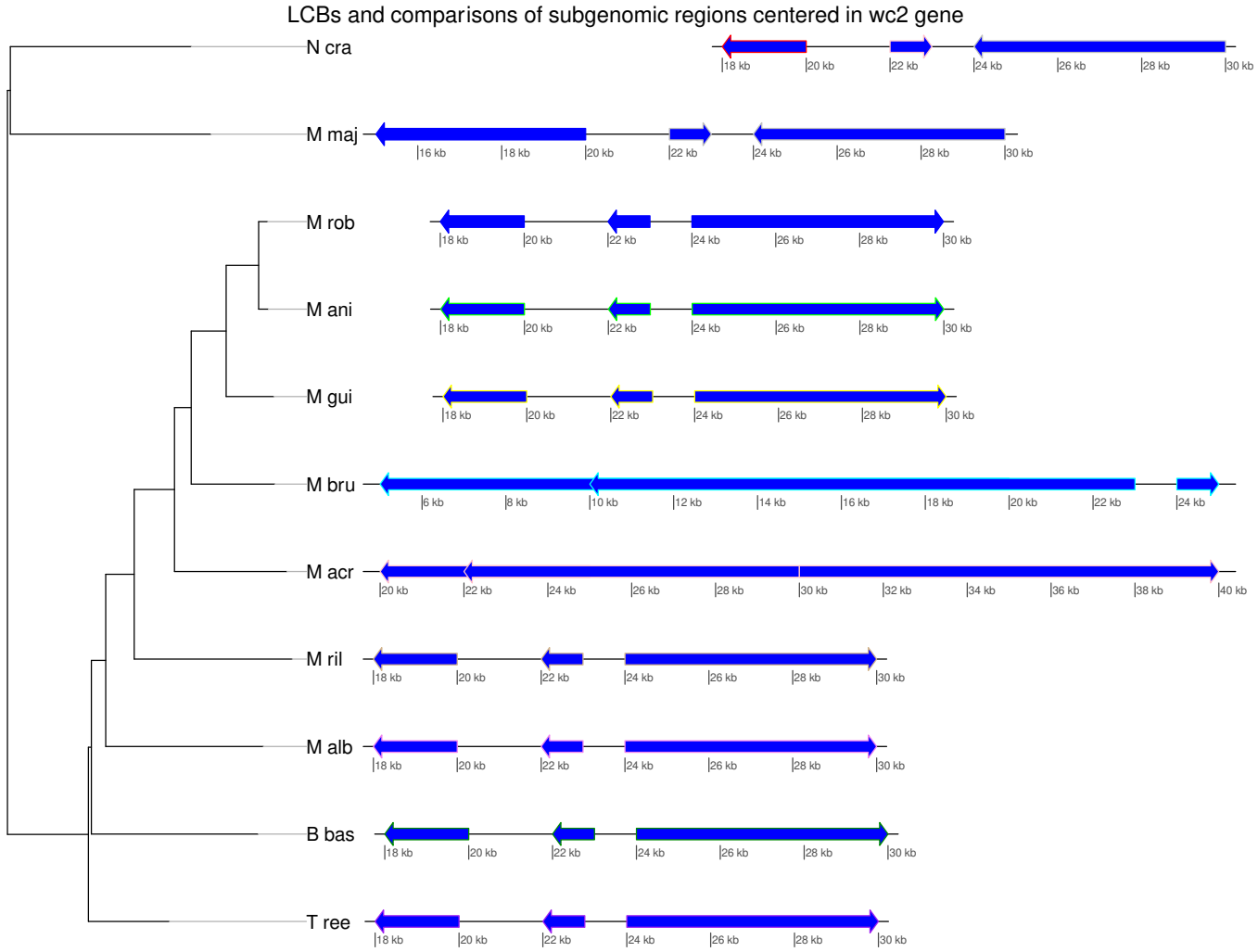


Figure 16. LCB identification and phylogenetic reconstruction in sub-genomic regions centered in *wc2.*

Black triangles represent the specific localization of the *wc2* gene in 10 kbp segments.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941



Fig*ure* 17. Identification on LCBs, phylogenetic reconstruction and comparison between the sub-genomic regions centered in the *wc2* gene.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

The sub-genomic region where we can find the *wc2* gene seems not to be highly conserved among the species in the filamentous fungi analyzed as indicated in Figure 16 where we can observe a mismatch of the LCBs identified in these regions. The discrepancies are visible when observing the exclusion of *M. majus* in the phylogenetic reconstruction since in this species the gene homologous to *wc2* is a transcription factor that doesn't have a PAS domain, this is exclusive in this species, since all homologous proteins analyzed in the remaining *Metarhizium* species contain the zinc finger domains and the PAS domain, like WC2 in *N. crassa.* Despite this, in the remaining *Metarhizium* species, comparisons in this specific region appear to inconstant, demonstrating a degraded synteny or not conserved in this specific region among the *Metarhizium* species analyzed.

vvd **Gene**

A 50 kbp sub-genomic region was selected for genomes of the specific species and strains *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297*, M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297*, M. album* ARSEF 1941. A progressive alignment was performed in the Mauve software. The files .bbone and .guide_tree, were selected for the analysis with the GenoPlotR in RStudio Software.

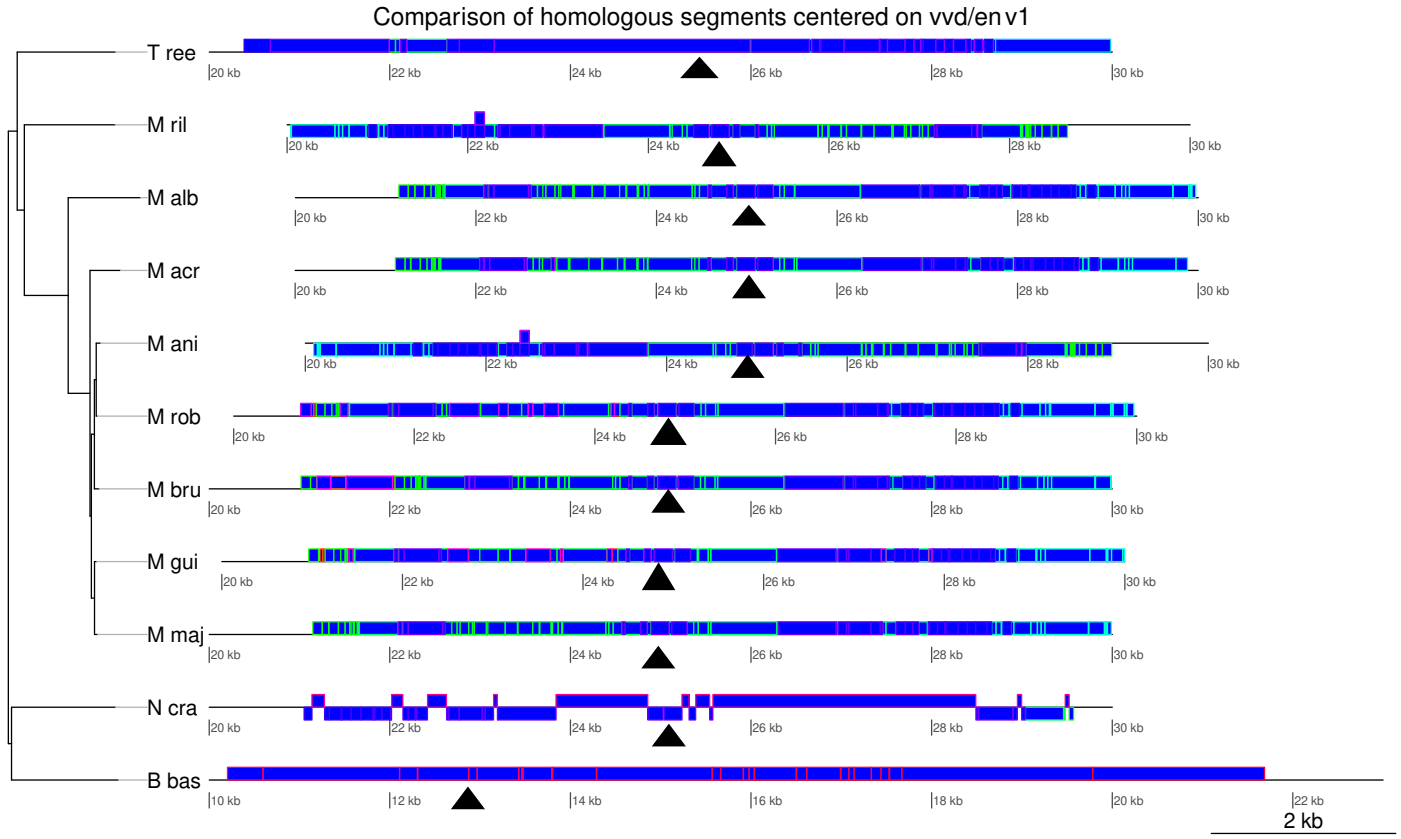


Figure 18. Identification of linear coaligned blocks in the sub-genomic region where *vvd/env1* are located.

Segments under the blocks correspond to the comparison between the blocks.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

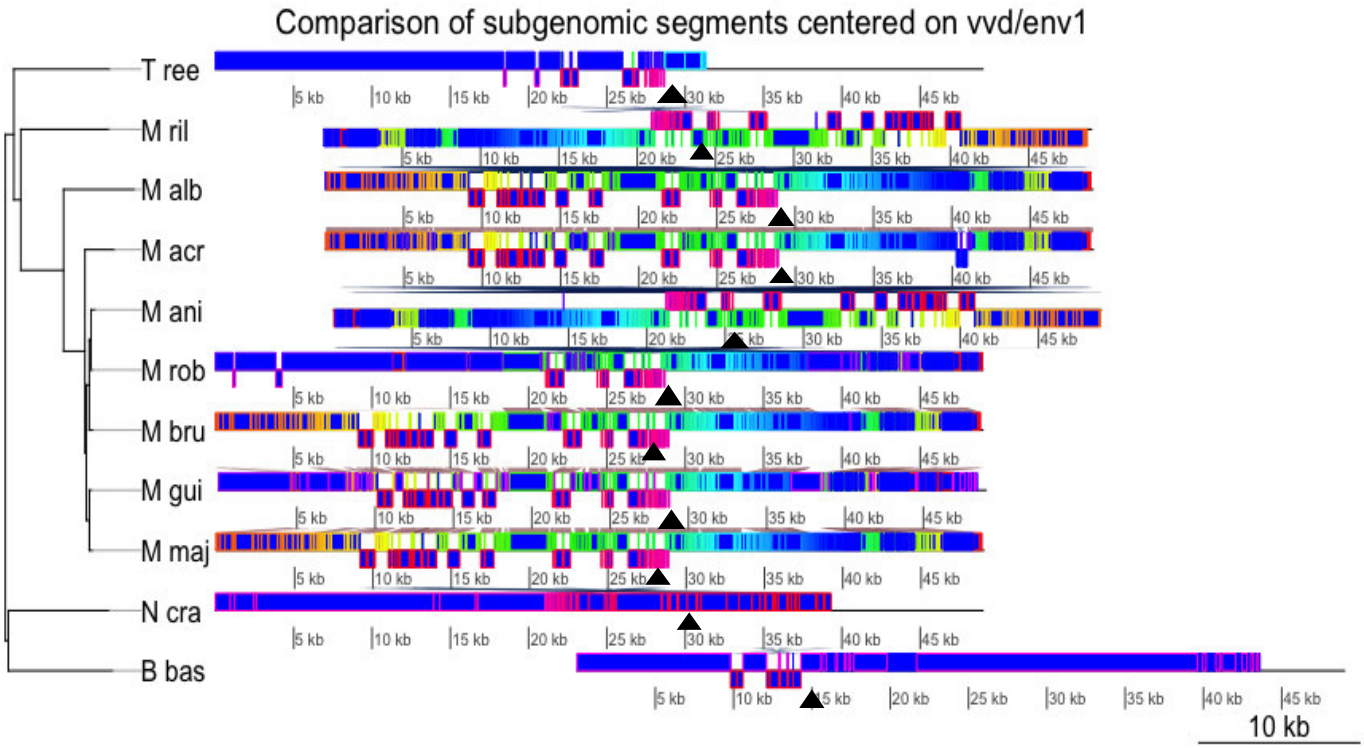


Figure 19. Comparison of sub-genomic segments centered on *vvd/env1*

Phylogeny generated by the local and global alignments generated by Mauve. LCBs identification and comparison between the species. Black triangles indicate the specific localization of *vvd/env1* genes.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

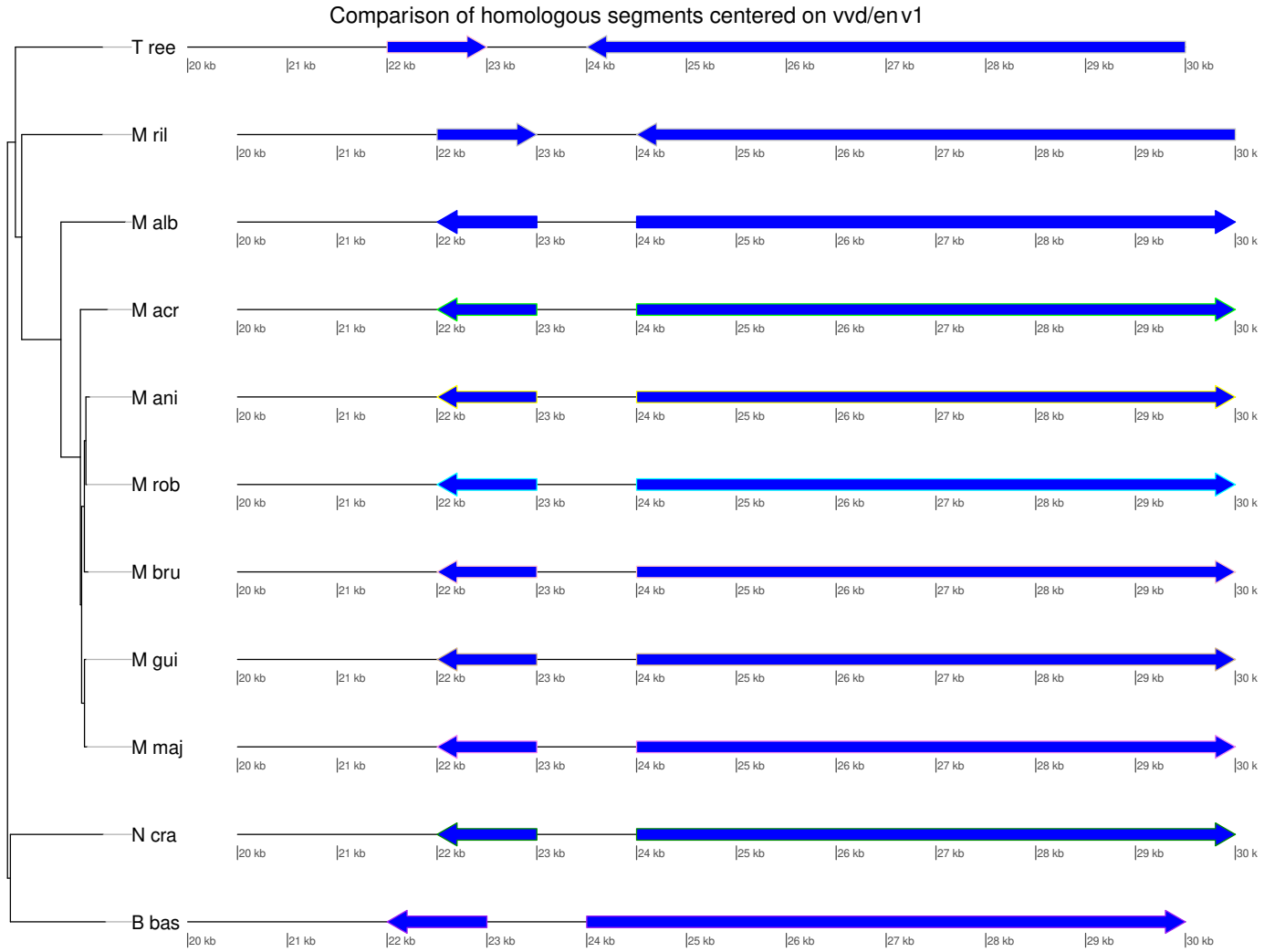


Figure 20. Comparison on 10kbp segments where *vvd/env1* is located.

Comparison on 10kbp segments selected from the analyzed subgenomic regions (50kbp) where the *vvd/env1* gene is located (23-27kbp approximately). Arrows correspond to the LCBs identified and the segments under the arrows are the comparison between them.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

The synteny index appears to be high. Similarities between the sub-genomic localization of the genes *vvd/env1* are evidence by the comparison analysis, indicated by the red and blue coloration under the LCBs in Figures 18 and 20. However, the PAS domain containing proteins in *Metarhizium* differs from *vvd i*n some species (*M. rileyi, M. album, M. acridum*) and are structurally and functionally similar to the ENVOY protein, which is involved in cellulase signaling. Notably, the *Metarhizium* species that show the greatest similarity to ENVOY are organisms that do not act as endophytic mycorrhizae.

In *B. bassiana*, the existence of the VIVID protein has already been determined and as expected, it is also involved with the conidiation of this fungus (Tong et al., 2018). therefore, its sub-genomic location is expected to be similar to that of *vdd* gene at *N. crassa* genome.

frq **Gene**

In this analysis, the reference genome used was *N. crassa* OR74A. The colors are a reference to the segments. To generate a complete visualization, the guide tree file given by Mauve illustrates the phylogenetic relationship generated by the analysis in the 50 kbp sub-genomic regions, paired with the blocks obtained by the previous analysis.

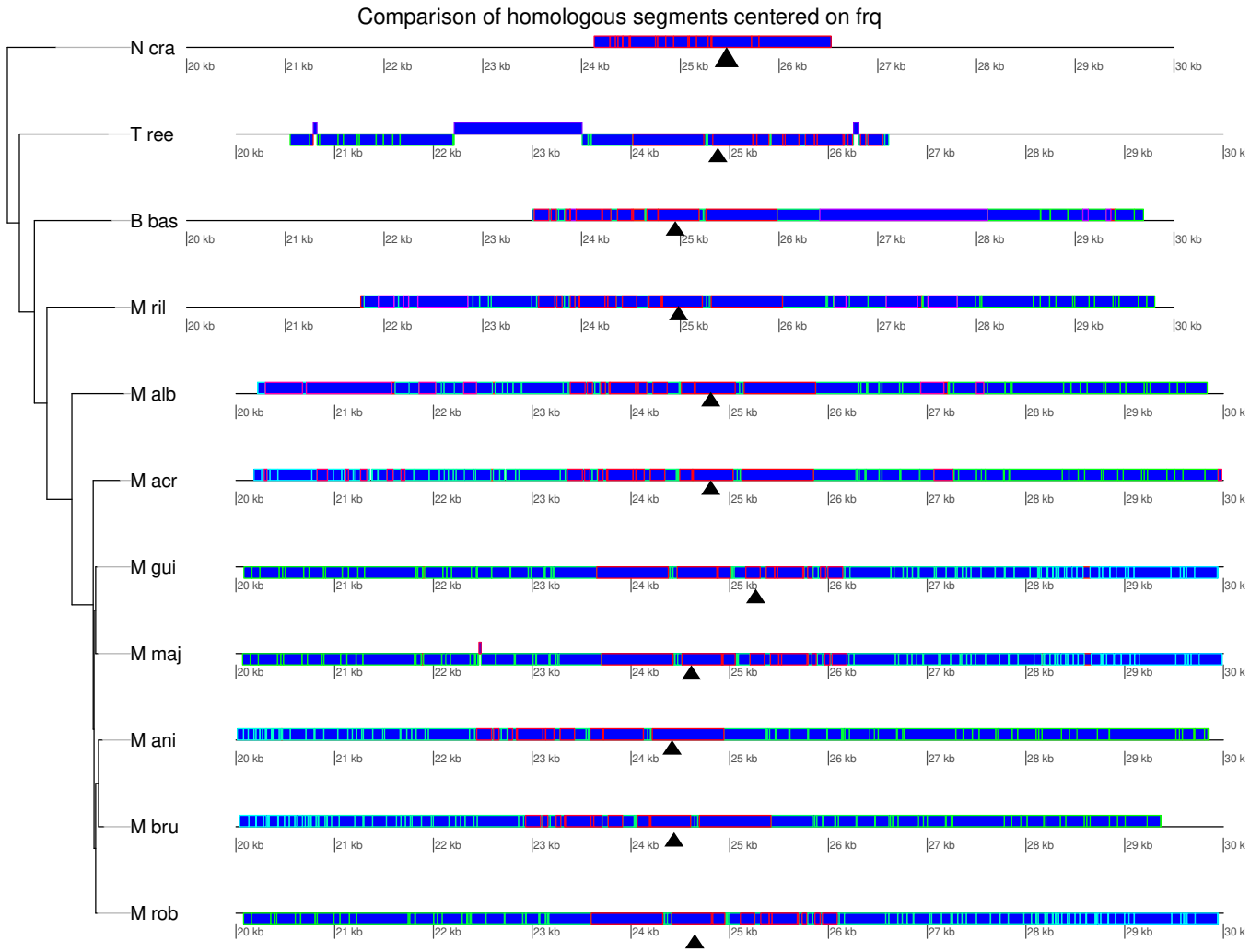


Figure 21. Localization and comparison of the LCBs where *frq* is located.

