LysM receptors in *Coffea arabica*: identification, characterization, and gene expression in response to *Hemileia vastatrix*

Mariana de Lima Santos[1*], Mário Lúcio Vilela de Resende[2*¶], Bárbara Alves dos Santos Ciscon[2#a¶], Natália Chagas Freitas[2&], Matheus Henrique de Brito Pereira[2#a], Tharyn Reichel[2&], Sandra Marisa Mathioni [2#b]

[1] Programa de Pós-graduação em Biotecnologia Vegetal, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil,

[2] Departamento de Fitopatologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil,

[#a]Current Address: BASF Corporation, Brazil

[#b]Current Address: Syngenta Crop Protection, Uberlândia, Minas Gerais, Brazil

* Corresponding authors

E-mail: santos-ml@outlook.com (MLS)
E-mail: mlucio@dfp.ufla.br (MLVR)

[¶]These authors contributed equally to this work.
[&]These authors also contributed equally to this work.

Abstract

Pathogen-associated molecular patterns (PAMPs) are recognized by pattern recognition receptors (PRRs) localized on the host plant cell wall. These receptors activate a broad-spectrum and durable defense, which are desired characteristics for disease resistance in plant breeding programs. In this study, candidate sequences for PRRs with lysin motifs (LysM) were investigated in the *Coffea arabica* genome. For this, approaches based on the principle of sequence similarity, conservation of motifs and domains, phylogenetic analysis, and modulation of gene expression in response to *Hemileia vastatrix* were used. The candidate sequences for PRRs in *C. arabica* (*Ca1-LYP*, *Ca2-LYP*, *Ca1-CERK1*, *Ca2-CERK1*, *Ca-LYK4*, *Ca1-LYK5* and *Ca2-LYK5*) showed high similarity with the reference PRRs used: *Os-CEBiP*, *At-CERK1*, *At-LYK4* and *At-LYK5*. Moreover, the ectodomains of these sequences showed high identity or similarity with the reference sequences, indicating structural and functional conservation. The studied sequences are also phylogenetically related to the reference PRRs described in Arabidopsis, rice, and other plant species. All candidates for receptors had their expression induced after the inoculation with *H. vastatrix*, since the first time of sampling at 6 hours post-inoculation (hpi). At 24 hpi, there was a significant increase in expression, for most of the receptors evaluated, and at 48 hpi, a suppression. The results showed that the candidate sequences for PRRs in the *C. arabica* genome display high homology with fungal PRRs already described in the literature. Besides, they respond to pathogen inoculation and seem to be involved in the perception or signaling of fungal chitin, acting as receptors or coreceptors of this molecule. These findings represent an advance in the understanding of the basal immunity of this species.

Introduction

The interaction between plants and pathogens can be understood as a co-evolutionary "molecular war," in which each opponent uses their biological weapons as necessary, causing a successful infection by the pathogen or resistance in the host [1]. Currently, the study of pathogen perception by plants is divided into two lines. The first line is based on the recognition of conserved microbial molecules, denominated pathogen-associated molecular patterns (PAMPs), activating PAMP-triggered immunity (PTI). The second, on the other hand, recognizes the pathogen effectors by resistance proteins (R proteins), leading to effector-triggered immunity (ETI) [2,3].

The PAMPs recognition is performed by pattern recognition receptors (PRRs). These receptors are membrane proteins that usually have an extracellular domain involved in the perception of the ligand, the transmembrane or glycosylphosphatidylinositol (GPI) anchor domain that anchors the protein in the plasma membrane, and an intracellular kinase domain that is involved in the defense response signaling [4]. Adapted pathogens can suppress this first line of defense by secreting specific effectors. In response to this suppression, R proteins, encoded by resistance genes, recognize these effectors triggering ETI [5]. In spite of identifying different ligands, ETI and PTI lead to similar signaling pathways [6]. This signaling involves changes in calcium levels in the cytoplasm, production of reactive oxygen species (ROS) and signaling cascades involving protein kinases, MAPKs (mitogen-activated protein kinases) and CDPKs (calcium-dependent protein kinases) [7–10].

Comparing these two lines of defense, many studies indicate that the responses from the ETI occur more quickly and are more efficient than those from the PTI [11,12] since the former is associated with a hypersensitive response (HR), which involves programmed cell death and also systemic acquired response (SAR). For these reasons, the resistance conditioned by one or a few

resistance genes has been the focus of breeding programs for several cultivated species. Nonetheless, the PTI is effective against pathogens, insects and parasitic plants and constitutes an important factor in non-host resistance [13,14]. In addition, it leads to a durable and broad-spectrum resistance [15,16]. The ETI, on the other hand, being characterized as a resistance against specific pathogens is quickly overcome, due to the emergence of new races of the pathogen [17].

Due to these defense characteristics, broad spectrum and durable, in which the PRRs are involved, currently they have been the target of studies aiming at a greater use of these receptors in plant breeding [16,18]. These studies focus on the possibility of combining (pyramiding) PRRs and increasing resistance to a broad spectrum of pathogens. The best characterized PRRs are the leucine-rich repeat receptor kinases (LRR-RKs). These receptors are involved in the recognition of bacterial structures. An example of this is *FLS2* (Flagellin sensing 2), which detects a conserved epitope of 22 amino acids, flg22, existing in the N-terminal region of the flagellin protein [18,19] and *EFR* (EF-Tu receptor), which detects the elf18 epitope, corresponding to the 18 conserved residues in the N-terminal region of the elongation factor Tu (EF-Tu) [20]. For fungi, well-described receptors are those that recognize chitin and have in common extracellular domains with lysin residues (Lys) [4,21], such as *CERK1* (chitin elicitor receptor kinase 1) [22], *CEBiP* (chitin elicitor binding protein) [23], *LYK4, LYK5* (LysM-containing receptor-like kinase 4 and 5) [24,25], *LYP4* and *LYP6* (LysM domain-containing protein 4 and 6) [26].

Genetic alterations in the PRRs that recognize both fungal and bacterial PAMPs reduce the plant ability to properly perceive and defend against pathogens. Gene knockouts such as *Os-CERK1* [21,22] and mutations in *At-LYK5* [24] lead to a loss of ability to respond to chitin and initiate defense responses to adapted pathogens. In addition, it allows some degree of disease progression by non-adapted pathogens, displaying failures in non-host resistance [15]. These

studies demonstrate that the PTI and ETI form a continuum, which is necessary for a durable and efficient defense response [11]. Therefore, programs that seek to enable resistance to phytopathogens, with a focus on increasing the capacity of the recognition system, are successful by adding the PTI and ETI as the main strategy for obtaining resistant cultivars [15,27].

Few non-model plants, such as barley [28], apple [29,30] and mulberry [31], had PRRs characterized. *Coffea arabica* is an important coffee species cultivated in countries such as Brazil, Vietnam, Colombia, and Indonesia [32]. PAMP receptors have been scarcely studied in *Coffea spp*., therefore, it is crucial to identify the receptors that are present in their genome, and whether there is a response induced by the inoculation of pathogens, thus allowing the use of PRRs in coffee breeding programs.

The rust is the main coffee disease, causing severe losses in productivity in all regions where coffee is cultivated [33,34]. In Brazil, the biotrophic fungus *Hemileia vastatrix* Berk. & Br, the etiological agent of coffee rust, has caused damage since the 1970s [35,36]. In regions with favorable conditions for the pathogen, the decline in productivity can reach 50% [36]. To circumvent such damage, chemical control has been used, however, the use of tolerant or resistant cultivars is a viable alternative to reduce costs and possible environmental damage [33,37,38]. Therefore, the goals of this study were (i) to identify the pattern recognition receptors (PRRs) for fungi in the *C. arabica* genome, (ii) to characterize these sequences for protein domains and motifs and (iii) to analyze the gene expression of these PRRs in cultivars of *C. arabica* contrasting to rust resistance inoculated with *H. vastatrix*. The data obtained suggested that *C. arabica* has LysM receptors that act as fungal PAMP receptors, and that the expression of these receptors is stimulated after *H. vastatrix* inoculation. Our results contribute to the understanding and future employment of PRRs in coffee breeding programs.

Materials and Methods

Identification and characterization of specific PRRs for fungi in the *C. arabica* genome

The reference PRRs described in the literature for fungal PAMPs recognition in *Arabidopsis thaliana* and in *Oryza sativa* were selected: *At-CERK1*, *At-LYK4*, *At-LYK5* and *Os-CEBiP* (Table 1). To identify these receptors, the *C. arabica* genome (accession UCG-17, variety Geisha) sequenced by the University of California (UC Davis Coffee Genome Project) and partially available in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) was used. The search was based on sequence similarity and domain conservation. For this, a BLASTp (Align Sequences Protein BLAST) with default parameters was performed in Phytozome. The *C. arabica* sequences returned by BLASTp were selected based on the following criteria: e-value $\leq 10^{-5}$, extracellular domain corresponding to the reference sequence used (Lysin motifs -LysM), and transmembrane or GPI anchor domain. The domains were analyzed using the SMART (http://smart.embl-heidelberg.de/), the TMHMM2.0 (http://www.cbs.dtu.dk/services/TMHMM/) and the PredGPI (http://gpcr.biocomp.unibo.it/predgpi/pred.htm).

Table 1. Reference PRRs and homologues.

Name	Type	ID*	Botanical species	PAMP	References
*OsCEBiP**	RLP	XP_015630176.1	*Oryza sativa*	chitin	Kaku et al. (2006)
AtLYP1 (LYM2)	RPL	AT2G17120.1	*Arabidopsis thaliana*	chitin	Shinya et al. (2012)
MtLYM2	RLP	-	*Medicago truncatula*	chitin	Fliegmann et al. (2011)
MmLYP1	RLP	AXQ60477.1	*Morus multicaulis*	chitin	Lv et al. (2018)
HvCEBiP	RLP	BAJ92081.1	*Hordeum vulgare*	chitin	Tanaka et al. (2010)
AtLYP2 (LYM1)	RPL	AT1G21880.2	*Arabidopsis thaliana*	PGN	Willmann et al. (2011)
AtLYP3 (LYM3)	RPL	AT1G77630.1	*Arabidopsis thaliana*	PGN	Willmann et al. (2011)
OSLYP4	RPL	XP_015610852.1	*Oryza sativa*	chitin/ PGN	Liu et al. (2012)
OsLYP6	RPL	XP_015641500.1	*Oryza sativa*	chitin/ PGN	Liu et al. (2012)