Localization and comparison of the LCBs in the sub-genomic region determined for 11 filamentous fungi. Sub- genomic localization of the gene *frq* (black triangles):
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

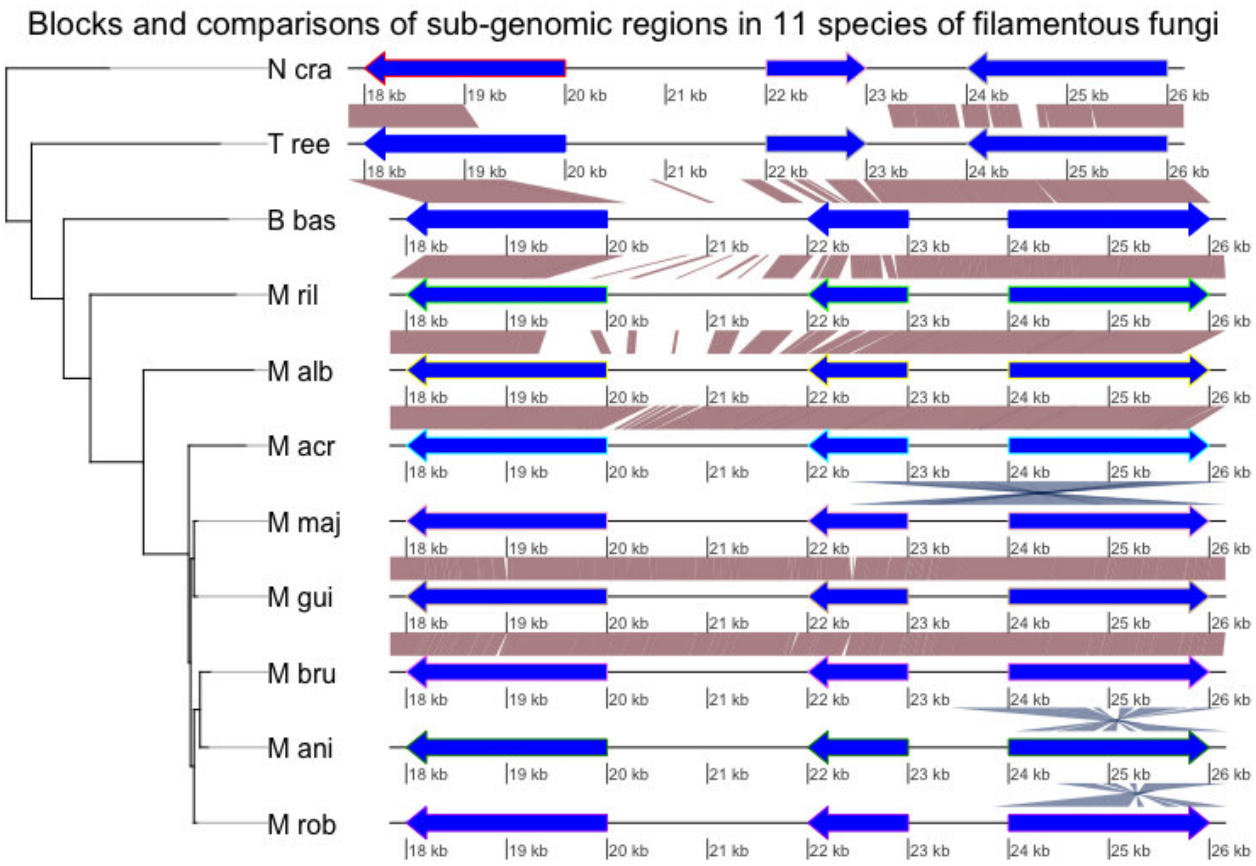


Figure 22. Localization of frq gene and its corresponding LCB.

Localization of *frq* gene and its corresponding LCB and co-adjacent blocks to the gene. The arrows do not represent the orientation of the blocks or genes.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

Using the approach described above, this plot was generated using the guide tree and the blocks identified in Figure 21. In this case the red and grey bars below the blocks represent the highly similar regions between the species compared.

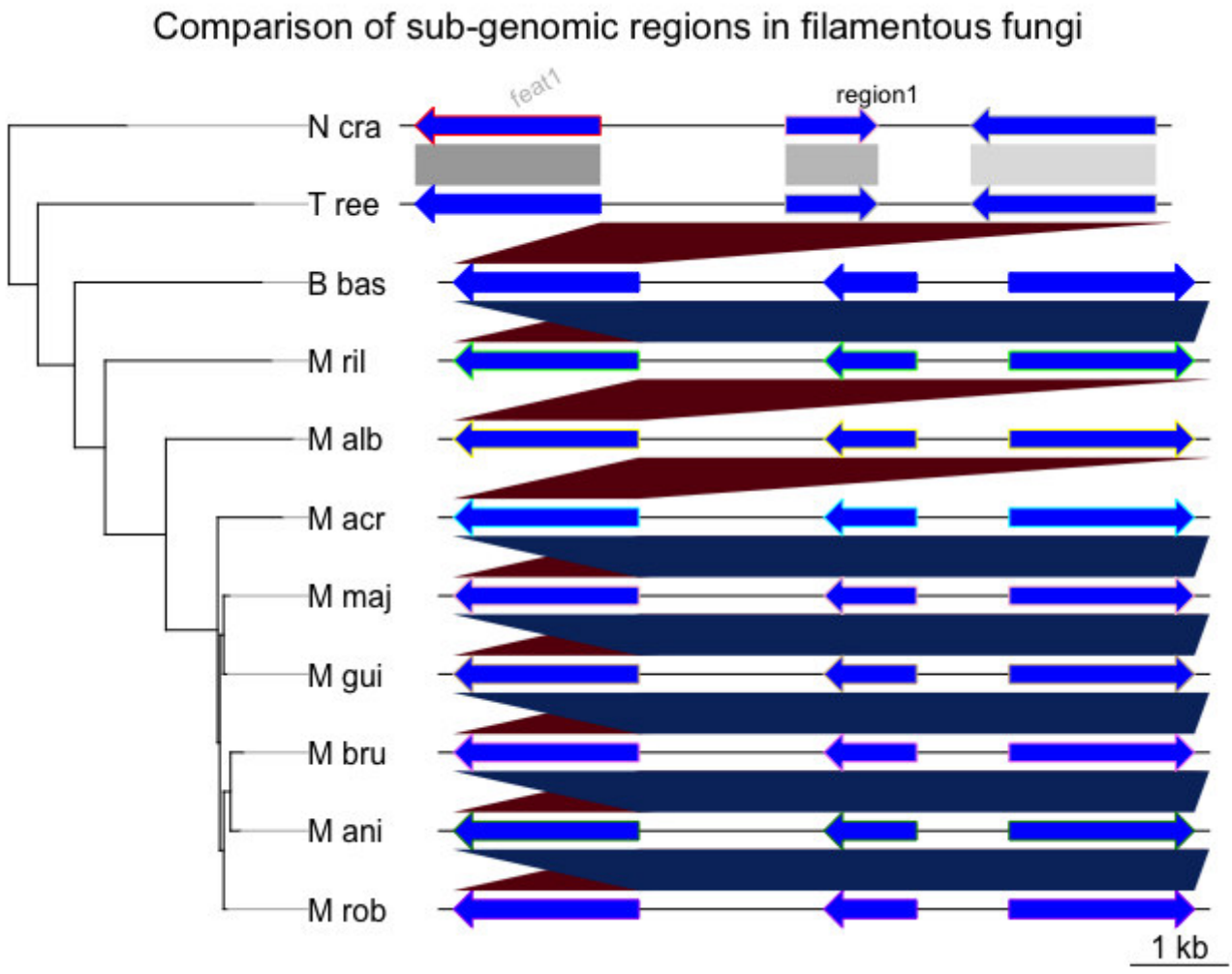


Figure 23. Localization of *frq* gene.

Localization of *frq* gene (right arrow) and comparison between co-adjacent blocks. Intense coloration is representative of the synteny index.
N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941 *M. anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

In Figure 23, the arrows do not represent the orientation of the LCBs; it only differentiates the blocks displayed in the regions delimited by the xlims function, where it was sought to center the display to *frq* and the blocks adjacent to it. The red and blue lines below the arrows represent the comparison analysis between the genomes to which they are connected, increasing the coloration when the similarity between the sub-genomic regions is more significant.

The similarities represented by the previous analysis comparing the nucleotide sequence of the gene *frq* in the same species agree with the results given by the synteny analysis. Phylogenetic trees are similar in both analyses, locating the similarities between *M. robertsii* ARSEF 23 and the other species and strains that are entomopathogenic generalists, *M. brunneum* ARSEF 3297 and *M. anisopliae* ARSEF 549; distancing with entomopathogenic specialist, such as *M. majus* ARSEF 297 and *M. acridum* CQMa102.
This region has a characteristic mycrosinteny pattern, since the similarities between the genes adjacent to *frq* in all the *Metarhizium* species and *T. reesei* are evidence that this homologous region is conserved within these analyzed species.

frh Gene

A 50 kbp sub-genomic region was selected from the genomes of the specific species and strains *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941. A progressive alignment was performed in the Mauve software. The files .bbone and .guide_tree were selected for the analysis with the GenoPlotR in RStudio Software.

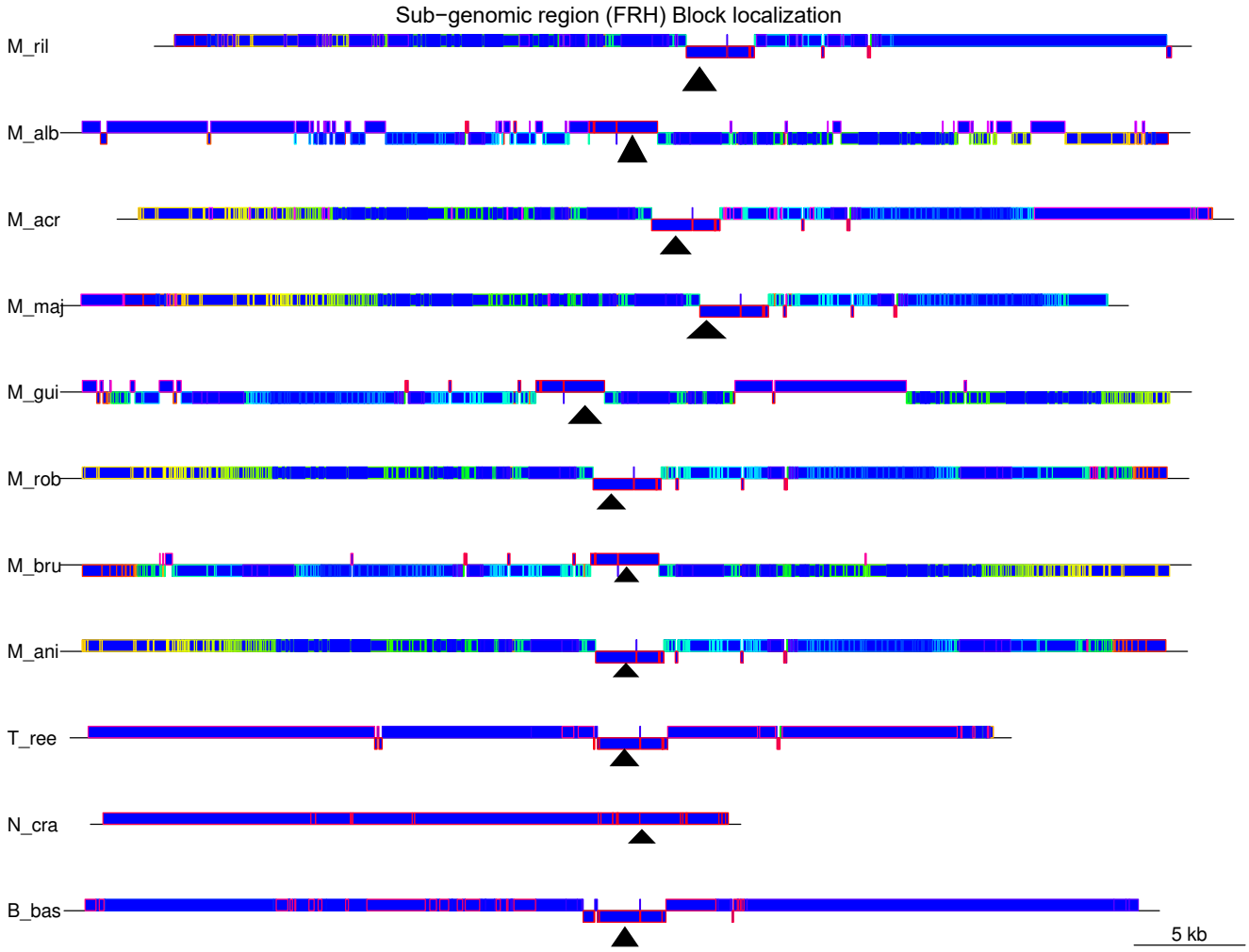


Figure 24. *frh* gene localization.

LCBs localized by the Mauve software. A subgenomic sample (50 kbp) was used. *frh* gene is located between 20-26 kbp region (black triangles indicate de *frh* gene).*N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

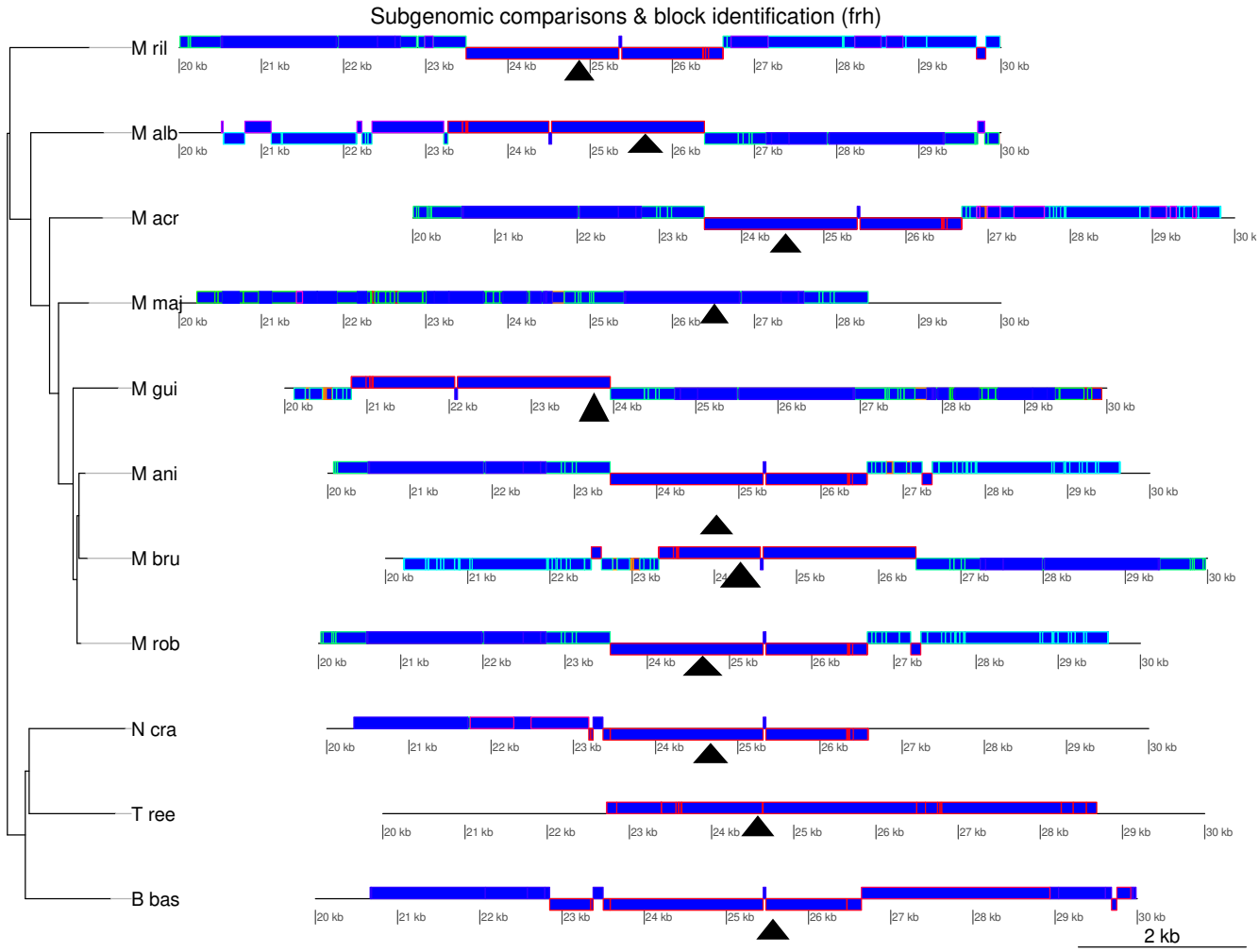


Figure 25. Comparison on sub-genomic regions where *frh* is located.

Delimitation on the sub-genomic region (10kbp) where *frh* is centered (black triangles). Comparisons between the sub-genomic regions demonstrate that the similitude is high, at least in the LCB were *frh* is located. *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

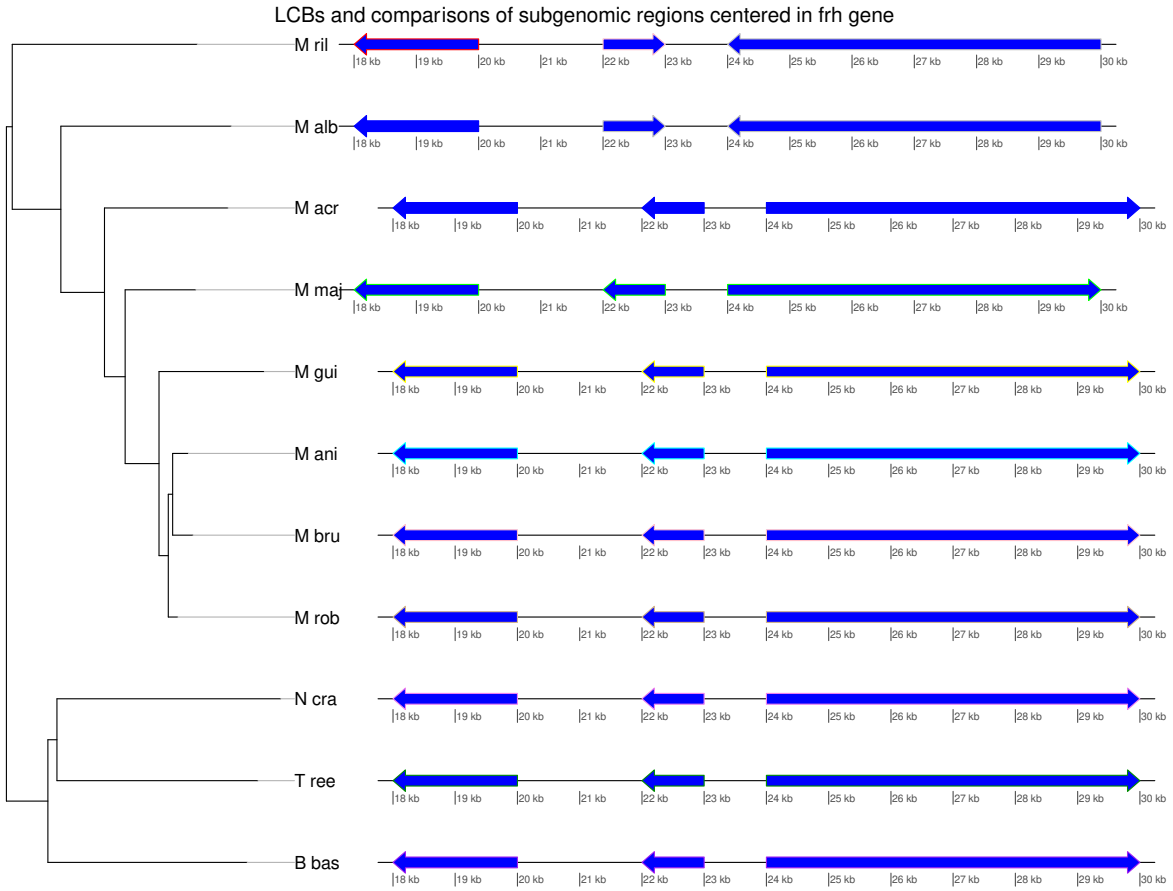


Figure 26. LCBs localization centered in the *frh* gene.

LCBs identification (arrows) in a 10kbp delimitation. *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

The phylogenetic tree calculated by Mauve indicates that the RNA helicase in *N. crassa* is similar to those found in the *Metarhizium* genome, RNA helicases are ATP dependent; meanwhile in *T. reesei* is a nuclear exosomal RNA helicase. Apart from *N. crassa*, the formation of the functional FRQ-FRH complex has not been described in any analyzed RNA helicases in *Metarhizium*.

Similar to the previous results, this analysis also groups the *Metarhizium* species that act as generalist (*M. brunneum, M. anisopliae* and *M. robertsii*) and specialist entomopathogens, highlighting the closeness that could exist between *M. robertsii* and *N. crassa* in at least this specific analysis, indicated by the red and blue coloration under the LCBs and the phylogenetic reconstruction based on the alignments performed by Mauve in figures 25 and 26.

6.5 Promoter characterization

Light response genes

wc1 Gene

Summary on the putative consensus sequences identified in the promoter of the *wc1/blr1* gene in 11 species. The complete upstream region of the putative ORF of each gene was analyzed, searching for the consensus sequences of the CREs.

Table 11. Putative CREs identified in the *wc1*/*blr1* genes.