*AtCERK1**	RLK	AT3G21630.1	*Arabidopsis thaliana*	chitin	Miya et al. (2007)
OsCERK1	RLK	BAJ09794.1	*Oryza sativa*	chitin	Shimizu et al. (2010)
SILYK1(Bti9)	RLK	Solyc07g049180	*Solanum lycopersicum*	-	Zeng et al. (2012)
VvLYK1-1	RLK	XP_010657225.1	*Vitis vinifera*	chitin	Brulé et al. (2019)
VvLYK1-2	RLK	XP_010655366.1	*Vitis vinifera*	chitin	Brulé et al. (2019)
MdCERK1	RLK	ATD50586.1	*Malus domestica*	chitin	Zhou et al. (2018)
MdCERK1-2	RLK	MD17G1102100	*Malus. domestica*	chitin	Chen et al. (2020)
MmLYK2	RLK	AXQ60478.1	*Morus multicaulis*	chitin	Lv et al. (2018)
PsLYK9	RLK	-	*Pisum sativum*	chitin	Leppyanen et. (2018)
*AtLYK4**	RLK	AT2G23770.1	*Arabidopsis thaliana*	chitin	Wan et al. (2012)
VvLYK4-1	RLK	XP_002269408.1	*Vitis vinifera*	chitin	Brulé et al. (2019)
VvLYK4-2	RLK	XP_010649202.1	*Vitis vinifera*	chitin	Brulé et al. (2019)
BdLYK4	RLK	Bradi3g06770.1	*Brachypodium distachyon*	chitin	Tombuloglu et al. (2019)
*AtLYK5**	RLK	AT2G33580.1	*Arabidopsis thaliana*	chitin	Cao et al. (2014)
VvLYK5-1	RLK	XP_002277331.3	*Vitis vinifera*	chitin	Brulé et al (2019)

RLP: Receptor like protein, RLK: Receptor like kinase, PGN: Peptidoglycan. *Reference sequences.

After selecting the sequences of *C. arabica*, they were again compared to the reference sequences by phylogenetic analysis. This analysis enabled to identify which peptide sequences had the greatest phylogenetic similarity to the reference PRRs, thus allowing the selection of candidate sequences. Additionally, considering that these PRRs present protein domains very close, a joint phylogenetic tree, with the candidate sequences in *C. arabica*, the reference PRRs and homologs (Table 1), was also created to confirm the separation of these groups and the homology of these sequences. The databases used to retrieve the reference sequences were: the GenBank from the National Center for Biotechnology Information (NCBI) sequence database, the Arabidopsis Information Resource (TAIR), the Sol Genomics Network, the Apple Genome and Epigenome, and Phytozome. The complete amino acid sequences were aligned by the CLC Genomics Workbench software version 11.0.1 (QIAGEN) (default parameters with very accurate) and the

phylogenetic tree was generated by the Mega software version 10.1.8 [39] using the Maximum Likelihood method with a bootstrap of 1000 replications.

To characterize the extracellular regions of the candidate sequences, the lysin motifs (LysM) were used for multiple alignments between the candidate and reference sequences. The LysM motifs of each sequence were predicted by SMART using the extracellular region and aligned by the MAFFT program online version (https://mafft.cbrc.jp/alignment/server/) [40]. After the alignment, the visualization and calculation of the identity and similarity of each of the candidate sequences against the reference sequences were obtained by BioEdit version 7.2.5 [41].

Considering the fact that *C. arabica* is an allotetraploid (2n = 4x = 44 chromosomes), the result of the cross between *C. canephora* and *C. eugenioides* [42,43], the sequences selected as PRR candidates for the arabica coffee were also analyzed by BLASTp in the database of the NCBI (https://www.ncbi.nlm.nih.gov/) against each ancestral subgenome. This analysis aimed to verify the possible genomic origin of the studied PRRs.

Primer design

The *C. arabica* sequences selected as candidates by the phylogenetic analysis were used for primer design. The primers were designed using the Primer Quest software and their quality was analyzed using the Oligo Analyzer software, both available online by IDT (Integrated DNA Technologies, USA). After the primers were designed, they were blasted (BLASTn - Standard Nucleotide BLAST) against the NCBI and Phytozome database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to attest their specificity through the identification of non-complementarity with nonspecific sequences.

Fungal inoculum preparation

The inoculum used was obtained from leaves of *C. arabica* naturally infected with *H. vastatrix*. The pustules of these leaves were scraped and placed in microtubes, were frozen in liquid nitrogen, and stored in a freezer at -80ºC. To prepare the inoculum, the stored spores were submitted a 40ºC thermal shock for 10 min and were tested for viability by the spore germination test. The viability was verified by observing the spore germination in glass cavity slides. After preparing the suspension (1 x 10^6 urediniospores/ml) for plant inoculation, three drops were transferred to glass cavity slides, which were incubated at 25°C for 48 hours. After the incubation, the spores were visualized under an optical microscope, so their germination could be observed (S1 Fig).

Plant materials, experimental design, and inoculation

Aiming to analyze the gene expression of the PRR selected candidates, seedlings of four cultivars of *C. arabica* were used, being two rust susceptible cultivars, Catuaí Vermelho IAC 144 (CV) and Mundo Novo IAC 367-4 (MN), and two rust resistant, Aranãs RV (AR) and Iapar-59 (IP). The experiment was conducted in a randomized complete block design (RCBD) with three replicates and an experimental plot consisting of three plants. The treatments were arranged in a 2 x 3 x 4 factorial scheme, the factors being: condition (inoculated and not inoculated); evaluation times (06, 24 and 48 hours post-inoculation - hpi) and cultivars (Catuaí Vermelho IAC 144, Mundo Novo, Aranãs RV, and Iapar-59). The experiment was repeated twice independently.

Young plants (3-4 pairs of leaves) were inoculated in a growth chamber with a controlled environment (temperature of 22 ± 2°C, relative humidity of 90%) favoring the disease development. The suspension was sprayed on abaxial leaf surfaces and the inoculated plants were kept in the dark in a humid chamber according to a previously published methodology [44]. The control plants (sprayed with pure water only) were also sampled at all the evaluated time points.

All the leaves collected were immediately frozen in liquid nitrogen and subsequently stored in a freezer at -80°C. After the treatment and sampling, the plants were kept in a greenhouse until the first symptoms and signs of the pathogen were seen to make sure the inoculation was effective (S2 Fig).

RNA extraction and quantification

Following the RNA extraction, the samples were ground with liquid nitrogen until a fine powder was obtained. The ground material was stored in a ultrafreezer at -80°C until the RNA extraction was performed. The extraction was performed using the Plant RNA Purification Reagent (Thermo Fisher). Subsequently, the RNA was treated with DNase (RQ1 RNase-Free DNase, Promega) to remove any residual DNA in the sample. These procedures were performed according to manufacturer's instructions. The integrity of the RNA was verified on 1% agarose gel and quantified on the NanoDrop One spectrophotometer (Thermo Fisher). All samples used showed a ratio reading 1.8-2.0 of absorbance at 260/280 nm and 260/230 nm for high-quality RNA.

cDNA synthesis and RT-qPCR

An aliquot containing 1 µg of total RNA (treated with DNase) was used for cDNA synthesis using the High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor (Thermo Fisher). After the synthesis, the cDNA was diluted 5x and stored at -20 °C. The RT-qPCR were performed in the QuantStudio® 3 Real-Time PCR System (Applied Biosystems) using the SYBR® Green detection system. The amplification conditions were: 50°C for 2 min and 95°C for 10 min, 40 cycles: 95°C for 15 s, 60°C for 1 min and a final step of 95°C for 15 s (melting curve). The final reaction volume was 10µL contained the following components: 1.0 µL of cDNA (~ 10 ng), 0.4

209 µL of each primer (forward and reverse) at a concentration of 10 µM (400 nM in the reaction),
210 except for the *Ca2-CERK1* (Scaffold 2193.164 and 476.38), which used 0.2 µL (200 nM in the
211 reaction), 5.0 µL of Platinum SYBR Green qPCR SuperMix-UDG with ROX (Thermo Fisher),
212 and 3.4 µL of ultrapure water (free of nucleases).

213 For each of the three biological samples, technical triplicates were used and for each plate
214 an inter-assay sample was used to ensure the reproducibility of the technique. The relative
215 quantification was calculated according to the formula by Pfaffl, 2001 [45]. Referring to the data
216 normalization, the expression stability of four reference genes was analyzed: protein 14-3-3 (*14-*
217 *3-3*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ribosomal protein 24S (*24S*) and
218 factor elongation 1α (*EF1-α*) [46–49]. The efficiency correction of these genes in Cq values was
219 performed by the GenEx Enterprise program (version 7.0) and the stability was verified by the
220 RefFinder tool [50]. The two most stable genes were *14-3-3* and *GAPDH* (S3 Fig), which were
221 used to normalize the transcription levels of the target genes. The samples with the lowest
222 expression were used as calibrators. The MN 48 hpi was used as calibrator sample, except for the
223 *Ca1-CERK1* (experiment 2), which was used the IP 48 hpi sample. The PCR amplification
224 efficiencies and linear regression coefficients were determined using the LinRegPCR software
225 version 2018.0 (Table 2) [51]. The average expression was obtained by the ratio of the sample
226 inoculated with *H. vastatrix* compared to the average of the control treatment (without
227 inoculation).

228

229 **Table 2. Sequence of primers used for candidate sequences of *C. arabica* PRRs and reference**
230 **genes.**

Gene	Target sequence (Scaffold)	Primer	Amplification efficiency	R^2

LYK4	612.376 and 952.320-1	AAAGGCCACAAACAGATGCGACAG (F)	Exp1 - 1,855	0,929
	(*Ca-LYK4*)	AGGTGGGATGGATCAGCTGCTAAG (R)	Exp2 - 1,866	0,961
LYK5	628.522	TTTGGTTCCTGCGGTATAGG (F)	Exp1 - 2,056	0,974
	(*Ca1-LYK5*)	TCTGGCAAAGCCCTGTAAAC (R)	Exp2 - 2,095	0,988
	1841.91	TTGCAGCATGCCACAGGTTCTTTC (F)	Exp1 - 1,920	0,961
	(*Ca2-LYK5*)	ATCACTCAGGCCACCTTTCTCTGC (R)	Exp2 - 1,898	0,952
CERK1	1805.113 and 539.592 **(*Ca1-CERK*)**	CGAGACATTAAGCCAGCTAAC (F)	Exp1 - 1,881	0,990
		GCATGTAACCGAAAGTACCC (R)	Exp2 - 1,887	0,965
	2193.164 and 476.38	CAGTTCCAGTTAGCTGCTCCA (F)	Exp1 - 1,899	0,999
	(*Ca2-CERK*)	GGAGAAGTTCCTTCAGCAACAC (R)	Exp2 - 1,885	0,992
LYP **(*CEBiP*-like)**	439.212	ACCACCGCCGATGTTCTGTTGC (F)	Exp1 - 1,898	0,992
	(*Ca1-LYP*)	GAGGAACATCGAGAATAGCGCCGG (R)	Exp2 - 1,887	0,994
	1196.90	TCCAGACCCTCTTCAACGTC (F)	Exp1 - 1,824	0,983
	(*Ca2-LYP*)	CAGGCGAAAGGAATCTTGAG (R)	Exp2 - 1,829	0,997
14-3-3	SGN-U347734	TGTGCTCTTTAGCTTCCAAACG (F)	Exp1 - 1,983	0,943
		CTTCACGAGACATATTGTCTTACTCAAA (R)	Exp2 - 2,001	0,933
GAPDH	SGN-U356404	TTGAAGGGCGGTGCAAA (F)	Exp1- 2,007	0,993
		AACATGGGTGCATCCTTGCT (R)	Exp2 - 2,060	0,995
24S	GR986263.1	ACGGCATCAAAGGAGACAAT (F)	Exp1 - 1,893	0,998
		ATGCAGAACATCGATCACGA(R)	Exp2 - 1,902	0,994
EF1-α	GW466696.1	CTCTCTCGCCTCCTGTCTTC (F)	Exp1 - 1,912	0,983
		CAGAGTCGACGTGACCAATG (R)	Exp2 - 1,932	0,972