Gene	Species & strain	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
White collar 1	*N. crassa* OR74A	-691 -1266 -1605 -1634 -2166 -2326 -2423 -3677 -3690	-393 -3487	-1850	*	-2524	*	*
White collar 1	*M. brunneum* ARSEF 3297	-137 -1095 -1276	-1280	*	-248 -253	-1036	*	*
White collar 1	*M. robertsii* ARSEF23	-143 -1089 -1270 -1750	-412 -1274	*	*	-1030	*	*
White collar 1	*M. acridum* CQMa 102	-646 -1057 -1716	-1835 -1868	-401 -1124	-1253	-998	*	*
White collar 1	*M anisopliae* ARSEF 549	-143 -1092 -1273 -1743	-896 -1277 -1544	*	-253	-1033	*	*
White collar 1	*M. rileyi* RCEF 4871	-144 -1440	-303 -394	*	*	*	*	-267

White collar 1	*M. guizohuense* ARSEF 977	-143 -1073 -1255 -1730	-595 -1259 -1882	*	-253	-1014	*	*
White collar 1	*M. majus* ARSEF 297	-143 -1050 -1232	-1236 -1508	*	-252	-991	*	*
White collar 1	*M. album* ARSEF 1941	-111 -611 -896 -964 -1160 -1755	-975	*	*	*	*	-360
Blue light regulator	*T. reesei* QM6a	-81 -1778 -2034 -2093	*	*	*	*	*	*
White collar 1	*B. bassiana* ARSEF 2860	-20 -192 -575	-370 -587	-1198 -1210	-513 -600		*	*

*Consensus sequence not identified in the promoter

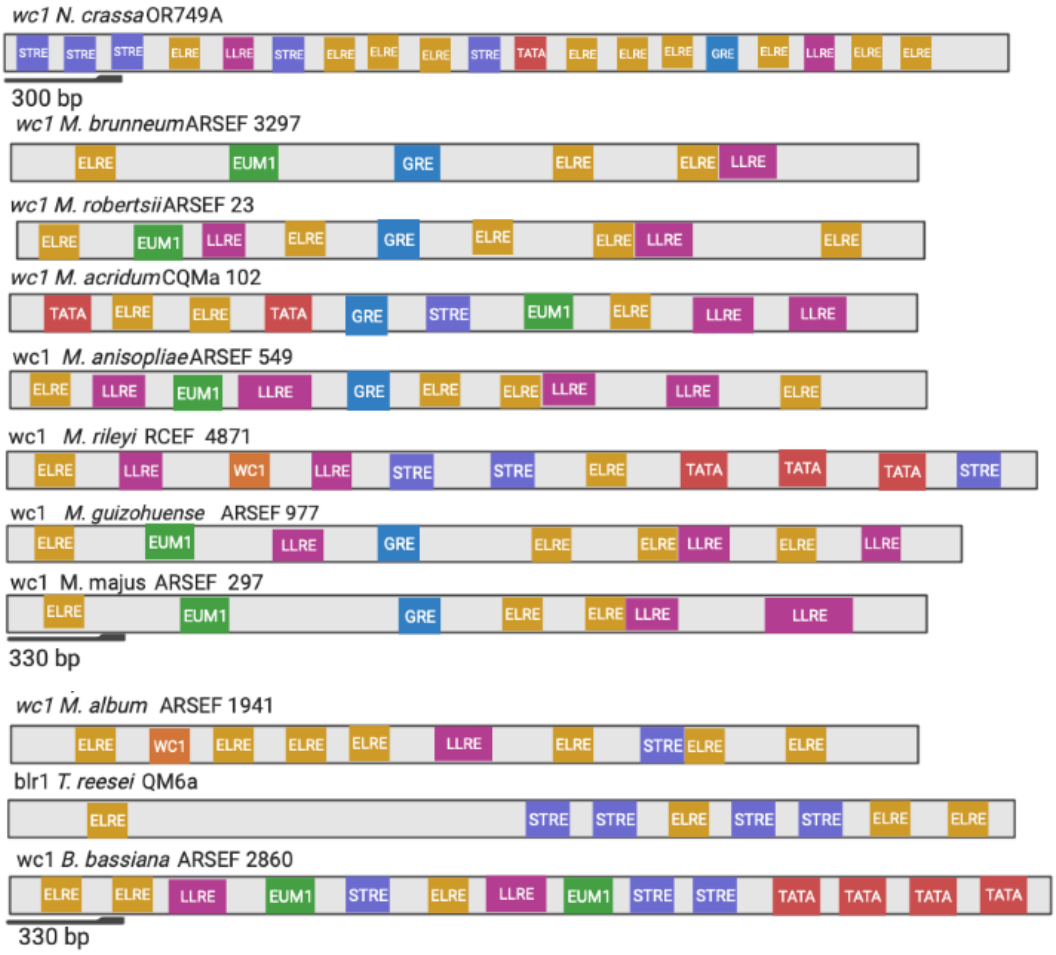


Figure 27. Representation of the cis-regulatory elements identified in the promoter of the wc1 / blr1 genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

By visually inspecting the promoters of the *wc1* gene and homologues, we can see that there are many CREs distributed on these promoters (Figure 27), highlighting the presence of LREs without an appreciable difference between them.

wc2 gene

Summary on the putative consensus sequences identified in the promoter of the *wc2/blr2* gene in 11 species. The complete upstream region of the putative ORF of each gene was analyzed, searching for the consensus sequences of the CREs.

Table 12. Putative CREs identified in the *wc2*/*blr2* genes

Gene	Species & strain	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
WC2	*N. crassa* OR74A	-38 -248 -409	-123 -1001	*	*	-425	-657 -676	*
Cutinase protein palindrome-binding protein	*M. brunneum* ARSEF 3297	-311 -1395 -1768 -2036 -3013	-1077 -1134 -1158 -2304 -2353 -2994	-2522	-428 -2215 -3266	-3169	-2086 -2376 -3183	-307
Cutinase protein palindrome-binding protein	*M. robertsii* ARSEF23	-311 -1398 -1783 -2765	-1137 -1161 -2319 -2368 -2510 -3010 -3032	-2537 -2549	-428 -3274	-3177	-2100 -2391 -3191	*
Cutinase protein palindrome-binding protein	*M. acridum* CQMa 102	-297 -377 -1389 -2091 -2735 -3070	-144 -1149 -1696 -2369 -2415 -2661 -3051 -3199	-1786 -2601	-117 -2280	-3212	-2137 -2438 -3226	
Cutinase protein palindrome-binding protein	*M anisopliae* ARSEF 549	-311 -1394 -1782 -2037 -3028	-326 -2316 -2365 -2505 -3346	-2532 -2544	-428 -2226 -3263	-3164	-2087 -2388 -3179	*

Cutinase protein palindrome-binding protein	*M. rileyi* RCEF 4871	-1145 -1757 -2845	-366 -1934 -2154 -2776 -2928 -3077 -3150 -3286	*	-461 -472	-3398	-2874 -3412 -3417	*
Cutinase protein palindrome-binding protein	*M. guizohuense* ARSEF 977	-307 -378 -2056 -2679	-1139 -1163 -1400 -1789 -2327 -2376 -2518 -2657 -3013 -3031 -3222	-2545	-425 -2238 -3270	-3173	-2106 -2399 -3187	*
GATA-type sexual development transcription factor NsdD	*M. majus* ARSEF 297	-231 -280 -610 -718 -955 -1149 -1224 -1254 -1450 -1638 -2214 -2232 -2537 -2693 -3519 -3562 -3974 -4280	-464 -1156 -1985 -2671 -2831 -2935 -2994 -3061 -3189 -3406 -3667 -4241 -4266	-534 -759 -3144	-3279	*	-3732	-2043
Cutinase protein palindrome-binding protein	*M. album* ARSEF 1941	-388 -934 -1317 -1869 -2844 -2792 -3198 -3390	-636 -1018 -1842 -1952 -2484 -2518 -3439	-2862	-1400 -3344	-3242	-1121 -1520 -2367 -2560	*

Blue light regulator 2	*T. reesei* QM6a	-1056 -1586 -2597 -3064 -3495 -3641 -3923	-90 -905 -1369 -1571 -1617 -2038 -2792 -3102 -3241 -4082	*	-547 -881 -2739	-2823	*	-399 -450
Cutinase protein palindrome-binding protein	*B. bassiana* ARSEF 2860	-134 -234 -916 -962 -1208 -2143 -3035 -3046	-636 -1954 -2116 -2422 -2801 -2902	*	-1740	-2923	-2700	*

*Consensus sequence not identified in the promoter

Figure 28. Representation of the cis-regulatory elements identified in the promoter of the *wc2/blr2* genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

LREs seem widely distributed in the promoter of *wc2/blr2* genes (Figure 28), in this way, we can infer the importance of the transcription of these genes after immediate and late light stimulus. In the species of *Metarhizium*, except for *M. majus*, the location of these CREs is highly similar.

vvd Gene

To identify and characterized the promoter region on the VIVID/ENVOY, PAS domain containing proteins, a manual search was carried out in the complete upstream region of the putative ORF of each gene identifying for the putative *cis*-regulatory elements.

Table 13. Localization of the consensus sequences for the putative CREs in promoter of the *vvd*/*env1* genes.

Gene	**Species & strain**	**Early light response element ELRE** GATC	**Late light response element LLRE** TGA---TCA	**TATA Box**	**ENVOY Upstream motif 1 EUM1** CTGTGC--CTGTGC	**Glucose Response Element (GRE)** CACGTG	**Stress response elements (SRE)** AGGG	**WC1 binding site** (A/T)GATA(A/G)
VIVID (vvd)	*Neurospora crassa* OR749A	-29 -250 -281 -291 -961 -1830 -2203 -2289 -2773 -2787 -2856 -3113	-1347 -2101 -2114 -2187 -2449 -2542 -2798 -2881 -2829 -3211	*	-1745		-865 -1272 -1494 -2531	*

Element involved in conidiation	*M. brunneum* ARSEF 3297	-218 -361 -577 -702 -790 -902 -1066 -1214 -1854	-255 -285 -1052 -1261 -1420 -1513	-122			*	*
PAS domain containing protein	*M. robertsii* ARSEF23	-219 -362 -578 -607 -779 -891 -1030 -1073 -1221	-286 -1268 -1427 -1531	-123		-	*	*
Cellulose signaling associated protein ENVOY	*M. acridum* CQMa 102	-221 -330 -360 -583 -612 -706 -790 -907 -1054 -1097	-1471 -1575 -1914	-123	-	-	*	-690
Element involved in conidiation	*M anisopliae* ARSEF 549	-507 -650 -866 -895 -950 -992 -1067 -1179 -1318 -1361 -1509	-132 -574 -1556 -1715 -1818	-411	-		*	*
Cellulose signaling associated protein ENVOY	*M. rileyi* RCEF 4871	-89 -247 -651 -878 -907	-1489 -2162	-290	*	*	*	*

		-1139 -1106 -1268 -1458 -1464 -1505 -1570 -1677						
Element involved in conidiation	*M. guizhouense* ARSEF 977	-219 -361 -577 -606 -703 -786 -898 -1037 -1081 -1229	-286 -1276 -1429 -1533 -1572	-123	-	-	*	*
Element involved in conidation	*M. majus* ARSEF 297	-362 -578 -607 -704 -787 -899 -1038 -1082	-256 -286 -1196	-123	-	-	*	*
Cellulose signaling associated protein ENVOY	M. album ARSEF 1941	-108 -216 -382 -600 -721 -751 -961 -990 -1183 -1228 -1257 -1300 -1361 -1512 -1558 -2057	-27 -333 -1337 -1544 -1899 -2300	*	-431 -1965	*	-2142	*

Glycoside hydrolase family 15, cellulose signaling associated protein ENVOY	*T. reesei* QM6a	-153 -404 -616 -623 -719 -805 -837 -1154 -1239 -1311 -1552 -2132 -2155 -2214 -2362 -2391 -2576	-135 -860 -1873 -2148 -2289 -2372 -2497 -2643	-2274	*	*	-1633	*
Vivid PAS protein VVD	*B. bassiana* ARSEF 2860	-227 -419 -478 -601 -1261	-271 -708 -987	*	-	-	-1190	*

*Consensus sequence not identified in the promoter.

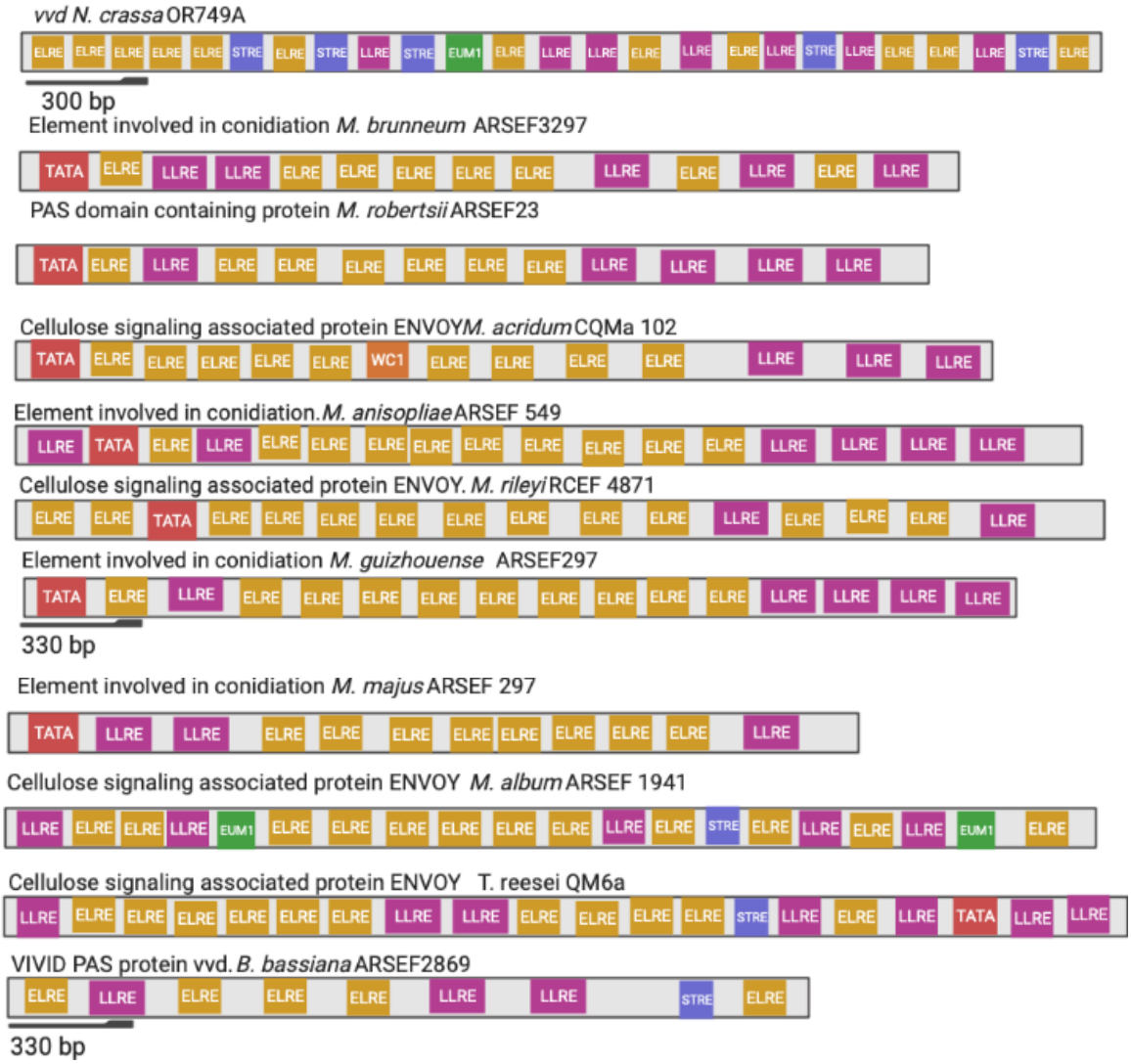


Figure 29. Representation of the cis-regulatory elements identified in the promoter of the vvd/env1 genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

The transcription of the proteins VIVID and ENVOY are mediated by the WCC, therefore it is expected that both LREs are present in the promoter region of these genes, since its transcriptions is dependent on the binding of WCC to this consensus sequences. Localization of the LREs in *Metarhiuzm* are highly similar, with exception of *M. album*.

Theoretically, the WCC complex could be expected to regulate other light response elements in *Metarhizium*, in addition to FRQ; like what happens *in N. crassa*; generating a biological pathway that could be highly similar to that already described in saprophytic models.

frq gene

The analysis was carried out using a 2500 bp sequence upstream from the putative ORF of the *frq* gene.

Table 14. Localization of the consensus sequences identified for each CREs in the promoter of *frq* genes

Gene	Species & strain	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
Frequency clock protein	*M. brunneum* ARSEF 3297	-82 -184 -385 -395 -399 -997 -1224 -1441 -1451	-91 -183 -385 -543 -735 -1143 -1473	-1548	-2270	*	*	-573
Frequency clock protein-like protein	*M. robertsii* CARO4	-82 -396 -400 -1001 -1167 -1228 -1445 -1455 -2060	-180 -386 -735 -964 -1144	-1550	-2366	*	*	*
Frequency clock protein - like protein	*M. robertsii* ARSEF23	-82 -396 -400 -1001 -1167 -1228 -1445 -1455 -2060	-180 -386 -735 -964 -1144	-1550	-2366	*	*	*

Frequency clock protein	*M. acridum* CQMa 102	-397 -401 -652 -1215 -1448 -1458 -1922 -2457	-177 -387 -736 -1045 -1129 -1209 -1675	-1553	-2126	*	*	*
Frequency clock protein	M anisopliae ARSEF 549	-83 -182 -396 -400 -1000 -1227 -1444 -1454 -2054	-59 -87 -181 -386 -445 -543 -736 -963 -1143	-1549	-2352	*	*	*
Frequency clock protein	*M. rileyi* RCEF 4871	-160 -382 -386 -390 -684 -795 -1452 -1505 -1662 -1672	-635	-1233	*	*	*	-442
Frequency clock protein	*M. guizohuense* ARSEF 977	-82 -92 -396 -400 -994 -1221 -1437 -1447	-91 -180 -386 -544 -572 -961 -1137	-1542	-2248	*	*	-575
Frequency clock protein	*M. majus* ARSEF 297	-393 -397 -401 -992 -1219 -1436 -1446 -2008	-91 -173 -383 -569 -732 -959 -2183	-220	-2279	*	*	-572 -647

Frequency clock protein	*M. album* ARSEF 1941	-85 -171 -389 -398 -402 -406 -553 -1001 -1117 -1435 -1665 -1675 -1985	-183 -446 -678 -1067 -1287 -1913 -1984	*	*	-802	-67	*
Frequency clock protein	*T. reesei* QM6a	-27 -348 -355-359 -676 -695 -1132 -1562 -1980 -2407	-43 -405 -640 -719	*	*	*	*	-1315
Frequency clock protein	*B. bassiana* ARSEF 2860	-337 -343 -699 -1846 -2155 -2252 -2352	-778 -874 -2150	*	-679 -1772	*	-1425	*
Frequency clock protein	*N. crassa* OR74A	-94 -122 -131 -197 -918 -1119 -1590 -1808 -2008 -2140	-247 -1632 -2409	*	-78 -1067	*	-1390 -2310	-911

* Consensus sequence not identified in the promoter.

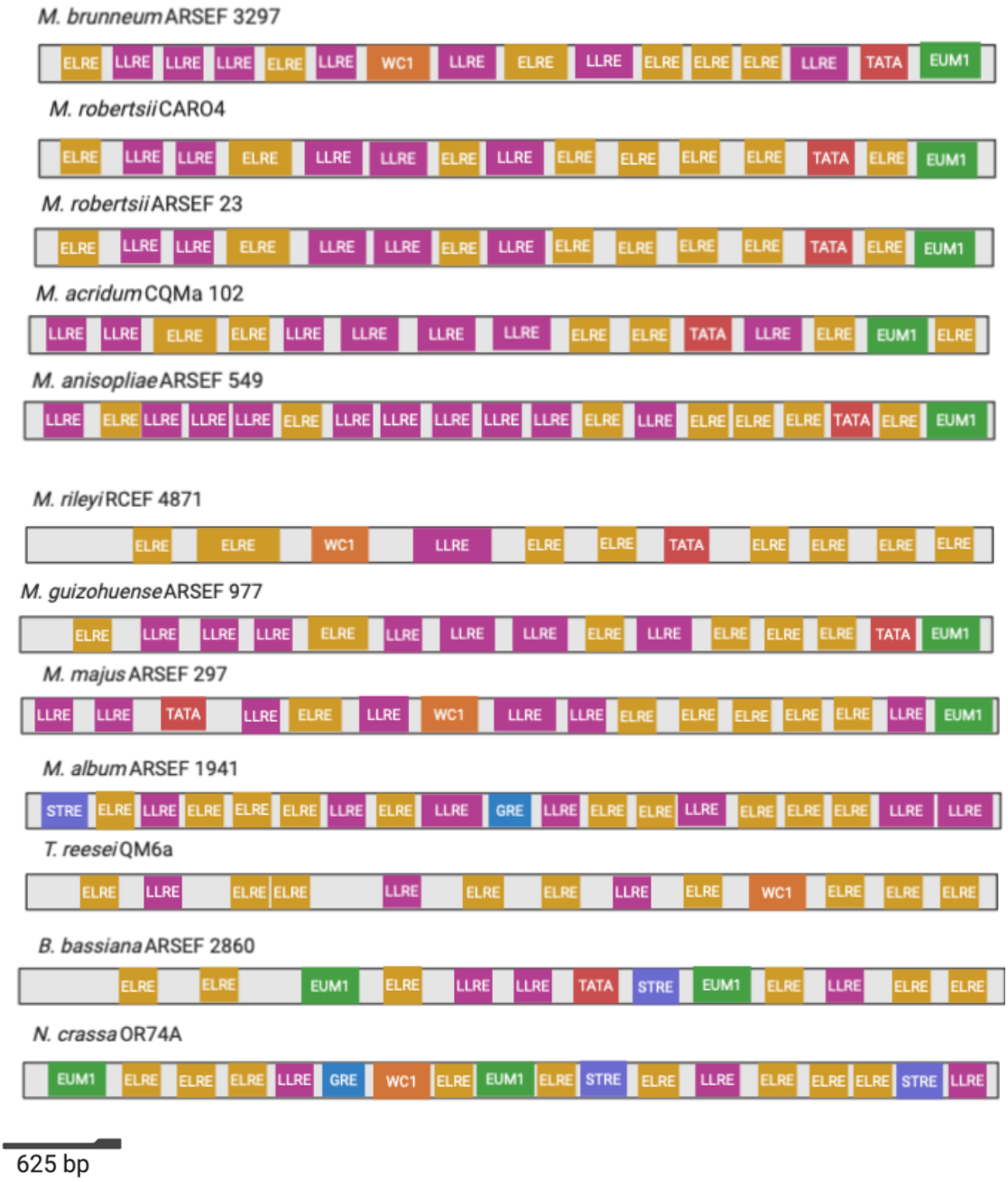


Figure 30. Representation of the cis-regulatory elements identified in the promoter of the frq genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

frh gene

For the *frh* gene a different length of the upstream region from the putative ORF was selected for each species

Table 15. Localization of the consensus sequences identified for each CREs in the promoter of the *frh* genes

Gene	Species & strain	Early light response element ELRE	Late light response element LLRE	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
ATP-dependent RNA helicase DOB1	M. brunneum ARSEF 3297	-419 -485	*	-112	*	*	-393 -526	*
ATP-dependent RNA helicase DOB1	M. robertsii ARSEF23	-102 -418	-415	-112	*	*	-393	*
ATP-dependent RNA helicase DOB1	M. acridum CQMa 102	-89 -213	-67 -402 -444	-99	*	*	-551	*
ATP-dependent RNA helicase DOB1	M. anisopliae ARSEF 549	-102	-415	-112	*	*	-393	*
ATP-dependent RNA helicase DOB1	M. rileyi RCEF 4871	*	-477	-17	*	*	*	*
ATP-dependent RNA helicase DOB1	M. guizhouense ARSEF 977	*	-415	-112	*	*	-393	*

-ATP dependent RNA helicase DOB1	M. majus ARSEF 297	*	-415	-112	*	*	*	*
ATP-dependent RNA helicase DOB1	M. album ARSEF 1941	*	-203	-110	*	*	*	*
Nuclear exosomal RNA helicase	T. reesei QM6a	-373	-170 -394	*	*	-496	*	*
DSHCT domain-containing protein	B. bassiana ARSEF 2860	-245	*	*	*	*	*	*
FRQ-interacting RNA helicase	N. crassa OR74A	-26	*	-123	*	*	*	*

* Consensus sequence not identified in the promoter.

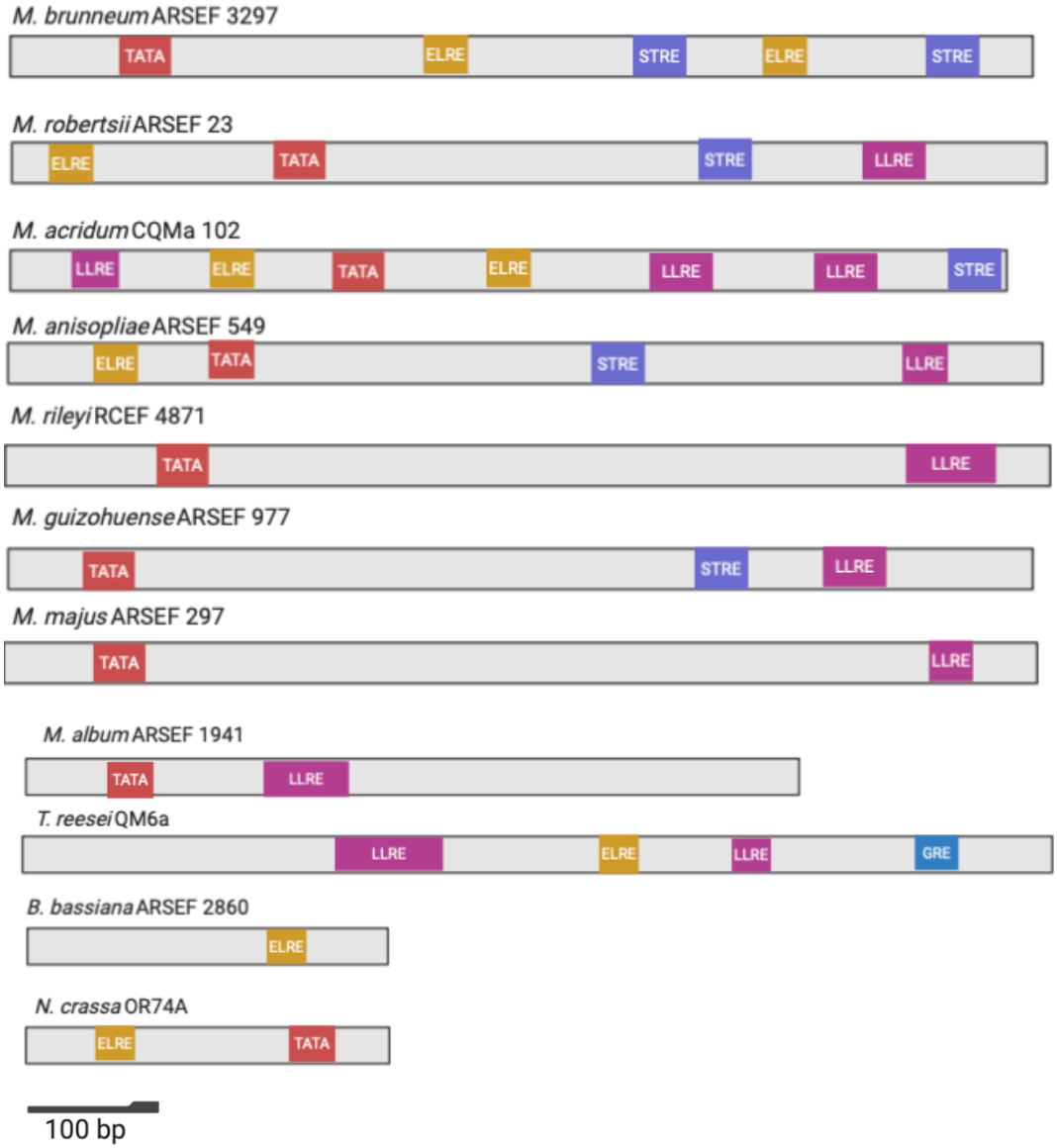


Figure 31. Representation of the cis-regulatory elements identified in the promoter of the *frh* genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

In the upstream region of the putative ORF of the *frq* and *frh* genes, consensus sequences of *cis*-regulatory elements were identified. Remarkably, the light response elements appear to be the prominent ones, including at least one of them in the promoter region in the *frq* genes (Figure 30), therefore, the transcription of said genes as a response to the light stimulus is feasible. The detection of early and late light response elements in the promoter regions of the genes, supports the proposal of expression mediated by light stimulation. Mainly in the promoters of the *frq* gene, where its role as a regulator of this biological process is primordial, acting in the light-mediated transcription activity. The several ELREs and LLREs could be evidence of the sensitivity required for both early and late expression of *frq* in filamentous fungi and the importance of the response to light.
Identifying the EUM1 consensus sequence in the promoter region of the *frq* gene in *Metarhizium*, open the possibility of the control expression of this gene by proteins homologous to the *T. reesie* ENVOY protein.

In the case of *frh*, the location of the late light response elements in the promoter region (Table 15) suggested that the transcription of this gene might not be induced after light stimulation. The promoters of the *frh* genes in *Metarhizium* contain stress response elements; the role of these in the transcriptional induction mediated by different forms of environmental stress has been described.

Although not all the functions of *frh* are known in *N. crassa*, different roles have been proposed in addition to its interaction with *frq.* The deletion of this gene is not viable in this species. In *B. bassiana* the deletion of *frh* is viable but the double deletion of both genes, *frq1* and *frq*2, is not. However, both *frh* promoters in *N. crassa* and *B. bassiana* are similar in size and contain early-light response elements (Figure 31); despite having contrasting consequences in their deletion. In *Metarhizium* the role of *frh* in circadian rhythms has not been described since the study of photobiology and chronobiology in this clade is relatively new.

6.5a LREs are conserved among light response genes.

To determine the motifs of the plausible binding sites of WCC in the promoter of light response genes (*wc1, wc2, vvd* and *frq*) and their homologous of each species, a search using the MEME suite was carried out.

WC1/BLR1

The conservation on LREs consensus sequences is maintained among the promoter of the *wc1/blr1* genes; therefore, this could indicate that the expression of this gene occurs continuously, ensuring a response to the light stimulus and the gene transcription mediated by the WCC.

Figure 32. Analysis results on the motif identification and alignments on the sequences of the *wc1* promoters.

Yellow stars indicate the ELREs in the promoter sequences, pink stars indicate the LLREs in the promoter sequences. Motifs shown have a p-value < 0.0001

WC2/BLR2

In contrast to the promoters of the *wc1/blr1* genes; the promoters of *wc2/blr2* demonstrate a marked preference for the LLRES in the *Metarhizium* species analyzed; this could indicate the possibility that the expression *wc2* doesn't occur after immediate light stimulus. Recalling that WC2 does not participate directly in the detection of light, since it does not have a LOV domain, its function is to stabilize and allow the WCC interaction with the promoters of the light response genes (Corrochano, 2007). In this way, it is possible that the presence of large amounts of the WC2 protein in the cytoplasm is not immediately required as in the case of WC1.

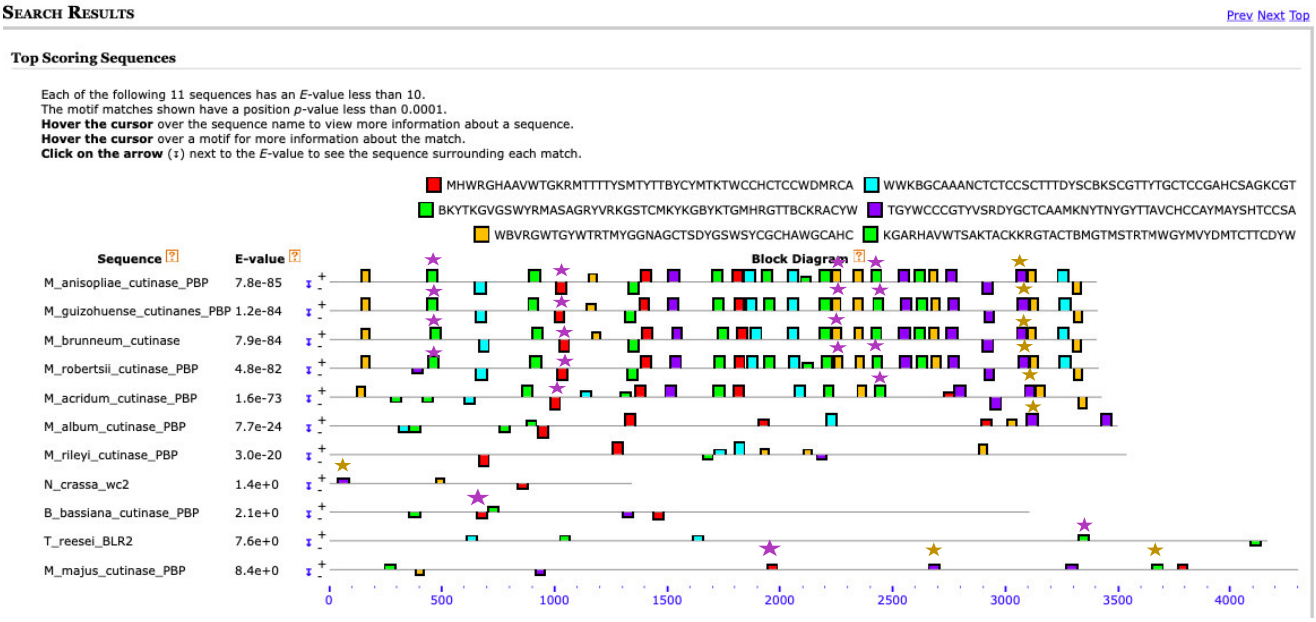


Figure 33. Analysis results on the motif identification and alignments on the sequences of the *wc2* promoters.

Yellow stars indicate the ELREs in the promoter sequences, pink stars indicate the LLREs in the promoter sequences. Motifs shown have a p-value < 0.0001

VVD/ENV1

In *N. crassa* the function of VIVID is regulating the WCC interacting with the WC1 protein avoiding the photoaduct formation with the chromophore FAD; in this way, the transcription induced by light is regulated (Fuller *et al*., 2015). Therefore, VIVID must be found in the cytoplasm to act quickly on the WCC.

In *T. reesei*, ENVOY acts in the metabolic regulation and in the expression of several genes through different signaling routes (Casas & Herrera, 2013). Again, the participation of this protein in the metabolism mediated by light is of great importance; therefore, the constant presence of this protein in the cytoplasm is necessary.

This can be observed in the distribution of the elements of light response (early and late) that are widely distributed in the sequences of the promoters of these genes, without a marked difference. In the *Metarhizium* species analyzed, there are differences in the characterization of homologous proteins to VIVID that contain this characteristic PAS domain; despite these differences, the presence of LREs is constant; that is, we can find both ELREs and LLREs (Figure 35).

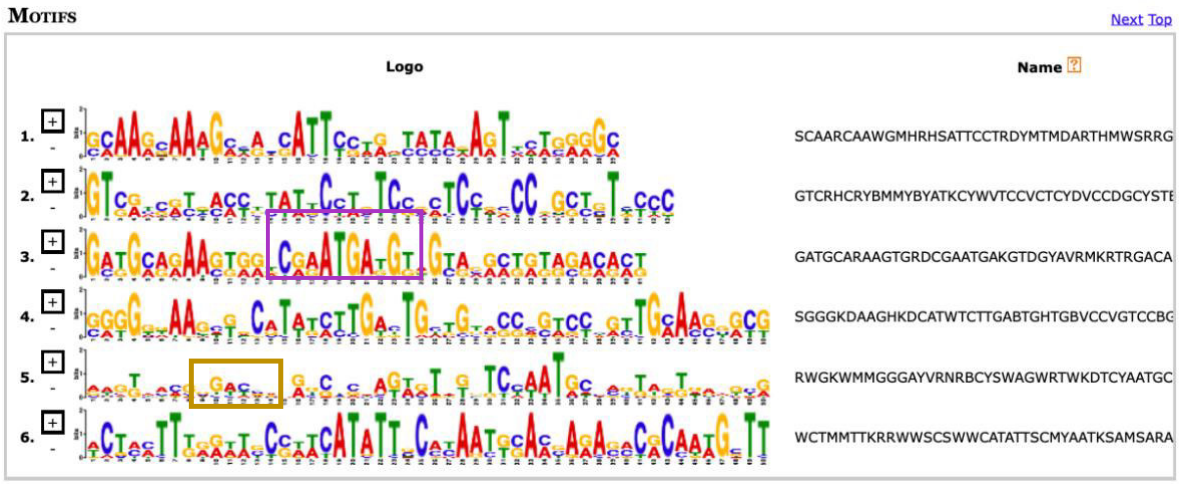


Figure 34. Identification of LREs.

Identification of consensus sequences of ELRES (yellow rectangle) and LLREs (Pink rectangle) in the 6 motifs identified by MEME Suite in the promoter sequences of *vvd/env1* genes in filamentous fungi.

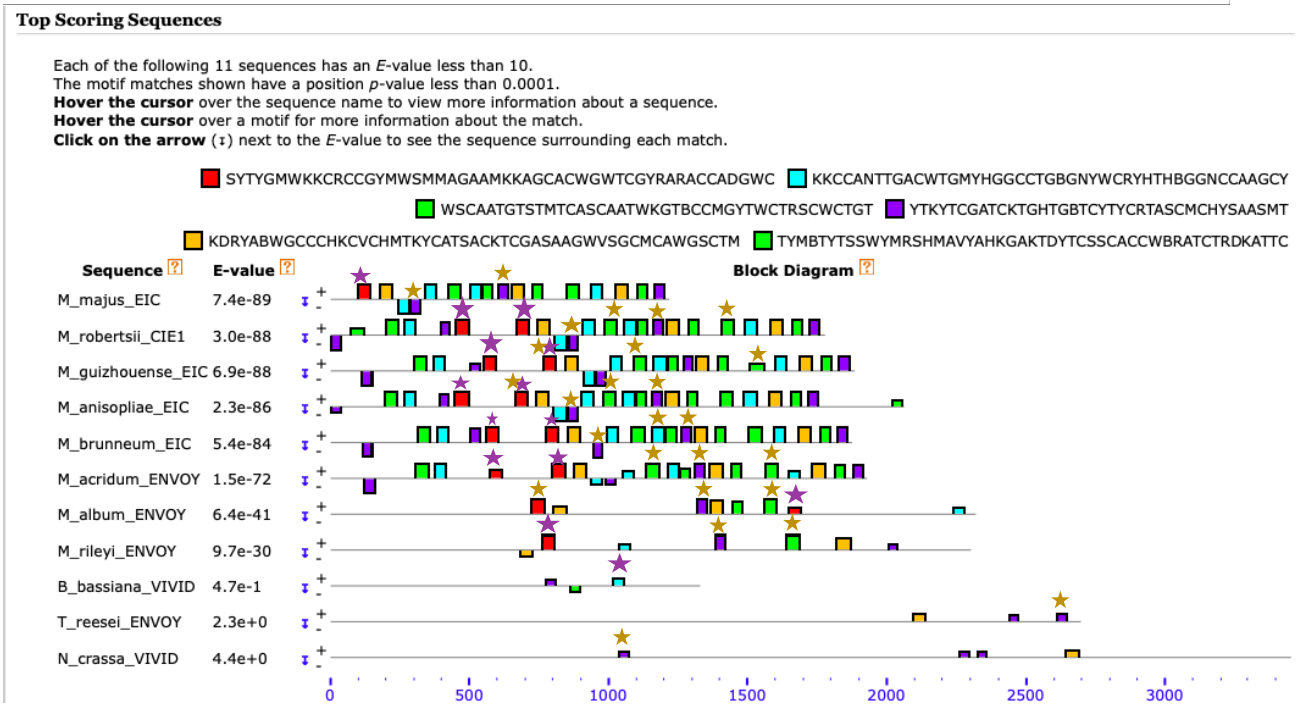


Figure 35.Specific localization of LREs in *vvd/env1* promoters.

Specific localization of ELREs (yellow stars) and LLREs (pink stars) in the *vvd/env1* promoter region of each species. Motifs shown have a p-value < 0.000

FRQ

The induction in the transcription of the *frq* gene is mediated by the WCC, therefore its binding to the LREs is necessary (Cheng *et al*., 2009). Two important LREs have been described in the promoter of this gene, one proximal (pLRE) and one distal (dLRE), to the start of the *frq*'s ORF. When pLRE is deleted, a 70% decrease in *frq* levels is generated and when dLRE is deleted a 50% decrease in these transcripts was obtained (Froehlich *et al*., 2002), therefore the presence of these LREs is of utmost importance.

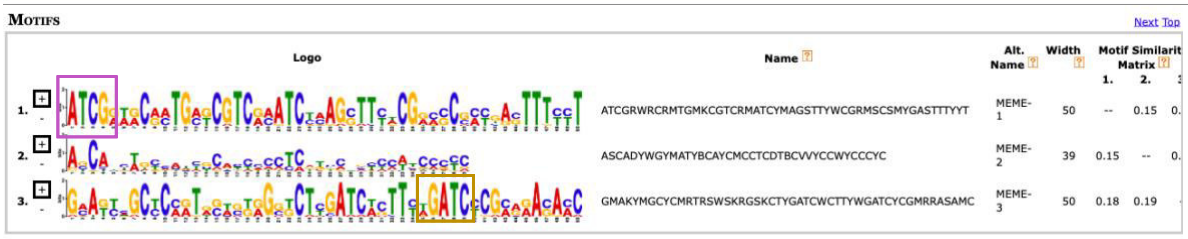


Figure 36. Motif identification in the promoter sequences of the analyzed species.

6 Motifs were determined to be conserved in the input sequences of the promoter region. Motifs 1 and 3 contain the consensus sequence for LREs (pink: LLRE, yellow: ELRE).

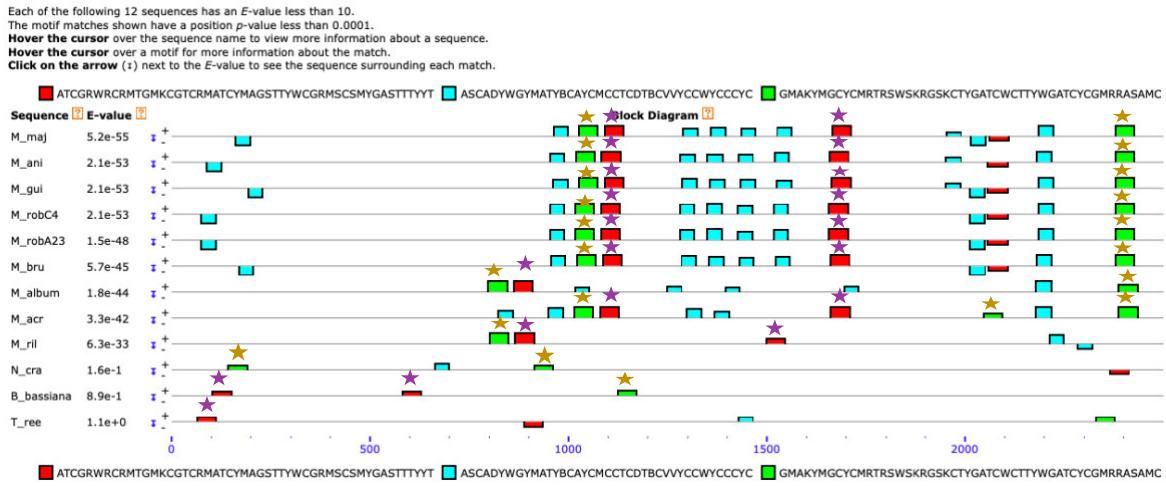


Figure 37. Motif identification on *frq* promoters.

Identification of each motif in the sequences. Red boxes contain the consensus sequences of LLRE and green boxes the consensus sequence of ELRE.
Motifs shown have a p-value < 0.0001.

Smith *et al*., (2010) determined the consensus sequences for the putative binding sites of the WCC in different light-responsive genes, including TFs (e.g., *sub1*) in *N. crassa*, determined that 20% of the genome on this species is light-regulated. These consensus sequences are conserved in the motifs identified in this analysis (Figure 37) and since only the statistically significant motifs are shown, we can conclude that these putative binding sites might be present in the promoter of the circadian regulation genes in the *Metarhizium* species that were analyzed.

frq adjacent genes

A visual inspection was carried out in each putative ORF upstream region to identify and characterize the promoter region of the *frq* adjacent genes, identifying for the putative *cis* elements. The analysis was carried out using different lengths of the upstream region from the putative ORF for each species.

Left flank genes

Table 16. Localization of the consensus sequences identified for each CREs in the promoter of the frq's left flank genes.