The candidate sequences and reference genes (Target sequence) were obtained from Phytozome and SOL Genomics Network. The primer sequences for the reference genes *14-3-3 and GAPDH* were obtained from Barsalobres-Cavallari et al. 2009 [46] and *24S* and *EF1-α* from Reichel 2021 [49]. Exp1: experiment 1, Exp2: experiment 2.

Statistical analysis

The relative expression data of the two experiments were subjected to analysis of variance, using the following model:

$$y = \mu + R/E_{b(k)} + E_k + C_i + T_w + (EC)_{ki} + (ET)_{kw} + (CT)_{iw} + (ECT)_{kiw} + e_{kiw}$$

in which $R/E_{b(k)}$ is the effect of block b within experiment k; E_k is the effect of experiment k, C_i is the effect of cultivar *i*, T_w is the effect of time w, $(EC)_{ki}$ is the effect of the interaction between experiment k and cultivar *i*, $(ET)_{kw}$ is the effect of the interaction between experiment *k* and time w; $(CT)_{iw}$ is the effect of the interaction between cultivar *i* and time w; $(ECT)_{kiw}$ it is the effect of the interaction between experiment k cultivar *i* and time w; e_{kiw} is the effect of the experimental error, $\cap$ N(0, σ_e^2). Checks for outliers and of the assumptions of residuals from models were accomplished using diagnostic plots within the R software [52].

The interaction between cultivar and time was decomposed and the means between the levels of the factors were analyzed by Tukey's test at 5% of probability. Data analysis was performed using the R software [52].

Results

Identification and characterization of specific fungal PRR in the *C. arabica* genome

The BLASTp analysis in Phytozome with the reference PRRs resulted in 4, 10, 12 and 14 sequences in the *C. arabica* genome for *Os-CEBiP*, *At-LYK5*, *At-CERK1* and *At-LYK4*, respectively (Fig 1 and S1 Table). These sequences were selected because they have e-value $\leq 10^{-5,}$ extracellular region containing lysin motif (LysM) and transmembrane domain or GPI-anchor. After the phylogenetic analysis, two candidate sequences were selected for *LYK4* (Scaffold 612.376 and 952.320) and *LYK5* (Scaffold 628.522 and 1841.91) (Fig 1B and 1D and S1Table) and four ones for *CERK1* (Scaffold 539.592, 1805.113, 2193.164 and 476.38) (Fig 1A and S1 Table). As the phylogenetic analysis for candidate sequences to the *CEBiP* protein did not result in a significant bootstrap (Fig 1C), other proteins belonging to the LYP clade (*CEBiP-like*)

261 described in Arabidopsis and rice were included in a new analysis: *At-LYP1* (*At-CEBiP / LYM2*),
262 *At-LYP2* (*LYM1*), *At-LYP3* (*LYM3*), *Os-LYP4* and *Os-LYP6* (Table 1).
263
264 **Fig 1. Phylogenetic analysis of the selected sequences for *C. arabica* by comparison with the**
265 **reference PRRs.** (A) *CERK1*, (B) *LYK4*, (C) *CEBiP*, (D) *LYK5*, (E) *CEBiP* and reference proteins
266 belonging to the *LYP* (*CEBiP*-like) group. The phylogenetic trees were constructed with complete
267 amino acid sequence alignments using the Maximum Likelihood method with a bootstrap of 1000
268 replications. The cluster clade of candidate sequences for *C. arabica* and reference sequences are
269 highlighted in blue.
270
271 The new phylogenetic analysis for *CEBiP* (Fig 1E) showed two distinct clades. The clade
272 one formed by the sequences Scaffold 506.17 and 1856.2, *At-LYP*2, *At-LYP3*, *Os-LYP4* and *Os-*
273 *LYP6*, and the clade two formed by *Os-CEBiP*, *At-LYP1*, Scaffold 439.212 and 1196.90. As the
274 Scaffold sequences 439.212 and 1196.90 showed greater similarity with the *Os-CEBiP* homologue
275 in *A. thaliana* (*At-LYP1),* they were selected as candidate sequences for the *CEBiP-like* (Fig 1 and
276 S1 Table). Moreover, the *At-LYP2* (*LYM1*) and *At-LYP3* (*LYM3*), belonging to clade one, are
277 described in the literature for their ability to recognize the peptidoglycan, a bacterial PAMP [53].
278 These sequences formed the nearest clade to the Scaffold 506.17 and 1856.2 sequences,
279 substantiating the choice of the two *C. arabica* sequences belonging to clade two. The *Os-LYP4*
280 and *Os-LYP6* that play a dual role, recognizing peptidoglycan and chitin [26], were not evaluated
281 in this study.
282 All the domains found in the coffee candidate sequences correspond to the characteristic
283 domains of the reference sequences. The description of these sequences such as identity and

similarity in relation to the reference sequences as well as the gene size, the CDS and the number of exons, are shown in Table 3. The candidate sequences for *CERK1*, *LYK4* and *LYK5* have an extracellular LysM domain (with three LysM), a transmembrane domain, and an intracellular Ser/Thr kinase domain. The sequences selected as *CEBiP-like* have two lysin motifs and a predicted GPI-anchor. The characterization of these domains, motifs and protein sizes are shown in Fig 2.

Table 3. BLASTp and nucleotide characterization of candidate sequences in *C. arabica*

Candidate sequence	Identity (%)	Similarity (%)	Gene (pb)	Exons	CDS (pb)
CERK1-Scaffold_539.592	56.109	70.3	6082	13	2511
CERK1-Scaffold_1805.113	55.145	69.0	4186	10	1815
CERK1-Scaffold_2193.164	57.546	73.0	10180	12	1860
CERK1-Scaffold_476.38	57.261	73.1	9921	12	1860
LYK4-Scaffold_612.376	46.154	64.1	1935	1	1935
LYK4-Scaffold_952.320	46.154	64.4	1935	1	1935
LYK5-Scaffold_628.522	58.036	76.5	2031	1	2031
LYK5-Saffold_1841.91	58.631	76.5	2031	1	2031
LYP-Scaffold_1196.90	42.258	56.1	2961	4	1098
LYP-Scaffold_439.212	35.385	49.6	3598	5	954

Percentage of identity and similarity refer to BASTp analysis of candidate sequences against reference sequences *At-CERK1*, *At-LYK4*, *At-LYK5* e *LYP* (*CEBiP*- like). Candidate sequences were obtained from Phytozome database.

Fig 2. Protein characterization of the candidate sequences for *CERK1*, *LYK4*, *LYK5,* and *CEBiP*-like in *C. arabica*. The signal peptide positions, lysin motifs (LysM) and transmembrane domains were identified by SMART, and the GPI anchor by PredGPI. The domains positions are represented by numbers at the beginning and end of each domain. Concerning the *CEBiP*-like candidate sequences, the putative signal sequences for the GPI anchor and their specificities are

shown. The numbers at the beginning of each sequence represents the scaffold (candidate sequence in *C. arabica*). The numbers at the end of each sequence represents the size of the proteins in number of amino acids. SP: signal peptide, LysM: lysin motifs identified as 1,2 e 3, TM: transmembrane domain, GPI: GPI-anchor.

The extracellular lysin motif regions (LysM1, LysM2 and LysM3) for these sequences ranged from 38 to 49 aa. The multiple alignments of these regions with the reference proteins showed high residue conservation but varied among the studied receptors (Fig 3). Out of eleven residues described as important for the chitin oligomer binding function in *At-CERK1* [54,55], eight ones displayed identity or similarity with the candidate sequences in *C. arabica*. For *Os-CEBiP*, from nine described [56], only three were present. In *At-LYK5*, only one of three described [24] showed similarity with *C. arabica* sequences. The tyrosine (Tyr) residue, located at position 128 in *At-LYK5*, considered as the fourth chitin-binding residue for this receptor, was not analyzed, as it is present between the LysM1 and LysM2 motifs, a region that was not analyzed in the alignment.

Fig 3. Alignment of the LysM motifs between reference sequences and candidate sequences in *C. arabica*. The LysM motif sequences were aligned using MAFFT and visualized by BioEdit. The numbers at the beginning of each sequence represents the scaffold (candidate sequence in *C. arabica*). The green line highlights the reference sequence. The purple and gray shading represent identical and similar amino acids, respectively. The percentages of identity and similarity between candidate sequences and references are indicated by * and **, respectively. In red are the critical residues that bind to chitin and the green arrows indicate residues identical or similar to these

regions present in the candidate sequences in *C. arabica*. The numbers at the end of each sequence represent the size of the LysM motifs in number of amino acids.

Joint phylogenetic analysis and BLASTp against the genomes of *C. canephora* and *C. eugenioides*

A joint phylogenetic tree was created to verify whether the candidate sequences would form distinct clades, including the reference sequences used. This tree was composed of the selected candidate sequences for PRRs in *C. arabica*, the reference sequences used to search for these PRRs in coffee (*At-CERK1*, *At-LYK4*, *At-LYK5* and *Os-CEBiP*) and homologs of these proteins described experimentally in the literature (Table 1). This analysis formed four clades that separated the candidate sequences in coffee with the respective reference proteins used, confirming their phylogenetic relationships (Fig 4).

Fig 4. Joint phylogenetic analysis of candidate sequences in *C. arabica*, reference sequences and homologs described experimentally. The phylogenetic tree was constructed with alignments of complete amino acid sequences using the Maximum Likelihood method with a bootstrap of 1000 repetition. The *CERK1*, *LYK5, LYK*4 and *CEBiP*-Like clades are highlighted in different colors: I- purple, II- red, III- green and IV- blue.

The clade I was composed of Scaffold 539.592, 1805.113, 2193.164 and 476.38, *At-CERK1* and their homologs *Md-CERK1*, *Md-CERK1-2*, *Mm-LYK2* (*CERK1*-like), *Ps-LYK9*, *SI-LYK1* (*Bti9*), *Vv-LYK1-1*, *Vv-LYK1-2* and *Os-CERK1*. In this clade, the candidate sequences in coffee, Scaffolds 476.38 and 2193.164 are closest to the homologs of *At-CERK1* in tomato, *SI-*

346 *LYK1* (*Bti9*), while the Scaffold 539.592 and 1805.113 sequences, formed a more distant subclade.
347 Clades II and III belonging to *LYK4* and *LYK5* formed closer clades. The coffee sequences were
348 grouped more closely to the *LYK4* homologues in grape and for the *LYK5* they formed a subclade
349 with the reference sequence *At-LYK5* and its homolog also in grape (*Vv-LYK5-1*). In clade IV,
350 belonging to the *CEBiP* cluster, it was observed that candidate sequences in coffee were
351 significantly grouped with the *Os-CEBiP* homologs.