Gene	Species & strain	Early light response element (ELRE) GATC	Late light response element (LLRE) TGA---TCA	TATA Box	ENVOY Upstream motif 1 (EUM1) CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
NCU_02267 Mitochondrial protein FMP25	*Neurospora crassa* OR749A	-35 -291 -1089 -1243 -1341 -1477 -1566 -1747 -2240 -2557 -2632	-2328 -2539 -3016	*	-1177	*	-302 -642 -985 -1085 -1118 -1292 -1439 -1913 -2076 -2386 -2448 -2550 -2721	*
TRIREDRAFT_77656 Phosphoglycerate mutase	*Trichoderma reesei* QM6a	-11	*	*	*	*	-335 -597	*
MAA_04672 Phosphoglycerate mutase	*Metarhizium robertsii* ARSEF23	-25 -467	*	*	*	*	*	*
MAM_04778 2-3, biphosphoglycerate-independent	*Metarhizium album* ARSEF1941	-372	*	*	*	*	*	*

phosphoglycerate mutase								
MAC_01915 2,3-phosphoglycerate-independent phosphoglycerate mutase	*Metarhizium acridum* CQMa102	-475	*	*	*	*	*	*
MAN_09558 2,3-phosphoglycerate-independent phosphoglycerate mutase	*Metarhizium anisopliae* ARSEF549	-25 -467	*	*	*	*	*	*
MBR_03082 2,3-phosphoglycerate-independent phosphoglycerate mutase	Metarhizium brunneum ARSEF 3297	-25	*	*	*	*	*	*
MGU_0475 2,3-phosphoglycerate-independent phosphoglycerate mutase	Metarhizium guizohuense ARSEF 977	-25	*	-313	*	*	*	*
MAJ_05789 2,3-phosphoglycerate-independent phosphoglycerate mutase	*Metarhizium majus* ARSEF297	-359 -801	*	*	*	*	*	*
NOR_03194 2,3-phosphoglycerate-independent phosphoglycerate mutase	Metarhizium rileyi RCEF 4871	-27 -324	*	-276	-358	*	*	*
BBA_08956 Tubulin beta chain	*Beauveria bassiana* ARSEF 2860	-257 -532 -553 -643 -755 -827 -1076 -1294 -1470	-202 -886	*	*	*	*	*

* Consensus sequence not identified in the promoter.

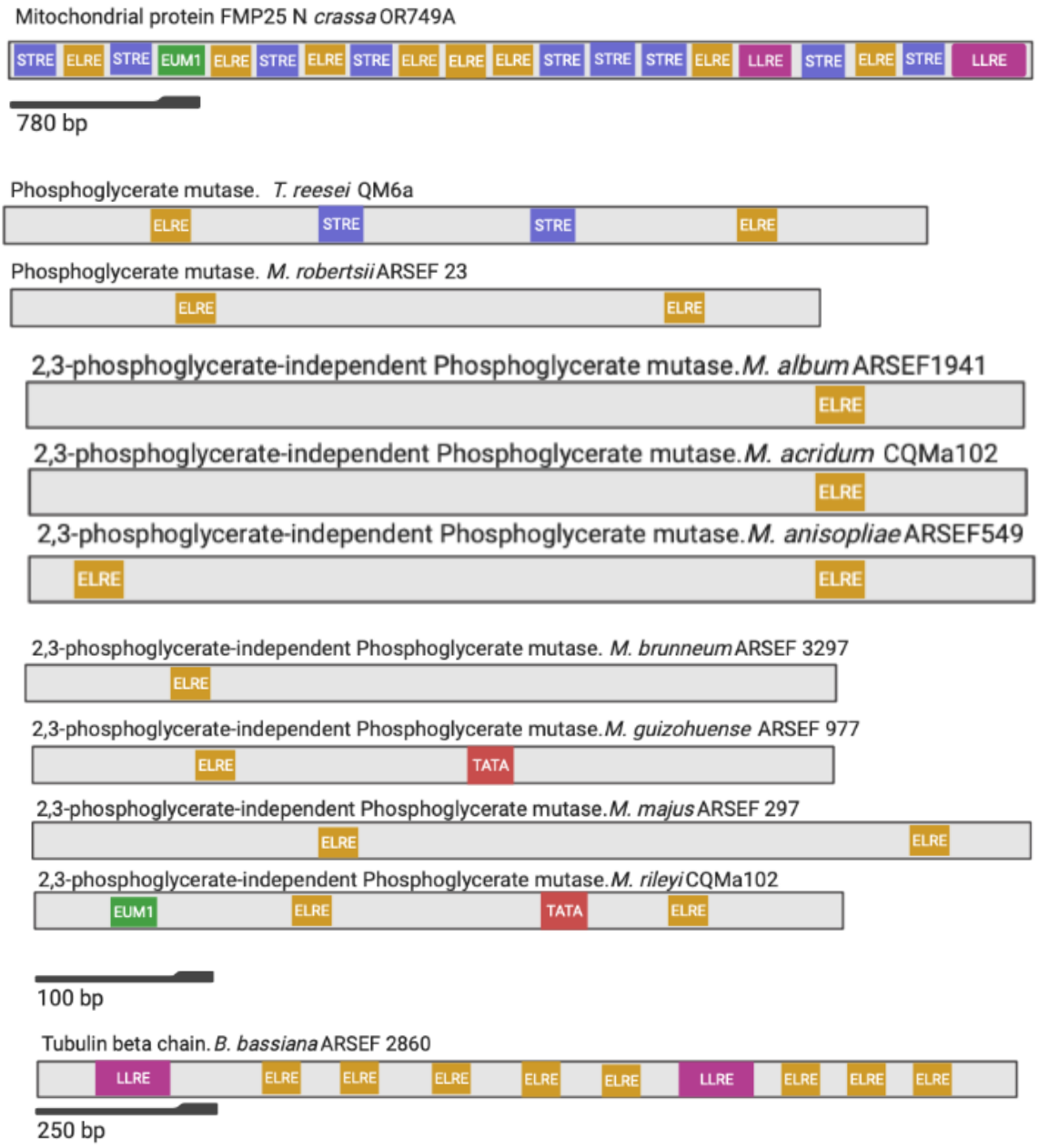


Figure 38. Representation of the cis-regulatory elements identified in the promoter of the frq's left-flank genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

Right flank genes

Table 17. Localization of the consensus sequences identified for each CREs in the promoter of the *frq*'s right flank genes.

Gene	Species & strain	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
NCU_02264 Prefoldin subunit 3	*Neurospora crassa* OR749A	-189 -214 -448 -523	-138 -616	*	*	*	*	*
TRIREDRAFT_47814 Uncharacterized protein (pblast= 97.45% similarity with Prenyl oxidase T. reesei)	*Trichoderma reesei* QM6a	*	*	-21	*	-150	*	*
MAA_04674 Prenyl-cystein oxidase	*Metarhizium robertsii* ARSEF23	*	*	*	*	*	*	*
MAM_04780 Prenyl cysteine oxidase	*Metarhizium album* ARSEF1941	-570	-704	*	-134 -207 -291 -887	-533	*	*
MAC_01917 Prenyl cysteine oxidase	*Metarhizium acridum* CQMa102	*	-101 -237	*	*	*	*	*
MAN_09580 Prenyl cysteine oxidase	*Metarhizium anisopliae* ARSEF549	-137	*	*	*	*	*	*
MBR_03084 Prenyl cysteine oxidase	*Metarhizium brunneum* ARSEF 3297	*	*	*	-310	*	*	*

MGU_0477 Prenyl cysteine oxidase	*Metarhizium guizohuense* ARSEF 977	*	*	*	*	*	*	*
MAJ_05791 Prenyl cysteine oxidase	*Metarhizium majus* ARSEF297	*	*	*	*	*	*	*
NOR_03196 Prenyl cysteine oxidase	*Metarhizium rileyi* RCEF 4871	-59 -1003 -1199	*	-89 -173	-106	*	*	*
BBA_08958 Casein kinase epsilon 1	*Beauveria bassiana* ARSEF 2860	-154 -457	*	*	*	*	*	*

* Consensus sequence not identified in the promoter.

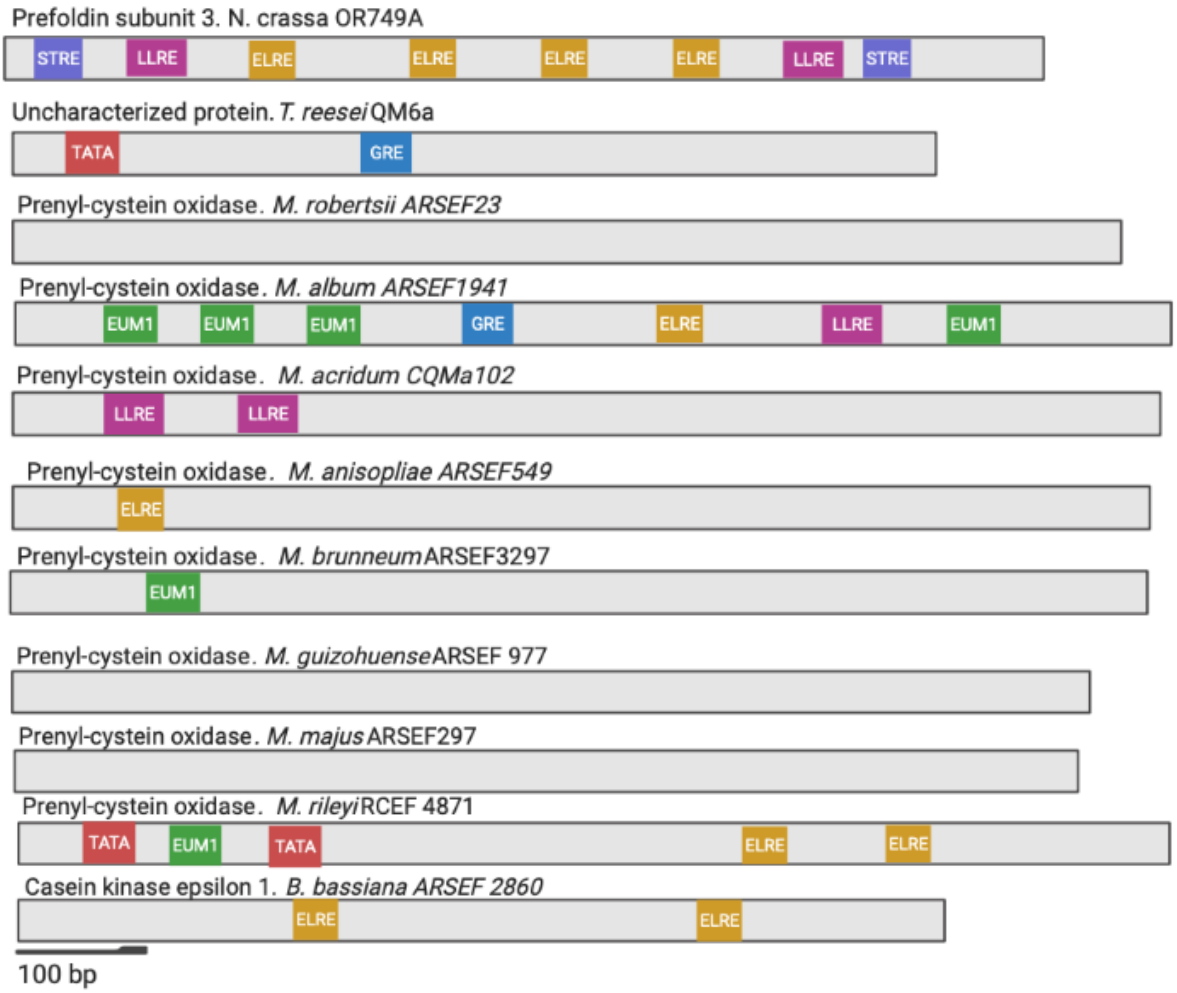


Figure 39. Representation of the cis-regulatory elements identified in the promoter of the frq's right-flank genes.

N. crassa OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

In the genes adjacent to *frq*, LREs in the gene promoters are not perceptible (Figure 38 and 39), therefore the expression mediated by the WCC may not be involved. In fact, the presence of possible cis-regulatory elements doesn't seem to be constant in these genes, mainly in those belonging to the *Metarhizium* species analyzed. In this case, the analyzed genes (prenil-cystein oxidase and phosphoglycerate mutase) are involved in important cellular processes: protein degradation and glycolysis; Therefore, it is expected that the expression of these genes is not mediated by the periodic presence of light in the environment where the fungus develops.

Promoter characterization on early and late responders in *M. robertsii* ARSEF 23

Aguilar-Gordillo (2010) determined several early and late response genes within the genome of *M. robertsii* CARO4, these include genes that are involved in pathogenicity, growth, and germination; among others. The TFIIF gene (an elongation factor) was used as a control, where it was observed that there is no difference in the transcription induction after light stimulation. Using this information, it was proposed to visually inspect the promoters of these genes to couple the experimental evidence collected with the analyzes generated here.

To characterize the promoter region of each gene, a manual search was performed in the complete upstream region of the putative ORF, searching for the consensus sequences of the putative cis-regulatory elements.

Table 18. Localization of the consensus sequences identified for each CREs in the promoter of the early and late response genes.

Gene	Classification (early or late responders) Aguilar-Gordillo, 2010.	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
TFIIF **Transcription factor IIF beta subunit MAA_03949**	No difference	-54 -284	-184	*	-12	*	-241 -498	*
cat1 **Catalase MAA_05879**	Late responder	-790 -865 -1120 -1435 -1893 -2081 -2194 -2996	-26 -79 -1635	-2086 -2911	*	*	-378 -943	-1053
chit2 **Chitinase II MAA_04700**	Early responder	-90	-484	*	*	*	-862	-755
mad2 **Adhesin protein MAA_03807**	Early responder	-359 -641 -1519 -1818	-241 -1069 -1937	*	-2006	*	-327 -1079 -1577	*
mmc (ccp-6 N. crassa) **Mmc protein MAA_06312**	Early responder	199 -392 -635 -681 -777 -1808 -1888	-408 -955	*	-3003	*	-447 -1017 -1345 -1632	*
odc **Ornithine decarboxylase MAA_06162**	Early responder	-12 -321 -1007 -1042 -1932 -2238 -3088	-351 -480 -597 -1059 -2142 -3710	*	*	*	-1674 -2331 -3141	-3693 -3907

		-3158 -3582 -3790 -3931						
Pr1D **Peptidase S8, subtilisin, Asp-active site protein MAA_08718**	Early responder	-443 -457 -706 -1194 -1508	-621 -1884 -1993	-152 -1823 -1833	-80	*	*	*
Pr1H **Vacuolar subtilisin-like protease MAA_10260**	Early responder	-1231 -1824 -1803 -1814 -2578 -3579 -3676 -3684 -4263 -4528 -6218 -6401	-137 -337 -875 -1375 -1869 -2303 -2494 -2703 -2937 -4679 -5134 -5336 -5770 -5841 -6159 -6505 -6543	-56 -177	-4745 -5150	-978	-1118 -1625 -3795 -4160 -4476 -4998	-3193
Pr1j **Peptidase S8, subtilisin, His-active site protein Pr1j MAA_10377**	Early responder	-169 -231 -841 -1325 -1427	-46 -445 -1794 -1884	*	-490	-1910	*	*

* Consensus sequence not identified in the promoter

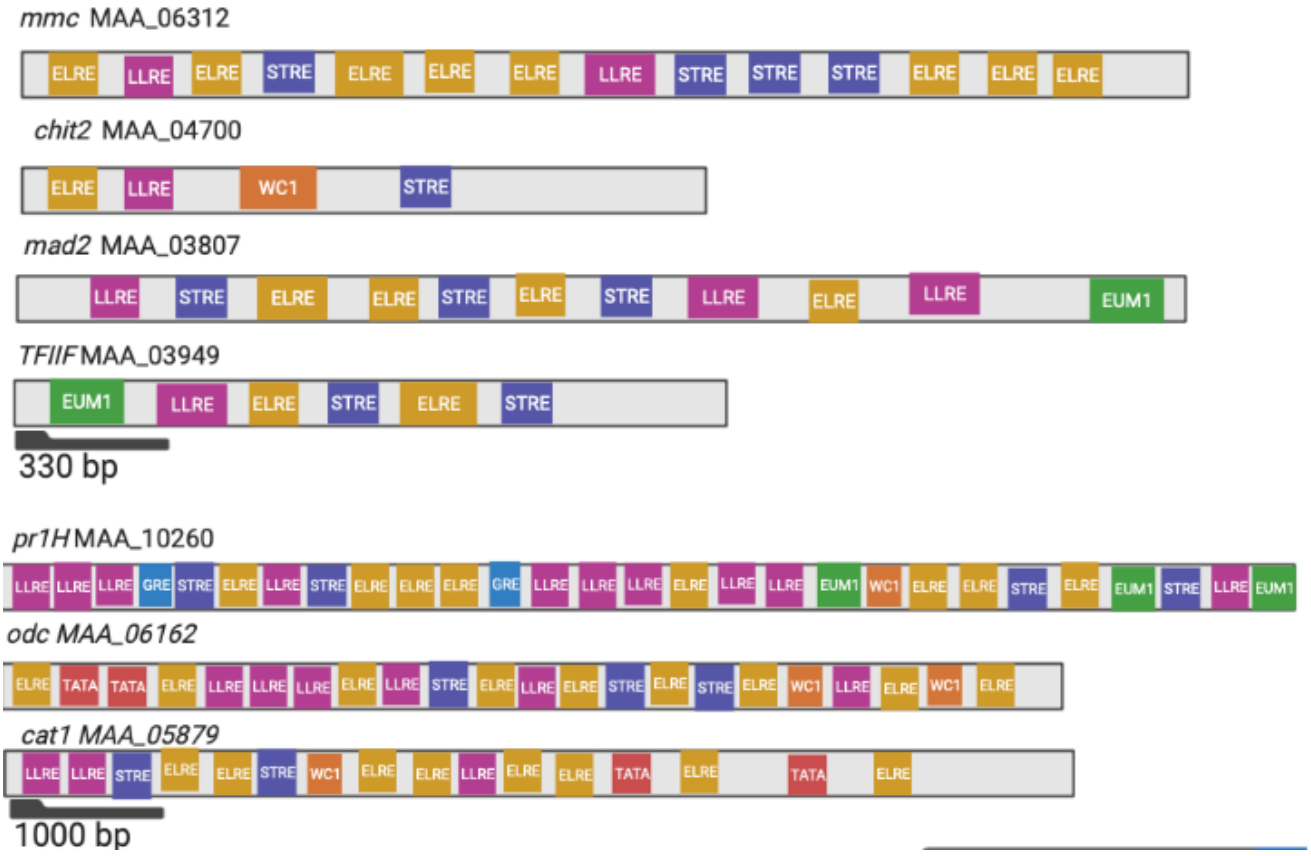


Figure 40. Representation of the cis-regulatory elements identified in the promoter of early and late response genes.

Genes were classified by Aguilar-Gordillo (2010). *N. crassa* OR74A, *T. reesei* QM6a, *B. bassiana* ARSEF 2860, *M. brunneum* ARSEF 3297, *M. robertsii* ARSEF23, *M. acridum* CQMa 102, *M anisopliae* ARSEF 549, *M. rileyi* RCEF 4871, *M. guizohuense* ARSEF 977, *M. majus* ARSEF 297, *M. album* ARSEF 1941

When characterizing the promoters of the genes that were already classified as genes of early or late response, we can see the elements of response to light (Figure 40), this may show that these consensus sequences can be involved in the transcript mediated by the light stimulation and the WCC, this is coupled to the experimental evidence compiled by Aguilar-Gordillo (2010).

5.6 Quantitative analysis of Cis-Regulatory elements in promoter sequences in light response genes.

The light response elements seem to vary in their quantity within the promoters of the different species analyzed. To visualize the differences that can be found within the putative *cis* regulatory elements, an analysis was generated using the ggplot2 package in RStudio. In this case, we analyzed the number of consensus sequences of each of the *cis* regulatory elements that were identified in the promoters.

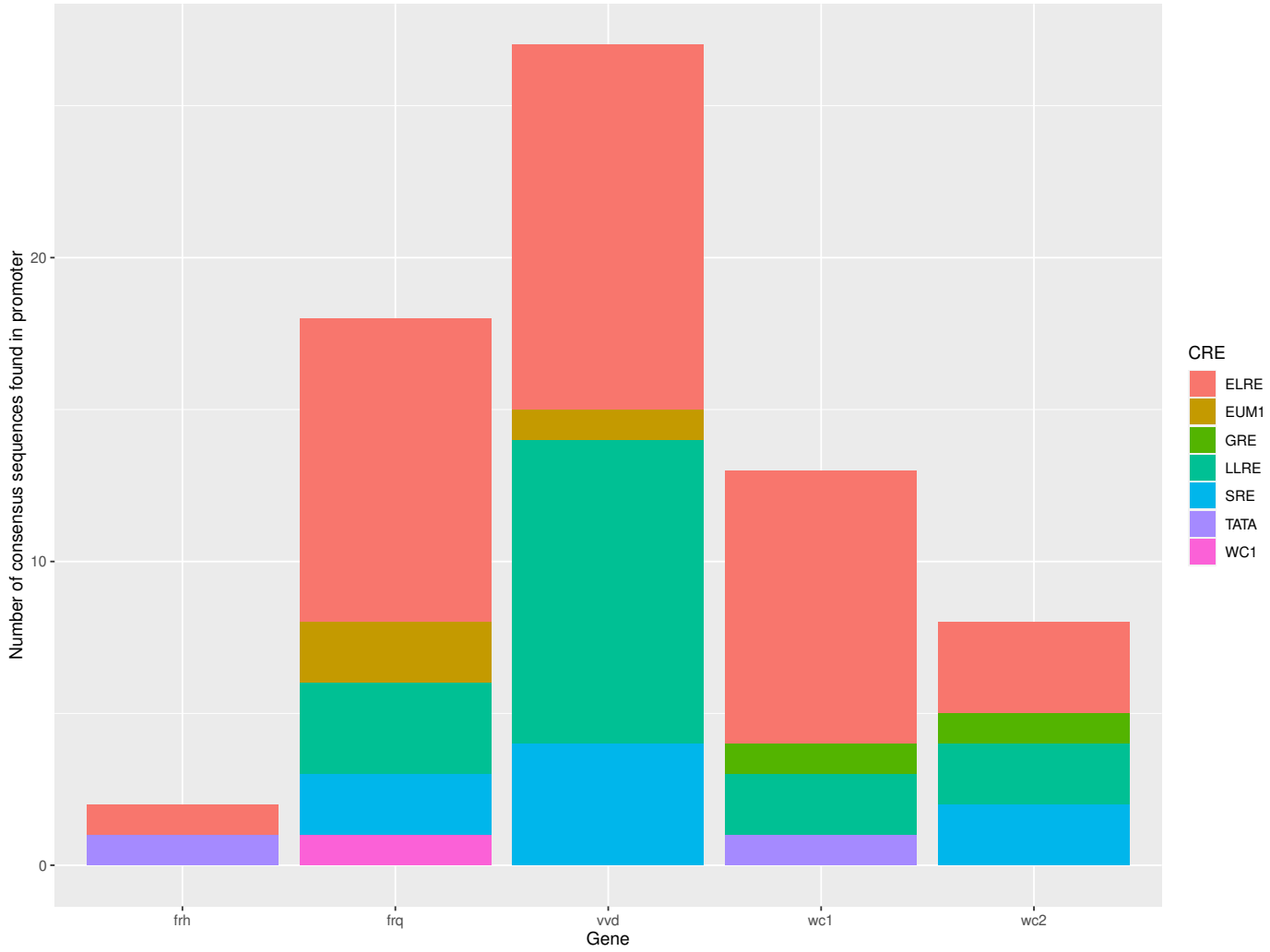


Figure 41. Representation of number of consensus sequences of cis-regulatory elements identified in the promoter of the light response genes in *N. crassa* OR74A.

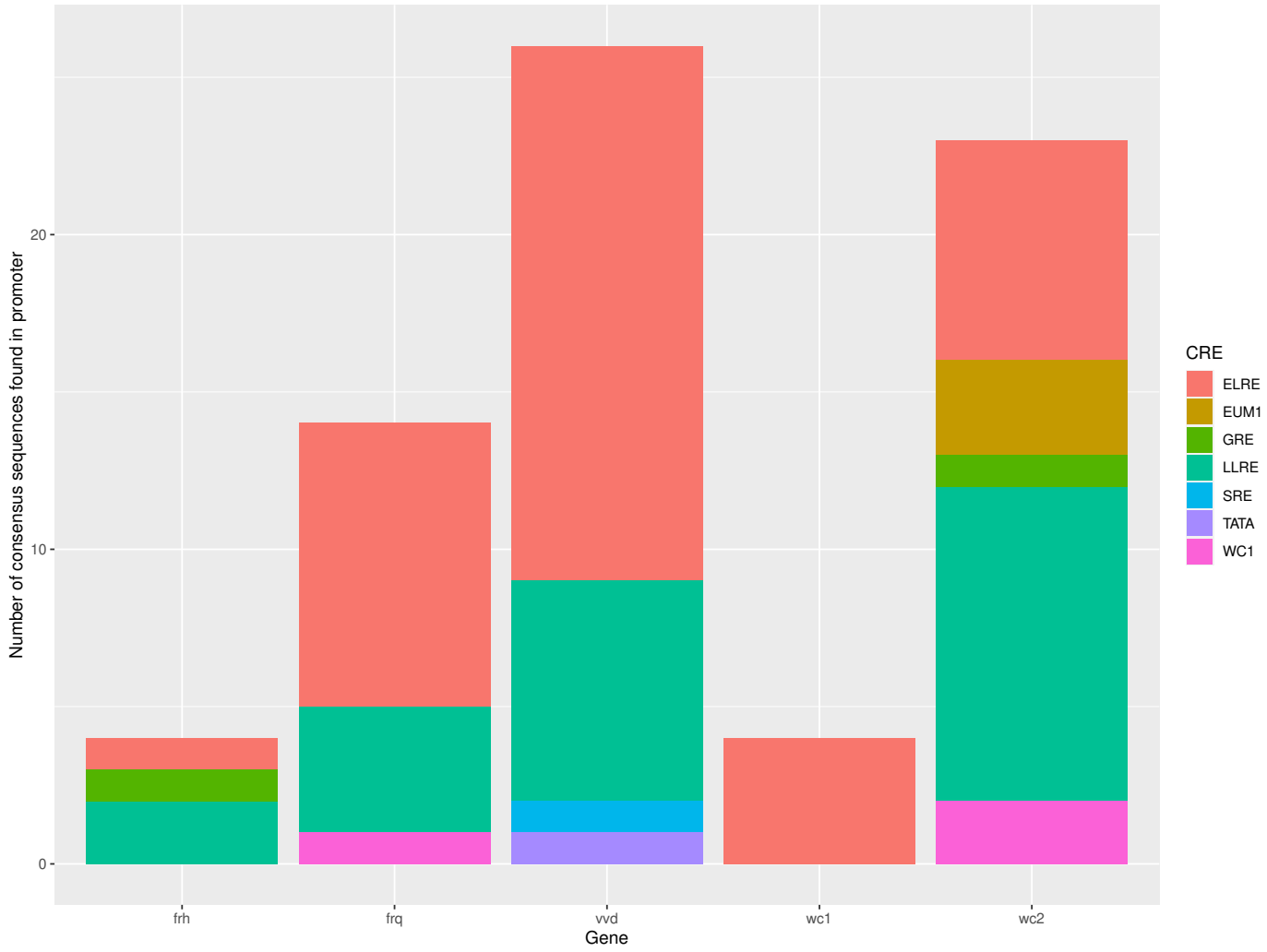


Figure 42. Representation of number of consensus sequences of cis-regulatory elements identified in the promoter of the light response genes in *T. reesei* QM6a.

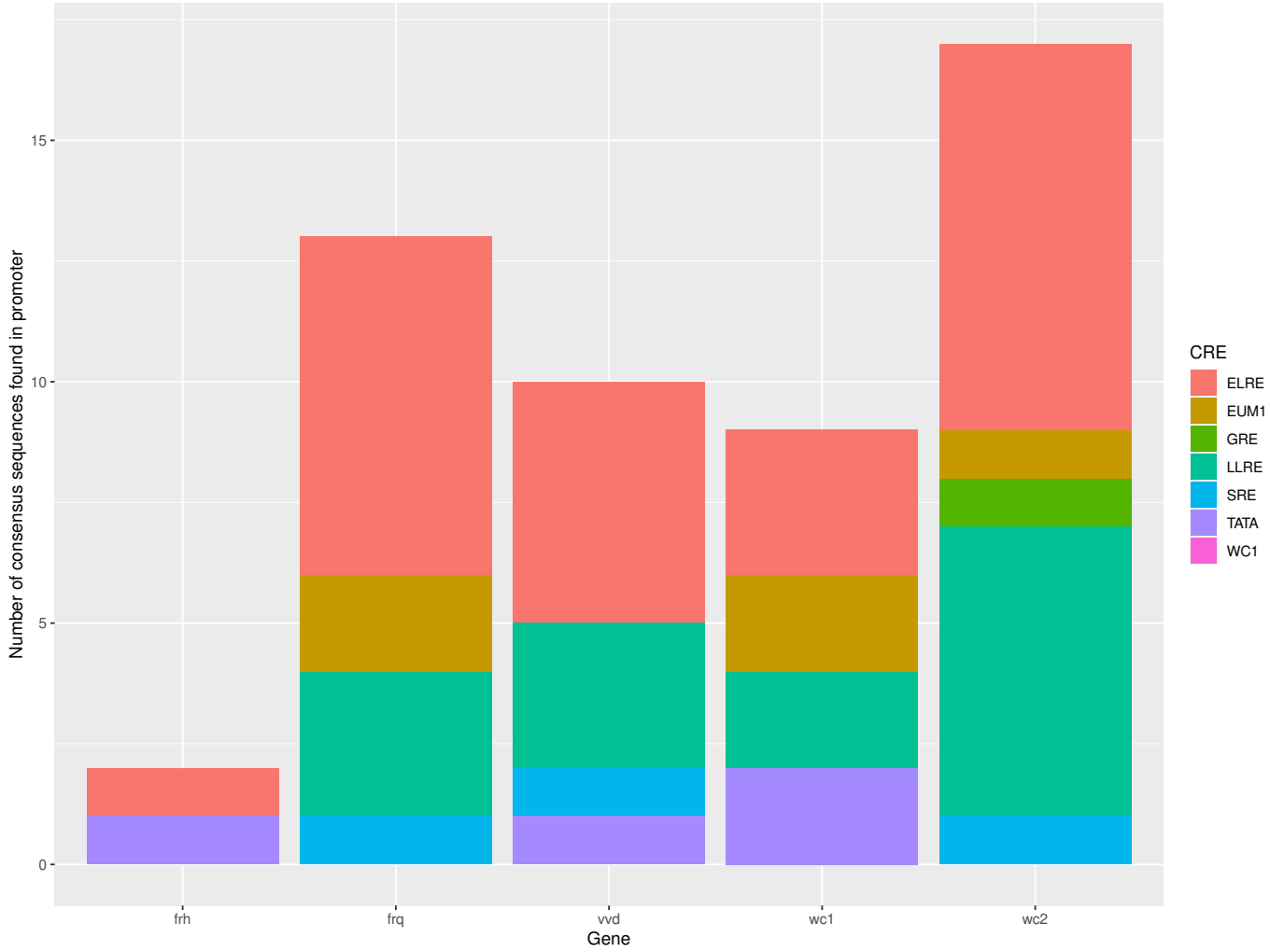


Figure 43. Representation of number of consensus sequences of cis-regulatory elements identified in the promoter of the light response genes in *B. bassiana* ARSEF 2860.

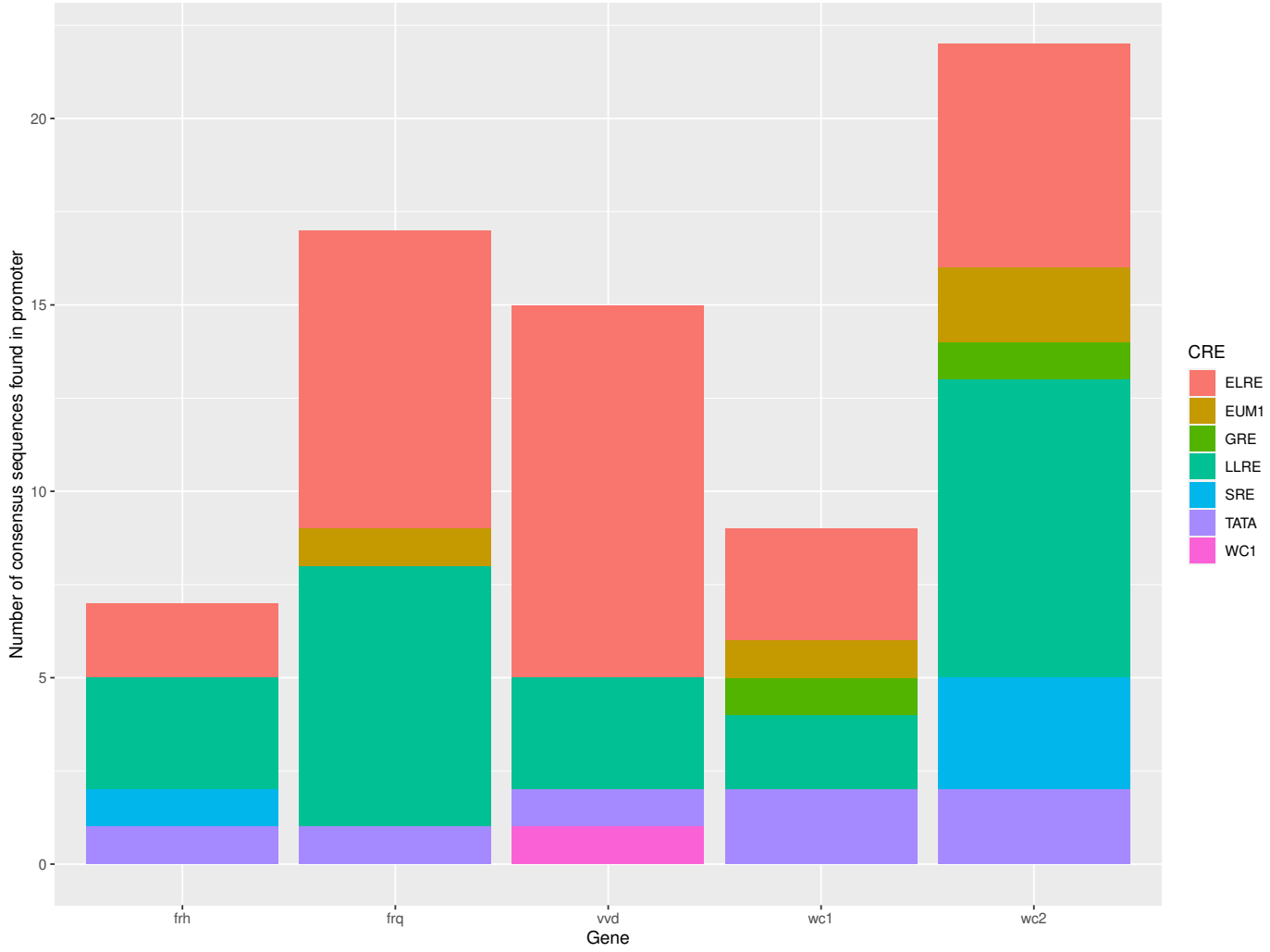


Figure 44. Representation of number of consensus sequences of cis-regulatory elements identified in the promoter of the light response genes in *M. acridum* CQMa102.

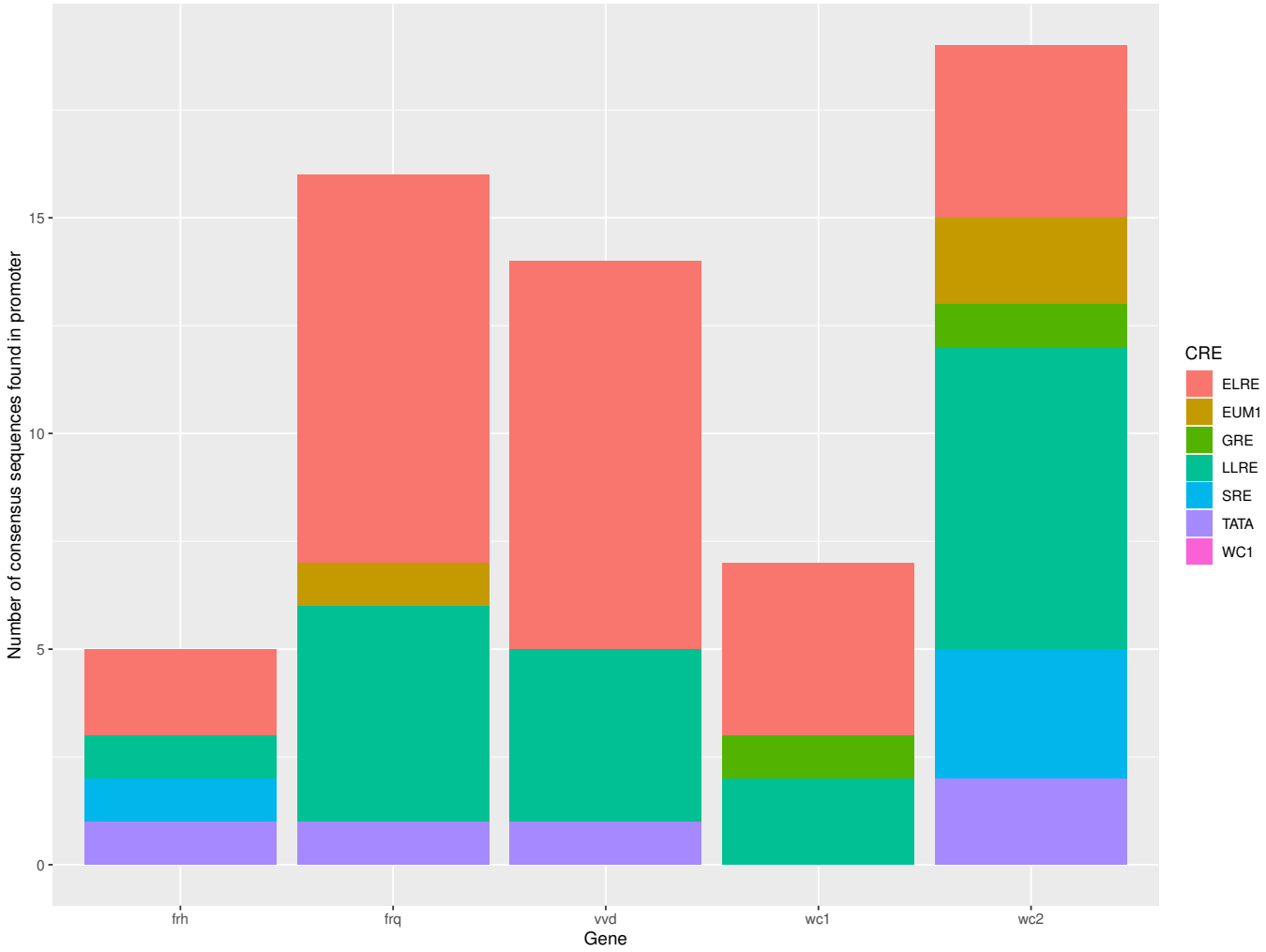


Figure 45. Representation of number of consensus sequences of cis-regulatory elements identified in the promoter of the light response genes in *M. robertsii* ARSEF23.

Differences persist between saprophytic organisms (*N. crassa* and *T. reesei*) and entomopathogenic organisms (*M. robertsii, M. acridum*, and *B. bassiana*). In the first case, it seems that the gene with the most significant evidence of light response elements is *vvd* (in *N. crassa,* Figure 41) and *env1* (in *T. reesei,* Figure 42). This is to be expected since these are among the first genes to be induced to be transcribed by the WCC.

There is a marked difference in the CREs in light response genes in the *Metarhizium* species analyzed (*M. robertsii* and *M. acridum*), the *wc2* gene that is characterized as cutinase protein palindrome-binding protein seem to have a greater amount of LLREs, followed by the ELREs in the *frq* gene, these results are similar to the quantitative analysis in *B. bassiana*, where the gene that seems to present a greater number of CREs is *wc2* followed by *frq*. This could be indicative of the importance of immediate transcription by the WCC once these entomopathogenic organisms detect the light.

5.7 Putative clock-controlled genes (ccg's) in *M. robertsii* ARSEF23

To determine the putative ccgs in the *M. robertsii* ARSEF 23 genome, a BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was performed using as reference the ccg's identified in the *N. crassa* genome (Bell-Pendersen *et al*., 1996). Then, a manual characterization of the promoters of each putative ccg's was generated.

Table 19. Summary on the homologous search for ccg's in the *M. robertsi* ARSEF 23 genome.

Clock-controlled genes in *N. crassa*	**Description**	**Gene in *M. robertsii* ARSEF23 genome**	**E value**	**% Identity**	**Accession**
Ccg-1	Glucose-repressible gene (grg-1)	Glucose repressible protein Grg1	4e-24	63.77%	XP_007818372.1
Ccg-4	Clock-controlled pheromone CCG-4	Clock-controlled pheromone ccg4	6e-27	38.05%	XP_007817373.1
Ccg-6	Clock-controlled gene-6	Mmc protein	1e-36	61.27%	XP_007819180.1
Ccg-7	glyceraldehyde-3-phosphate dehydrogenase-1	Glyceraldehyde/Erythrose phosphate dehydrogenase	0.0	85.37%	XP_007823864.1
Ccg-8	Clock-controlled protein-8	Clock-controlled protein	2e-78	37.47%	XP_007825917.2
Ccg-9	Trehalose phosphorylase	Trehalose synthase-like protein	0-0	56.56%	XP_007825512.2
Ccg-12	Cmt cooper metallothionein (CuMT), has a role on copper storage, detoxification and metal transfer to cooper containing proteins in *N. crassa*. Cmt gene expression is also induced by copper ions	GCN5-like protein	0.83	81.82%	XP_007818039.1

Table 20. Identification of the CREs in the promoters of the putative ccg's in *M. robertsii* ARSEF 23.