352 The BLASTp analysis in the NCBI database against the genomes of *C. canephora* and *C.*
353 *eugenioides* showed that seven candidate sequences for PRRs in coffee have the highest percentage
354 of identity with sequences belonging to the genome of *C. eugenioides* and only two (*LYK4*-
355 Scaffold 612. 376 and *LYP*- Scaffold 1196.90) showing greater identity with the *C. canephora*
356 sequences (Table 4).

357

358 **Table 4. BLASTp analysis of candidate sequences in *C. arabica* against the genomes of *C.**
359 ***eugenioides* and *C. canephora*.**

Candidate sequence *C. arabica*	*C. eugenioides*			*C. canephora*	
	Query Cover	Identity (%)	ID*	Identity (%)	ID*
CERK1-Scaffold_539.592	98%	97.09	XP_027185465.1	96.53	CDP04438.1
CERK1-Scaffold_1805.113	99%	96.76	XP_027185465.1	93.03	CDP04438.1
CERK1-Scaffold_2193.164	100%	97.09	XP_027172569.1	96.77	CDP11343.1
CERK1-Scaffold_476.38	100%	99.84	XP_027172569.1	94.67	CDP11343.1
LYK4-Scaffold_612.376	100%	97.67	XP_027154003.1	99.69	CDP02416.1
LYK4-Scaffold_952.320	100%	99.38	XP_027154003.1	97.20	CDP02416.1
LYK5-Scaffold 628.522	100%	98.82	XP_027147944.1	98.52	CDP11912.1
LYK5-Saffold 1841.91	100%	97.78	XP_027147944.1	96.15	CDP11912.1
LYP-Scaffold_1196.90	100%	96.99	XP_027185286.1	98.90	CDP14511.1
LYP-Scaffold_439.212	94%	78.80	XP_027185286.1	77.65	CDP14511.1

360 *GenBank National Center for Biotechnology Information (NCBI) sequence database. All e-
361 values were 0.0, except for XP_027185286.1 (3e-176) and CDP14511.1 (1e-173).
362

363 # Primer design

364 The four sequences selected as candidates for *CERK1* in the *C. arabica* genome by
365 phylogenetic analysis formed two distinct subclades (Fig 1A). The subclade I formed by the
366 Scaffold 539.592 and Scaffold 1805.113 sequences and the subclade II formed by the Scaffold
367 2193.164 and Scaffold 476.38 sequences. The coding sequences (CDS) of subclade I showed an
368 71.33% identity, with the 1805.113 sequence presenting a smaller CDS (1815bp) and shared
369 almost entirely with the Scaffold 539.592 sequence. The Scaffold 539.592 sequence, on the other
370 hand, presents a larger CDS (2511 bp) with two regions that are not present in 1805.113 (S4 Fig).
371 The Scaffold 2193.164 and 476.38 showed CDS of the same size (1860bp) and an identity 98.28%
372 (S5 Fig). For the primer design in the gene expression analysis, the formation of these two
373 subclades was considered, thus using a pair of primers for each of the formed subclades. They
374 were named *Ca1-CERK1* and *Ca2-CERK1* respectively and are referred to as such in the gene
375 expression analysis (Table 2).
376 Concerning the *LYK4* candidate sequences (Scaffold 612.376 and 952.320), a primer pair
377 was also designed for both candidate sequences. These showed a 98.45% identity (S6 Fig) and
378 were named as *Ca-LYK4*. Regarding the candidate sequences *LYK5* (Scaffold 628.522 and
379 Scaffold 1841.91) and *LYP* (*CEBiP*-Like) (Scaffold 439,212 and 1196.90), a primer pair was
380 designed for each sequence separately and they are referred to as *Ca1-LYK5*, *Ca2-LYK5*, *Ca1-*
381 *LYP*, *Ca2-LYP*, respectively (Table 2).

Transcriptional response of candidate receptors in *C. arabica*

To verify the transcriptional responses of the candidate sequences to the PRRs in *C. arabica,* four cultivars with contrasting rust resistance levels were inoculated with *H. vastatrix*. The inoculum used displayed viability in both tests: the one with the glass cavity slides (S1 Fig) and the other about the ability to cause the disease symptoms and signs in susceptible cultivars CV and MN (S2 Fig). The resistant cultivars AR and IP, presented no symptoms or signs of the disease. The fungal inoculation induced the expression of all candidate receptors in all cultivars and studied time points. To a greater or lesser degree, there was an increase in expression from 6 hpi (Fig 5), with the peak varying between 6 and 24 hpi, followed by a decrease at 48 hpi.

Fig 5. Relative expression of candidate genes for *CERK1*, *LYP* (CEBiP-like), *LYK5* and *LYK4* in *C. arabica*. (A) *Ca1-CERK1*, (B) *Ca2-CERK1*, (C) *Ca1-LYP*, (D) *Ca2-LYP*, (E) *Ca1-LYK5*, (F) *Ca2-LYK5*, (G) *Ca-LYK4*. Candidate genes were evaluated in *C. arabica* leaves at 6, 24 and 48 hours post-inoculation (hpi) with *H. vastatrix*. The average of relative expression was obtained by the ratio between the means of inoculated and control (not inoculated). Capital letters represent the statistical analysis of the times for each cultivar and lower letters between cultivars. Means followed by the same letter are not differentiated by Tukey's test at 5% probability. The data shown represents experiments 1 and 2. MN: Mundo Novo, CV: Catuaí Vermelho IAC 144, AR: Aranãs RV, IP: IAPAR-59.