Gene	Species & strain	Early light response element ELRE GATC	Late light response element LLRE TGA---TCA	TATA Box	ENVOY Upstream motif 1 EUM1 CTGTGC--CTGTGC	Glucose Response Element (GRE) CACGTG	Stress response elements (SRE) AGGG	WC1 binding site (A/T)GATA(A/G)
Glucose repressible protein Grg1	*M. robertsii* ARSEF 23	-594 -618 -779 -944 -1192 -1722	-281 -628 -699 -855 -1217 -1557 -1838 -1944	*	*	-1882	*	*
Clock-controlled pheromone ccg4	*M. robertsii* ARSEF 23	-398 -1444 -1468 -1638 -1706 -1712 -1746 -2038 -2073	-264 -292 -858 -1591 -1611 -1769 -1824 -1919	*	-761	-1070	-1333	*
Mmc protein	*M. robertsii* ARSEF 23	199 -392 -635 -681 -777 -1808 -1888	-408 -955	*	-3003	*	-447 -1017 -1345 -1632	*
Glyceraldehyde/Erythrose phosphate dehydrogenase	*M. robertsii* ARSEF 23	-749	-970 -1018 -1134	*	*	*	-1160	*
Clock-controlled protein	*M. robertsii* ARSEF 23	-717 -1235 -1883 -2731 -2889 -3047 -3233	-309 -1423 -1507 -2090 -3219 -3549	-2728	-636 -867 -1685	-1025 -3245	-134 -560	*
Trehalose synthase-like protein	*M. robertsii* ARSEF 23	*	*	*	*	-203	*	*
GCN5-like protein	*M. robertsii* ARSEF 23	-136 -433 -710	-332	*	*	*	*	*

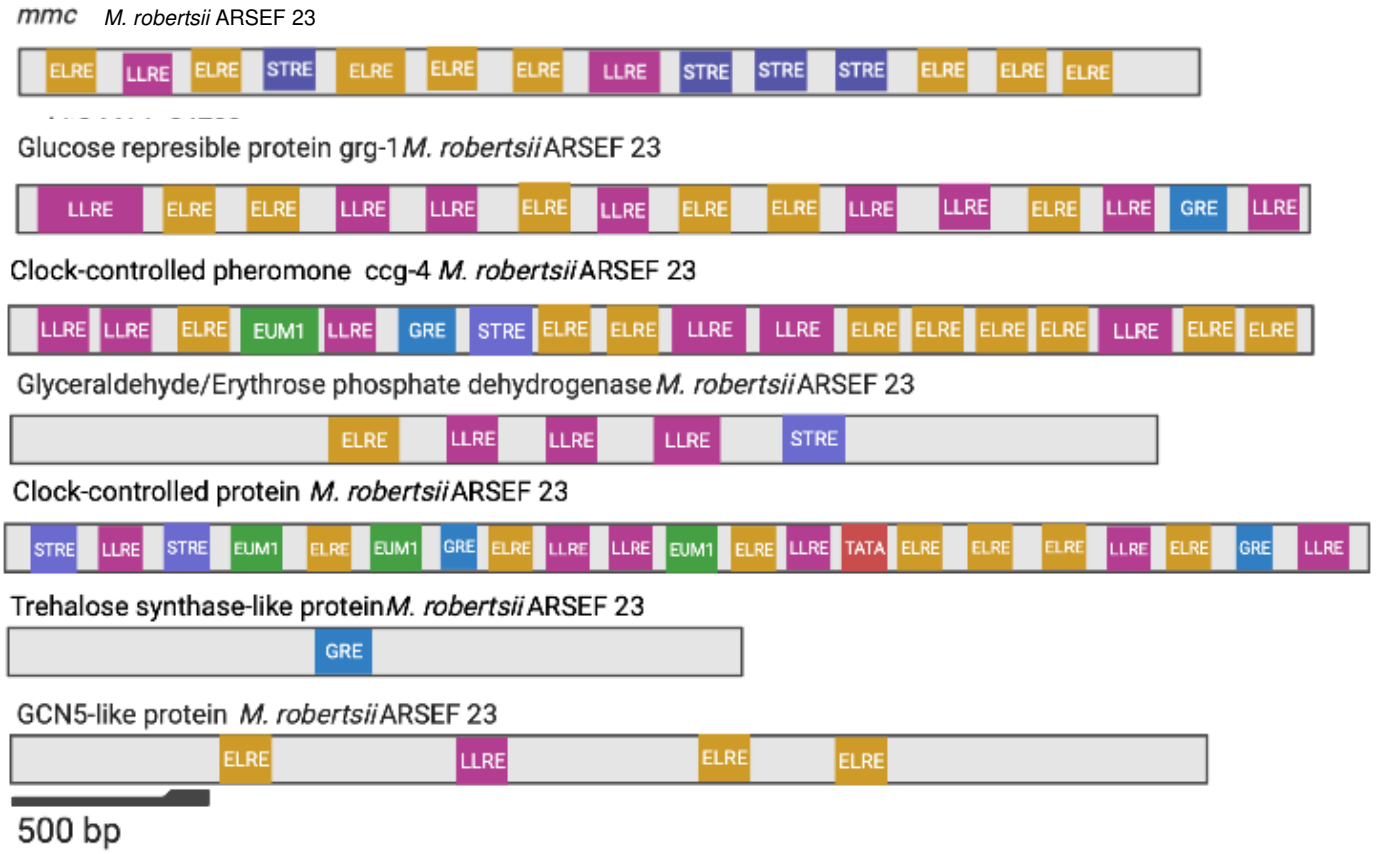


Figure 46. Representation of the promoters in the putative ccg's in *M. robertsii* ARSEF 23

The genes identified share homology with those reported in *N. crassa*. Previous studies in our working group (Aguilar-Gordillo, 2010) had already reported the possibility of the expression of these genes controlled by the clock; mainly the *mmc* gene that has already been characterized as an early response gene due to its induction time in transcription after the light stimulus is presented. However, the function of many of the putative "ccgs" has not yet been described in the *Metarhizium* genus, in addition to the fact that it is necessary to verify that their expression is mediated by the control of the circadian clock.

To characterize the promoter region and the putative -cis regulating elements of the genes involved in the stress responses that are up-regulated under the light stimulus; we choose 2042 bp sequence upstream from the putative ORF of the gene *phr* and for the *tps* gene 2707 bp of the upstream region from the putative ORF.

Table 21. Identification of the CREs in the promoters of the putative ccg's in *Metarhizium*

Up-regulated genes reported in Metarhizium	**Early light response element ELRE GATC**	**Late light response element LLRE TGA---TCA**	**Stress response elements (SRE) AGGG**	**WC1 binding site (A/T)GATA(A/G)**
tps (glycosyltransferase family 20) ***M. robertsii* ARSEF23 MAA_04676 Dias *et al.*, 2019**	-1204 -1578 -2122	-1004 -1985	-153 -1229 -1479 -2126	*
phr **(photolyase) *M. acridum* CQMa 102 MAC_05491 Brancini *et al.*, 2018**	-81 -217 -259 -324 -358 -1379 -1691 -1751 -1939	-438 -1599	-1260 -1532	-1466

Figure 47. Representation of CREs identified in the promoter of putative ccg's in *Metarhizium*

Reports of upregulation in the photolyase gene (*phr*), involved in the photoreactivation, the repair system for UV-damage DNA in fungi under the light stimulus are recent (Brancini *et al*., 2018). In this work, we determinate the presence of early light response elements in the promoter sequence (Table 21, Figure 47); thus, their transcription might be induced by the immediate light stimulus, becoming a plausible early light-respond gene.

Meanwhile, the putative trehalose synthase in *M. robertsii* contains ELLREs, LLREs, and STREs (Table 21, Figure 47) so that the regulation might be mediated by both the light stimulus and stress conditions. It has been reported that conidia of *M. robertsii* accumulate higher amounts of trehalose when is exposed to stressful conditions (Rangel *et al*., 2006), and higher levels of trehalose had been correlated with the stabilization of proteins in their native state and maintaining the membrane's integrity (Singer and Lindquist, 1998); furthermore, trehalose synthase in *N. crassa* is under circadian control; therefore Shinohara *et al*., (2002) have proposed a relationship between *ccg*s and responses to stress. In *Metarhizium*, *tps* might be a

ccg based on the promoter characteristics and the apparent light-induced transcription.

7. Discussion

Homology analysis and phylogenetic reconstructions have shown that the preservation of the light response genes is maintained in *Metarhizium*, this is coupled with the evidence collected (Onofre *et al*., 2001, Brancini *et al*., 2016, Oliveira *et al*, 2017, Brancini *et al*., 2018 and Dias *et al*., 2019) where it has been shown that the light stimulus influences the fungus life cycle. The presence of these homologous in *Metarhizium* proteins suggest a similar function in the process of response to light as illustrated in Figure 48. Notably, FRQ-like proteins in the analyzed genomes draw attention since its function as a circadian regulator has only been described in biological models like *N. crassa;* this presents the probability of a similar system in *Metarhizium.*

Unfortunately, the 3D structure of the FRQ protein from *N. crassa* is not available, therefore designing a homologous model is impossible. However, in this work, we identified the possible domains that may be involved in the function of the FRQ-like proteins (Figure 11), including those that are necessary for their interaction with kinases and the RNA helicase that determines its stability and the import to the nucleus. These results determine that the structural similarities and their maintenance in *M. robertsii* ARSEF 23 bring us closer to the possibility of a similar function.

Regarding the RNA helicase, it is expected that it conserves the characteristic domains of these proteins (Walker A box and DEAD box). Although in *B. bassiana* there is already recent evidence of the importance of the interaction of this helicase with FRQ to maintain the periodic cycle in this fungus (Tong *et al*., 2020); it is necessary to determine its function within the circadian cycles in *Metarhizium* by determining its possible function in the regulation and stabilization of FRQ.

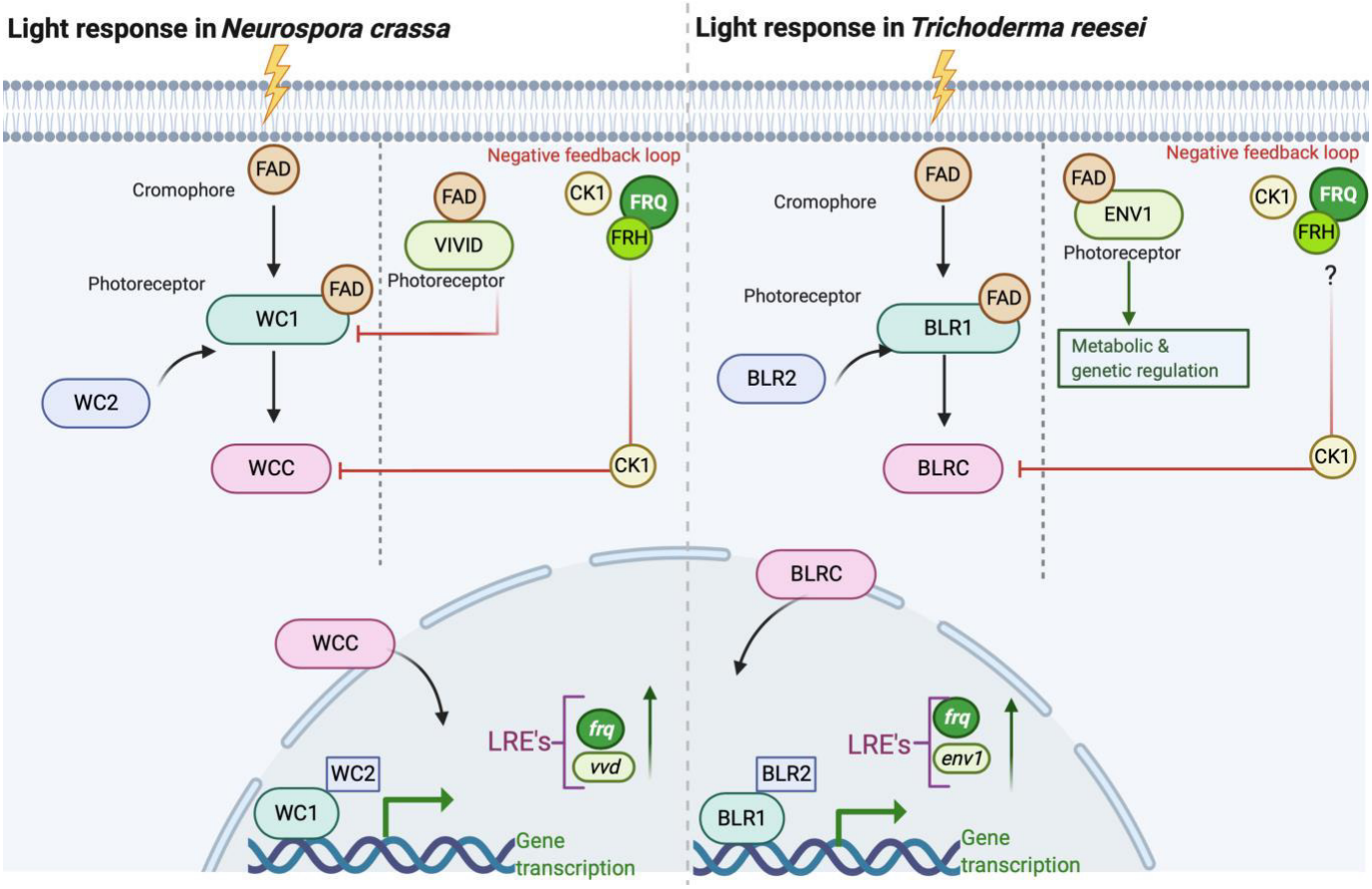


Figure 48. Summary of the light response mechanism in N. crassa and T. reesei.
Modified from Corrochano, 2007; Liu *et al.,* 2019, Casas-Flores & Herrera Estrella 2013

Comparisons involving the sub-genomic regions of the *frq* gene have shown the conservation and similarity of the genes that flank this gene in the analyzed *Metarhizium* species (Table 7). On the other hand, these genes are not involved in the transcription mediated by the light stimulus or by the WCC system.

The synteny analysis carried out with the light response genes does not show evident conservation in the co-linearity in the sub-genomic region in the genomes of the analyzed species, although the microsynteny is maintained among the *Metarhizium* species in the sub-genomic regions where *wc1, cie1, frq* and *frh* genes are located. In the sub-genomic context of the *frq* gene, we can observe that this similarity it is extrapolated to the saprophytic fungus *T. reesei*, surpassing the barrier of genus. The conservation of the collinearity in this region might be useful for future genome assemblies in *Metarhizium*.

It is known that the interaction of the WCC with the LREs in the promoter is necessary for the transcription of the light response genes (Cheng *et al*., 2009). When analyzing the genes that were characterized by Aguilar-Gordillo we can observe that these elements remain in the promoter of the early and late response genes (Figure 40 and Table 22). Although, there is no evidence of the interaction of this WCC to the LREs in *Metarhizium,* we can infer that this process may be happening similar to *N. crassa*. Therefore, analyzing and characterizing the putative cis- regulatory elements of the light response genes seems a viable approach. When characterizing the CREs of these specific genes, we can find, mainly in the genes homologous to *wc2,* that these LREs stand out. In the genes homologous to *vvd* the presence of ELREs is clear; this could be indicative of the immediate transcription of this gene.

Although there are discrepancies in the characterization of the genes coding this containing PAS domain protein, these elements appear constant. Synteny analysis place these genes in 2 groups (Figure 17), which are most similar to ENVOY (*M. rileyi, M. album* and *M. acridum*) and similar to VIVID (*M. anisopliae*, *M. robertsii*, *M.*

brunneum, *M. guizohuense and M. majus*). Therefore, the function of these genes could be specie- dependent.

Meanwhile, in the sub-genomic region where we find the genes homologous to *wc1*, in *Metarhizium*, we can observe that the synteny remains relatively preserved; in addition, apparently this similarity is shared within the sub-genomic region where the *wc1* gene is in the *N. crassa* genome (Figure 15).

The quantitative analysis of the CREs in the different light response genes determines a marked difference between saprophytic and entomopathogenic fungi. In the first case*, vvd/env1* seems to be the genes that the action of these LREs could strongly regulate; in both, the amount of LRES stands out. In entomopathogenic fungi (*B. bassiana, M. robertsii* and *M. acridum*) *wc2* seems to be the gene that contains a more significant number of CREs in its promoter (Figures 43, 44 and 45). However, the amount of LREs is more significant in *frq* and *vvd*; therefore, the transcription mediated by the light stimulus and by the WCC could be necessary for these genes. It is worth mentioning that there is no visible difference between the LREs between specialists (*M. acridum*) and generalist (*M. robertsii*) entomopathogenic fungi.

Finally, we sought to determine putative clock-controlled genes in *Metarhizium*, the genes were identified by homology analysis by those that have already been described and cataloged in the *N. crassa*. In addition to those genes, we selected genes that have shown overexpression in growth under light stimulation in species belonging to *Metarhizium*. In this way, the mmc, glucose repressible protein, clock-controlled pheromone, glyceraldehyde/erythrose phosphate dehydrogenase, clock-controlled protein, GNC5-like protein, photolyase and finally trehalose synthase, genes were analyzed, determining the possible transcription mediated by the light stimulus and the WCC (Tables 20 and 21).

Table 22. Summary of the characterization and function of early and late response genes by Aguilar-Gordillo (2010).

Gene	Classification (early or late responders) Aguilar-Gordillo, 2010.	Function
TFIIF **Transcription factor IIF beta subunit** **MAA_03949**	No difference	Elongation factor.
cat1 **Catalase** **MAA_05879**	Late responder	Detoxification. Involved in cellular detoxification processes.
chit2 **Chitinase II** **MAA_04700**	Early responder	Pathogenicity Conidia formation on host.
mad2 **Adhesin protein** **MAA_03807**	Early responder	Pathogenicity. Involved in conidiation process and favors conidial adherence to new host.
mmc ***(ccp-6 N. crassa)*** **Mmc protein** **MAA_06312**	Early responder	Putative involvement in circadian regulation.
odc **Ornithine decarboxylase** **MAA_06162**	Early responder	Conidiation. Appressoria formation and growth a conidiation.
Pr1D **Peptidase S8, subtilisin, Asp-active site protein** **MAA_08718**	Early responder	Pathogenicity Protease involved in early pathogenic events (penetration on host exoskeleton)
Pr1H **Vacuolar subtilisin-like protease** **MAA_10260**	Early responder	Pathogenicity. Protease involved in conidia formation in host exoskeleton
Pr1j **Peptidase S8, subtilisin, His-active site protein** **Pr1j** **MAA_10377**	Early responder	Pathogenicity. Protease overexpressed in fungi emersion process.

Table 23. Summary characterization and function of the putative clock-controlled genes and early and late response genes classified in this work.

Putative clock-controlled genes in *Metarhizium.*	**Early or late responder classification (This work)**	**Function**
tps (glycosyltransferase family 20) *M. robertsii* ARSEF23 MAA_04676 Dias *et al.,* 2019	Late responder	Stress response. Protein stabilization
phr (photolyase) *M. acridum* CQMa 102 MAC_05491 Brancini *et al.*, 2018	Early responder	UV tolerance. Photoreactivation process (UV repair mechanisms)
Glucose repressible protein Grg1 *M. robertsii ARSEF23* *MAA_02183*	Early responder	Protein increases during glucose deprivation. Its function its unknown.
Clock-controlled pheromone ccg4 *M. robertsii ARSEF 23* *MAA_01184*	Early responder	Function its unknown.
Glyceraldehyde/Erythrose phosphate dehydrogenase *M. robertsii ARSEF 23* MAA_07675	Late responder	Involved in glycolysis. This protein has adhesion-like activity and is believed to be involved in the adhesion of *M. anisopliae* conidia to the surface of the host insect (Broetto *et al.*, 2010).
Clock-controlled protein *M. robertsii ARSEF 23* MAA_09728	Early responder	Its function is unknown. CD identification shows an Opi1 domain (328-675aa), involved in transcription factors that negatively regulates phospholipid biosynthesis in yeast.
GCN5-like protein *M. robertsii ARSEF 23* MAA_01850	Early responder	Function unknown.

The functions of these genes within the metabolism of *Metarhizium* is varied, many of these proteins act as response genes to environmental stress (*Brancini et al*., 2018, Dias *et al.,* 2019) while others are involved in the growth of the fungus (Broetto *et al*., 2010). Coupling these results with the previous characterization (Aguilar-Gordillo, 2010) where several genes were determined that are also involved in these stress response processes, fungal growth and pathogenicity.

To specify these results, a putative expression model was generated where the action of the WCC and the putative circadian clock in *Metarhizium* is involved (Figure 50).

Putative circadian rhythms and blue light response mechanism in *Metarhizium.*

The compilation of bibliographic, bioinformatics, and experimental evidence leads us to propose a mechanism in the genetic regulation mediated by the response to light, involving putative circadian rhythms in *Metarhizium.* This arises from the subjective time; where we propose the synthesis of FRQ in the morning, forming the complex with FRH, acting as a regulator of the WCC when imported into the nucleus; where it favors the recruitment of kinases that will act on the WCC, inactivating the complex. Following this inactivation, accumulation of WCC happens in the cytoplasm. Meanwhile, the levels of *frq* transcripts decrease, so at night FRQ is degraded, favored by the hyperphosphorylation of the protein, releasing its inhibitory effect on the WCC, allowing to act as transcription factor favoring the transcription of *frq* (Figure 49).

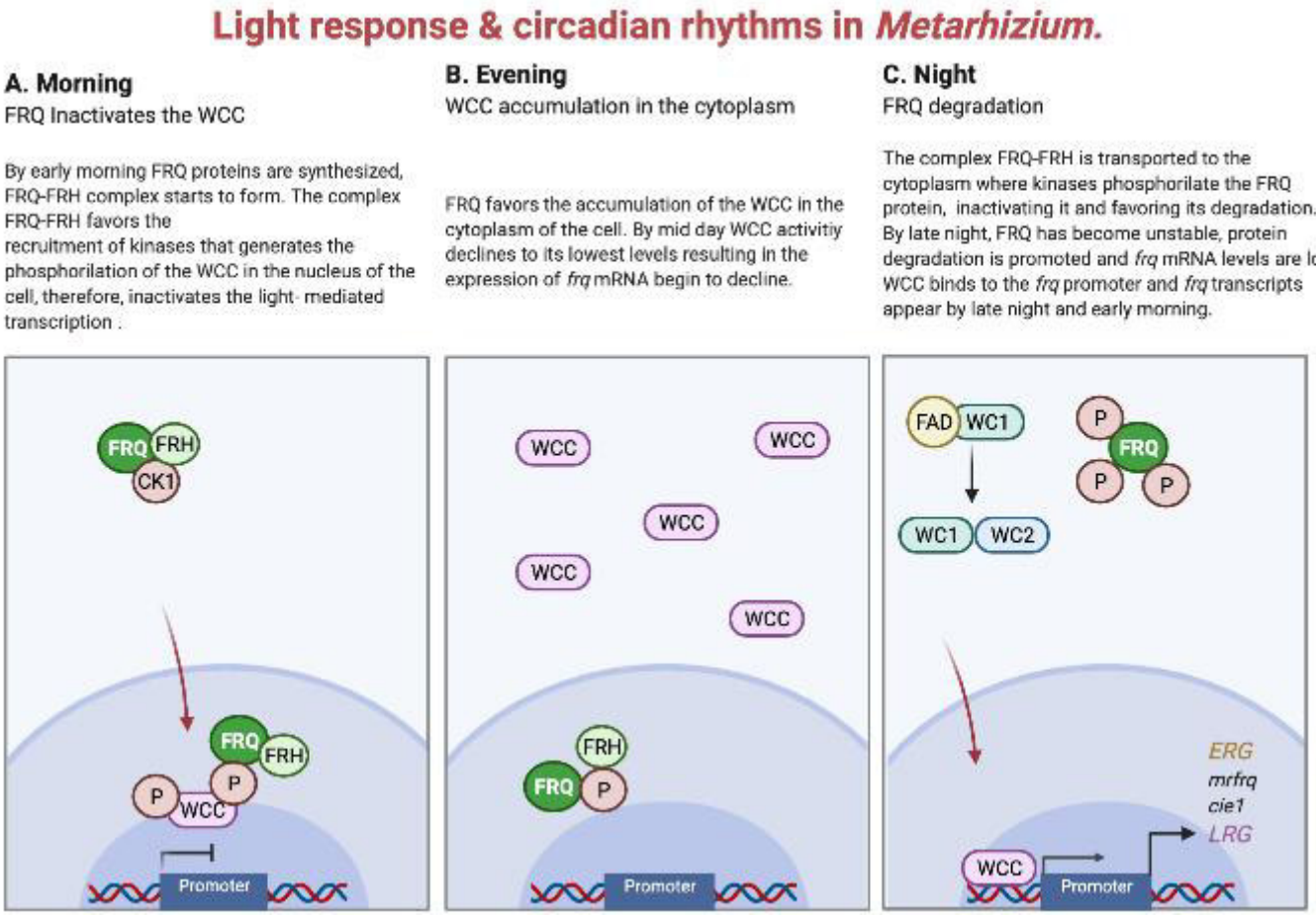


Figure 49. Proposed mechanism for the putative circadian rhythms and light response in *Metarhizium*.

In addition to this mechanism, a complement with the activity of WCC is proposed (Figure 50), mediated by the conditions of darkness or light, in the dark WCC acts favoring the transcription of *frq* and ccgs, that determines the function and maintain circadian rhythms.

Under light stimulus, *Mrvvd* (CIE1) could inactivate WCC, restarting the circadian clock by decreasing the levels of *frq*. At the same time, the action of WCC in the early response genes (ERG) would have a phenotypic effect that various authors have widely evidenced and characterized by i) an increase in resistance to UV radiation and osmotic stress, ii) an increase in the production and germination of conidia and iii) an increase in the virulence of the fungus.

To facilitate the resistance to environmental stress, the overexpression of photolyases may be favoring this characteristic in the case of being induced early by light through the binding of WCC to the ELREs that have been detected in its promoter region in this work, while the overexpression of the *tps* gene coding for trehalose synthase could be induced by light stimulation or by other conditions of environmental stress (osmotic, thermal stress, etc.), this could justify the high concentrations of trehalose in conidia resulting from growth under light and nutritional stress reported by Dias *et al*., (2019)

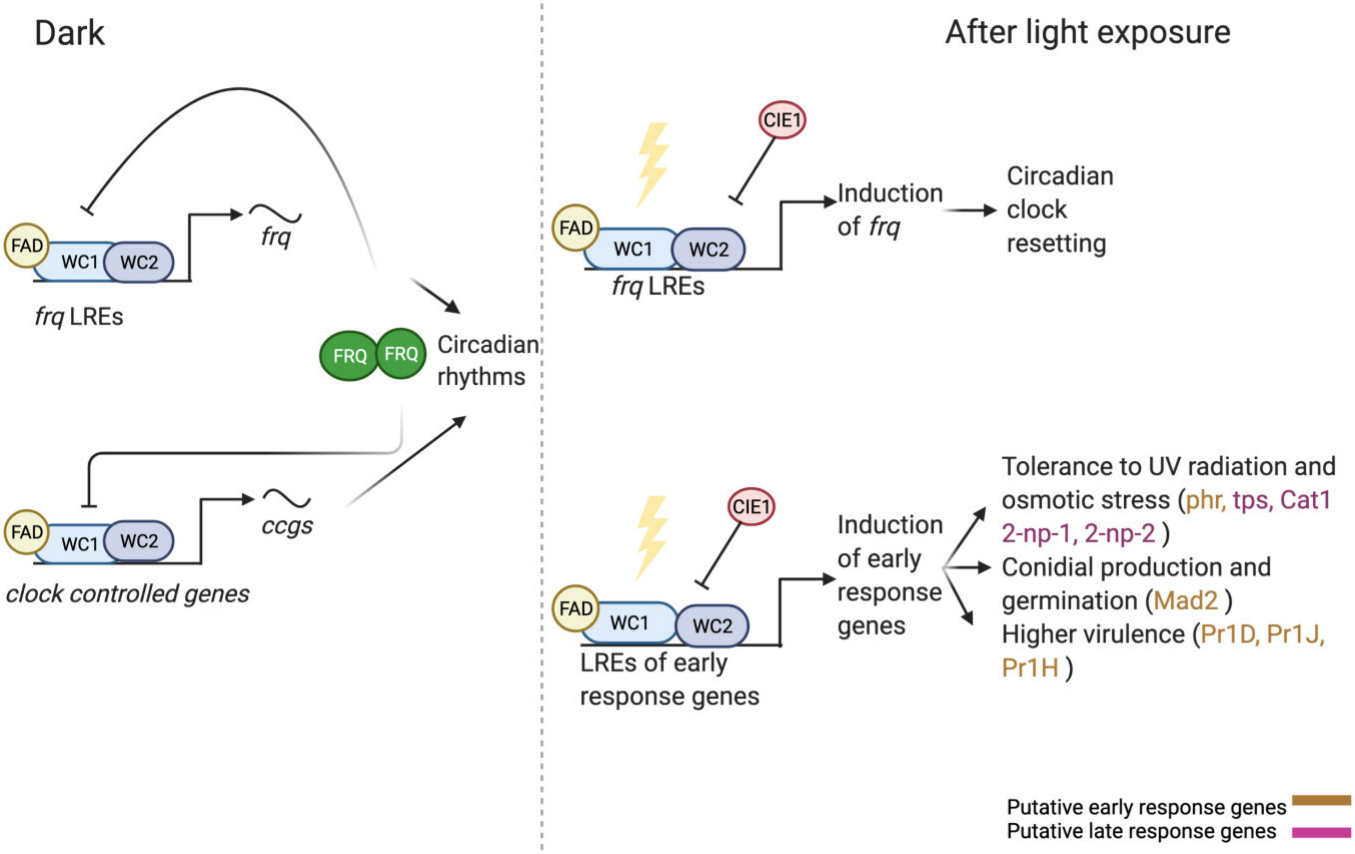


Figure 50. Proposed mechanism for the WCC and the phenotypic effects observed in *Metarhizium* (Onofre *et al.,* 2001, Brancini *et al.,* 2016, Oliveira *et al.,* 2017, Brancini *et al*., 2018, Aguilar-Gordillo, 2010 and Dias *et al.,* 2019) (modified from Liu et al., 2003)

The involvement of the gene control that influences the conidiation in *Metarhizium* has been broadly approached. One of the recent elements that mediated this biological process is the involvement of transcription factors (TF). Son *et al*., (2019) determined the role of the TF *MaCreA* in the regulation of the conidiation by deleting this gene; they were able to determine a decrease in the conidial production and yield, pointing out that this TF might be involved in the carbon catabolism regulation (CCR) that negatively regulates the genes in carbon sources.

Another example of the role of TFs in *Metarhizium* is the function of MrMsn2 in the anamorph fungi *M. rileyi*, where this TF controls the yeast to hyphae transition, the conidia and microsclerotia formation and influences the virulence of this specialist entomopathogen fungus. Deleting this TF generates a strain sensitive to stress, reducing the microsclerotia formation and abnormalities in the morphotype and virulence levels (Song *et al*., 2018). Considering these examples of the control in the fungus *Metarhizum* mediated by the action of transcription factors, the function of the WCC may be similar, generating the phenotypic effects described. However, it is essential to expand the study by compiling experimental evidence on the effects of the WCC in *Metarhizium* to demonstrate and support the function proposed here. Different regulatory mechanisms highly orchestrate the action of TFs; in this case, FRQ could be one of the central regulators.

Notably, recent studies (Peng *et al.,* 2021) have demonstrated the action of WCC in *M. robertsii;* managing to characterize its cytoplasmic and nuclear location dependent on light stimulation, mainly blue light. An important characteristic is that most of the orthologous proteins to WC1 in *Metarhizium* does not have a zinc finger domain as in *N. crassa*; therefore, it seems that WCC depends on the action of WC2 to act as a transcription factor. Peng *et al*., (2021) also demonstrated that the photoreactivation process in *M. robertsii* is dependent on the action of WCC, where two photolyases act to repair the damage caused by UVB radiation, highlighting that the expression of *phr2* is dependent on the WC1 and *phr1* depends on WC2. These recent results are coupled with the assertions made in this work by the bioinformatic analysis of the photolyase in *M. acridum*, where we proposed that its expression is very possibly dependent on the WCC by identifying putative CREs in its promoter.

However, experimental evidence is required to affirm this, as well as to identify the union of the WCC to the LREs in the genus *Metarhizium*.

8. Conclusions

- Homology analyses demonstrate the presence of light response genes in the genomes of different species of *Metarhizium*.
- In the sub-genomic context of these genes, we can find a microsyntenic conservation, mainly in the localization of the *frq* gene, and clear differences (*wc2* gene specifically). In addition, the similarities of the sub-genomic region where *frq* is located is not exclusive to the *Metarhizium* genus.
- Early and late light response genes demonstrate a wide distribution of putative CREs, mainly LREs; in their promoter region, this is coupled to the experimental results previously presented.
- Continue with the strategy of identifying the putative *cis*- regulatory elements allowed us to determine other possible clock-controlled genes and early or late response genes. These identified genes could be involved in the phenotype characterized by growth under the light stimulus.
- Finally, quantitative analyzes on the distribution of CREs among light response genes promoters demonstrate a difference between the saprophyte and entomopathogenic fungi analyzed, where the possible control is differential in its expression mediated by the CREs, highlighting the possible importance in the FRQ gene expression in these organisms.

9. Perspectives

This work has been based on evidence and bioinformatic analysis; therefore, it is necessary to generate experimental strategies that allow to refute or assert the inferences carried out.

i) Initially, it is necessary to determine the presence of the *mrfrq* gene within the genome of *Metarhizum* and characterize its encoding protein.

ii) Possibly generate the deletion and determine the existence of some characteristic phenotype under the deletion of this gene or determine the presence of independent FRQ oscillators (FLOs) in *Metarhizum*, as it happens in *N. crassa.*

iii) It is also necessary to demonstrate the periodic transport of this protein inside and off the nucleus for completion of its possible function as a WCC regulator.

Recently in our working group, a phenotype has been demonstrated that presents rings of conidiation (data unpublished) under the photoperiod of the CARO4 strain of *M. robertsii*, therefore, it is possible that a differential phenotype may be presented under different environmental conditions that work as signals for the circadian regulation in this fungus.

iv) Thus, determining the effects of environmental changes (temperature, light, and even in interactions with plants and insects) are necessary since they could present ourselves the possibility of the expansion of circadian rhythms as we know them.

10. References

1. Aguilar-Gordillo S. (2010). Estudio de la conidiación en *Metarhizium anisopliae* en respuesta luz. Tesis de Maestría, Universidad de Guanajuato.

2. Andreani, T. S., Itoh, T. Q., Yildirim, E., Hwangbo, D. S., & Allada, R. (2015). Genetics of circadian rhythms. Sleep medicine clinics, 10(4), 413-421.

3. Angelone, S., Piña-Torres, I. H., Padilla-Guerrero, I. E., & Bidochka, M. J. (2018). "Sleepers" and "Creepers": A theoretical study of colony oolymorphisms in the fungus *Metarhizium* related to insect pathogenicity and plant rhizosphere colonization. *Insects*, *9*(3), 104.

4. Aronson, B. D., Johnson, K. A., & Dunlap, J. C. (1994). Circadian clock locus frequency: protein encoded by a single open reading frame defines period length and temperature compensation. Proceedings of the National Academy of Sciences, 91(16), 7683-7687.

5. Ballario, P., Vittorioso, P., Magrelli, A., Talora, C., Cabibbo, A., & Macino, G. (1996). White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein. The EMBO journal, 15(7), 1650-1657.º

6. Bell-Pedersen, D., Shinohara, M. L., Loros, J. J., & Dunlap, J. C. (1996). Circadian clock-controlled genes isolated from N*eurospora crassa* are late night-to early morning-specific. Proceedings of the National Academy of Sciences, *93*(23), 13096-13101.

7. Brancini, G. T., Bachmann, L., Ferreira, M. E. D. S., Rangel, D. E., & Braga, G. Ú. (2018). Exposing *Metarhizium acridum* mycelium to visible light up-regulates a photolyase gene and increases photoreactivating ability. Journal of invertebrate pathology, 152, 35-37.

8. Brancini, G. T., Ferreira, M. E., Rangel, D. E., & Braga, G. Ú. (2019). Combining transcriptomics and proteomics reveals potential post-transcriptional control of gene expression after light exposure in *Metarhizium acridum*. G3: Genes, Genomes, Genetics, 9(9), 2951-2961.

9. Broetto, L., Da Silva, W. O. B., Bailão, A. M., De Almeida Soares, C., Vainstein, M. H., & Schrank, A. (2010). Glyceraldehyde-3-phosphate dehydrogenase of the entomopathogenic fungus *Metarhizium anisopliae:* cell-surface localization and role in host adhesion. *FEMS microbiology letters*, *312*(2), 101-109.

10. Casas-Flores, S., Herrera-Estrella, A., Mukherjee, P., Horwitz, B., & Singh, U. (2013). The influence of light on the biology of *Trichoderma*. Trichoderma: biology and applications, 43-66.

11. Corrochano, L. M. (2007). Fungal photoreceptors: sensory molecules for fungal development and behaviour. Photochemical & Photobiological Sciences, 6(7), 725-736.

12. Corrochano, L. M. (2019). Light in the fungal world: from photoreception to gene transcription and beyond. Annual Review of Genetics, 53, 149-170.

13. Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome research, *14*(7), 1394-1403.

14. Daub, M. E., & Ehrenshaft, M. (2000). The photoactivated *Cercospora* toxin cercosporin: contributions to plant disease and fundamental biology. Annual review of phytopathology, *38*(1), 461-490.

15. De Fabo, E. C., Harding, R. W., & Shropshire, W. (1976). Action spectrum between 260 and 800 nanometers for the photoinduction of carotenoid biosynthesis in *Neurospora crassa*. Plant Physiology, 57(3), 440-445.

16. Diernfellner AC, Schafmeier T, Merrow MW, Brunner M (2005). Molecular mechanism of temperature sensing by the circadian clock of *Neurospora crassa*. Genes & Development. 19:1968–1973. [PubMed: 16107616]

17. Dunlap, J. C., & Loros, J. J. (2006). How fungi keep time: circadian system in *Neurospora* and other fungi. Current opinion in microbiology, 9(6), 579-587.

18. Froehlich, A. C., Liu, Y., Loros, J. J., & Dunlap, J. C. (2002). White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. Science, 297(5582), 815-819.

19. Fuller, K. K., Loros, J. J., & Dunlap, J. C. (2015). Fungal photobiology: visible light as a signal for stress, space and time. Current genetics, 61(3), 275-288.

20. Galperin, M. Y., Koonin, E. V., & Bairoch, A. (1998). A superfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases. Protein Science, 7(8), 1829-1835.

21. Gao, Q., Jin, K., Ying, S.H., Zhang, Y., Xiao, G., Shang, Y., Duan, Z., Hu, X., Xie, X.Q.,Zhou, G., Peng, G., Luo, Z., Huang, W., Wang, B., Fang, W., Wang, S., Zhong, Y., Ma, L.J., St Leger, R.J., Zhao, G.P., Pei, Y., Feng, M.G., Xia, Y., Wang, C., (2011). Genome sequencing and comparative transcriptomics of the model entomo- pathogenic fungi *Metarhizium anisopliae* and *M. acridum*. PLoS Genet. 7, e1001264 https://doi.org/10.1371/journal.pgen.1001264.

22. Gasca-Venegas, A. (2016) Estudio de la conidiación en respuesta a un estímulo luminoso en *Metarhizium spp*. Tesis de Licenciatura, Universidad de Guanajuato.

23. Guo, J., Cheng, P., & Liu, Y. (2010). Functional significance of FRH in regulating the phosphorylation and stability of *Neurospora* circadian clock protein FRQ. Journal of Biological Chemistry, *285*(15), 11508-11515.

24. Guo, H., Wang, H., Keyhani, N. O., Xia, Y., & Peng, G. (2020). Disruption of an adenylate-forming reductase required for conidiation, increases virulence of the insect pathogenic fungus *Metarhizium acridum* by enhancing cuticle invasion. Pest management science, 76(2), 758-768.

25. Guy, L., Roat Kultima, J., & Andersson, S. G. (2010). genoPlotR: comparative gene and genome visualization in R. *Bioinformatics*, *26*(18), 2334-2335.

26. Hader, D. P. (2013). General photobiology. Elsevier.

27. Hernandez RR, Allen MF (2013) Diurnal patterns of productivity of arbuscular mycorrhizal fungi revealed with the soil ecosystem observatory. New Phytol 200(2):547–557

28. Herrera-Marcos, L. V., Lou-Bonafonte, J. M., Martinez-Gracia, M. V., Arnal, C., Navarro, M. A., & Osada, J. (2018). Prenylcysteine oxidase 1, a pro-oxidant enzyme of low density lipoproteins. Frontiers in bioscience (Landmark edition), 23, 1020–1037. https://doi.org/10.2741/4631

29. Hu, S., & Bidochka, M. J. (2020). DNA methyltransferase implicated in the recovery of conidiation, through successive plant passages, in phenotypically degenerated *Metarhizium*. Applied microbiology and biotechnology, 104(12), 5371-5383.

30. Kojima, M., Kimura, N., & Miura, R. (2015). Regulation of primary metabolic pathways in oyster mushroom mycelia induced by blue light stimulation: accumulation of shikimic acid. Scientific reports, 5, 8630.

31. Larrondo, L. F., & Canessa, P. (2018). The clock keeps on ticking: emerging roles for circadian regulation in the control of fungal physiology and pathogenesis. *Fungal Physiology and Immunopathogenesis*, 121-156.

32. Lauinger, L.; Diernfellner, A.; Falk, S.; Brunner, M. The RNA helicase FRH is an ATP-dependent regulator of CK1a in the circadian clock of *Neurospora crassa*. Nature Communications **2014**, *5*, 3598

33. Lee, S. J., Morse, D., & Hijri, M. (2019). Holobiont chronobiology: mycorrhiza may be a key to linking aboveground and underground rhythms. Mycorrhiza, 29(5), 403-412.

34. Lilly and Barnett, Lilly, VG; Barnett, (1951). Physiology of Fungi. McGraw Hill, New York, 1951, 463p. 33

35. Linden H, Ballario P, Macino G. (1997). Blue light regulation in *Neurospora crassa*. Fungal Genet Biology. Dec;22(3):141-50.

36. Linden H. (2002). A White Collar protein senses blue light. Science, 2002 Aug 2;297(5582):777-8.

37. Liu Y, He Q, Cheng P. (2003). Photoreception in *Neurospora*: a tale of two White Collar proteins. Cellular and Molecular Life Sciences. 2003 Oct;60(10):2131-8.

38. Liu, X., Chen, A., Caicedo-Casso, A., Cui, G., Du, M., He, Q., ... & Liu, Y. (2019). FRQ-CK1 interaction determines the period of circadian rhythms in *Neurospora*. Nature communications, 10(1), 1-13.

39. Liu, Y., & Bell-Pedersen, D. (2006). Circadian rhythms in *Neurospora crassa* and other filamentous fungi. Eukaryotic cell, 5(8), 1184-1193.

40. Lokhandwala, J., Hopkins, H. C., Rodriguez-Iglesias, A., Dattenböck, C., Schmoll, M., & Zoltowski, B. D. (2015). Structural biochemistry of a fungal LOV domain photoreceptor reveals an evolutionarily conserved pathway integrating light and oxidative stress. Structure, 23(1), 116-125.

41. Oliveira, A. S., Braga, G. U., & Rangel, D. E. (2018). *Metarhizium robertsii* illuminated during mycelial growth produces conidia with increased germination speed and virulence. Fungal biology, 122(6), 555-562.

42. Onofre, S. B., Miniuk, C. M., Barros, N. M. D., & Azevedo, J. L. (2001). Growth and sporulation of *Metarhizium flavoviride* var. *flavoviride* on culture media and lighting regimes. Scientia Agricola, 58(3), 613-616.

43. Peng, H., Guo, C. T., Tong, S. M., Ying, S. H., & Feng, M. G. (2021). Two white collar proteins protect fungal cells from solar UV damage by their interactions with two photolyases in *Metarhizium robertsii*. Environmental Microbiology.

44. Rangel, D. E. (2020). Outcome of blue, green, red, and white light on *Metarhizium robertsii* during mycelial growth on conidial stress tolerance and gene expression. Fungal biology, 124(5), 263-272.

45. Rangel, D. E., Fernandes, É. K., Braga, G. U., & Roberts, D. W. (2011). Visible light during mycelial growth and conidiation of *Metarhizium robertsii*

produces conidia with increased stress tolerance. FEMS microbiology letters, 315(2), 81-86.

46. Roberts, D. W., & St Leger, R. J. (2004). *Metarhizium* spp., cosmopolitan insect-pathogenic fungi: mycological aspects. Advances in applied microbiology, 54(1), 1-70.

47. Simon L, Bousquet J, Levesque CR, Lalonde M (1993) Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. Nature 363:67–69

48. Smith, K. M., Sancar, G., Dekhang, R., Sullivan, C. M., Li, S., Tag, A. G., ... & Freitag, M. (2010). Transcription factors in light and circadian clock signaling networks revealed by genomewide mapping of direct targets for *Neurospora* white collar complex. Eukaryotic cell, 9(10), 1549-1556.

49. Song, Z., Yang, J., Xin, C., Xing, X., Yuan, Q., Yin, Y., & Wang, Z. (2018). A transcription factor, MrMsn2, in the dimorphic fungus *Metarhizium rileyi* is essential for dimorphism transition, aggravated pigmentation, conidiation and microsclerotia formation. Microbial biotechnology, 11(6), 1157-1169.

50. Stone, L. B., & Bidochka, M. J. (2020). The multifunctional lifestyles of *Metarhizium*: evolution and applications. Applied Microbiology and Biotechnology, 1-11.

51. Tisch, D., & Schmoll, M. (2010). Light regulation of metabolic pathways in fungi. Applied microbiology and biotechnology, *85*(5), 1259-1277.

52. Tong, S. M., Wang, D. Y., Cai, Q., Ying, S. H., & Feng, M. G. (2020). Opposite nuclear dynamics of two FRH-dominated frequency proteins orchestrate non-rhythmic conidiation in *Beauveria bassiana*. *Cells*, *9*(3), 626.

53. Tong, S. M., Zhang, A. X., Guo, C. T., Ying, S. H., & Feng, M. G. (2018). Daylight length-dependent translocation of VIVID photoreceptor in cells and its essential role in conidiation and virulence of *Beauveria bassiana*. Environmental microbiology, 20(1), 169-185.

54. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, *25*(9), 1189-1191.

55. Yu, Z., & Fischer, R. (2019). Light sensing and responses in fungi. Nature Reviews Microbiology, 17(1), 25-36.

56. Zeng, G., Chen, X., Zhang, X., Zhang, Q., Xu, C., Mi, W., ... & Fang, W. (2017). Genome-wide identification of pathogenicity, conidiation and colony sectorization genes in *Metarhizium robertsii*. Environmental microbiology, 19(10), 3896-3908.