The two groups of candidate sequences for *CERK1* showed different expression profiles (Fig 5A and 5B) at 24 hpi. The *Ca1-CERK1* had higher expression than *Ca2-CERK1*. Concerning the former, the expression rate was seven times higher than that of the control in cultivar MN,

regarding the latter, the highest value did not reach twice as much for IP. When the time expression levels were analyzed for each cultivar in the two groups (Fig 5A and 5B), there was a significant difference for 24 hpi, except for CV *Ca2-CERK1*. Regarding the *Ca1-CERK1*, the analysis between cultivars (Fig 5A) showed that IP and MN displayed approximately 6- and 7-fold higher expression levels at 24 hpi, respectively, demonstrating significant differences compared to AR and CV. Respecting 6 and 48 hpi, there were no significant differences. Concerning *Ca2-CERK1* (Fig 5B), the analysis between cultivars showed that at 6 hpi it was the most expressed in CV and MN. At 24 hpi, the highest expression was in IP, and at 48 hpi the same cultivar showed a reduction in its expression, which was the least expressed among the cultivars.

A similar profile to *CERK1* was observed for the sequences studied as candidates for *LYP* and *LYK5* (Fig 5 C - 5 F). The *Ca1-LYP* and *Ca2-LYK5* obtained cultivars with higher expression levels at 24 hpi than *Ca1-LYK5* and *Ca2-LYP*, however, for these genes, the candidate sequences were studied apart. Considering *Ca1-LYP* and *Ca2-LYP* (Fig 5C and 5D), the expression patterns were different at 6 and 24 hpi. The *Ca1-LYP* expression levels did not reach twice as much compared to the control at 6 hpi, while for *Ca2-LYP* the highest averages were observed at that time. Moreover, regarding the *Ca1-LYP*, all cultivars showed an expression above twofold higher at 24 hpi. Therefore, the greatest inductions for *Ca2-LYP* occurred at 6 hpi while for *Ca1-LYP* they happened later at 24 hpi.

The expression differences in time for each cultivar considering *Ca1-LYP* (Fig 5C) showed that AR and IP have significant differences at 24 hpi, which did not occur in CV and MN. The analysis between cultivars showed that at 6 hpi and 48 hpi there were no differences, but that at 24 hpi, IP was the cultivar that showed the highest expression, reaching 6-fold higher. Considering *Ca2-LYP* (Fig 5 D), AR and CV showed higher expressions at 6 hpi. For IP and MN, the largest

428 expression occurred at 6 and 24 hpi, with no difference between these times. The analysis between
429 cultivars showed that at 6 hpi, AR obtained the highest expression while IP presented the lowest
430 expression. On the other hand, at 24 and 48 hpi, there were no differences between cultivars.
431 However, it was found that 48 hpi was the time with the lowest average observed, within and
432 between cultivars.

433 Regarding *Ca1-LYK5* (Fig 5E), there was a difference between the times for all cultivars,
434 except for AR. The MN cultivar had the highest average at 6 hpi, while IP obtained the highest at
435 24 hpi. For the cultivar CV, there were no differences between these times, only at 48 hpi.
436 Concerning the analysis between cultivars, the MN obtained the highest average at 6 hpi and IP at
437 24 hpi. At 48 hpi, there were no differences between cultivars and this time presented the lowest
438 average for all. Referring to *Ca2-LYK5* (Fig 5F), all cultivars showed differences between the
439 evaluated times, except for CV. The AR and IP cultivars showed significant differences in averages
440 at 24 hpi compared to the ones at 6 and 48 hpi, coming to express about six and eight times more
441 than the control, respectively. Regarding MN, the highest average was also detected at 24 hpi, but
442 this did not differ statistically from 6 hpi, only from 48 hpi. For the times between cultivars, there
443 were differences only in 24 hpi, with AR and IP having the highest expression.

444 The values for *Ca-LYK4* were the result of a single primer pair designed for two candidate
445 sequences. In this receptor, the expression levels at 24 hpi differed within and between the cultivars
446 evaluated. The IP cultivar obtained the highest average expression, reaching almost 19 times
447 higher than that of the control, followed by MN, which expressed ninefold higher. The lowest
448 averages for that time were observed for CV and AR, with an expression seven- and sixfold higher,
449 respectively. For 6 and 48 hpi there was no difference within and between cultivars, the averages
450 for those times reached at most twice as much.

Discussion

Fungal PRRs in the *C. arabica* genome

Understanding basal immunity has been the focus of several studies with the purpose of identifying the mechanisms governing this line of defense, enabling its use as another tool in the search for plant resistance to pathogens [18]. The description of the reference PRRs and studies of the modulation of their gene expression in response to *H. vastatrix*, one of the most devastating pathogens in coffee trees, presents an advance for understanding this crop basal immunity. In the present study, fungal PRR candidate sequences well described in the literature for model plants such as Arabidopsis and rice were studied in *C. arabica*. We observed that there is more than one candidate sequence for each receptor studied, which may be the result of the ploidy of this species or duplication of these receptors, a common mechanism in plant genomes [57]. Four candidate sequences for *CERK1* and two for *LYK5* in *C. arabica* presented higher percentages of identity compared to sequences of *C. eugenioides*, which may indicate duplications of this receptor in both species. Referring to *LYK4* and *LYP* (*CEBiP*-like), each candidate sequence showed greater identity with sequences of *C. eugenioides* or *C. canephora*. Therefore, it is possible to infer that those genes may have come from these genomes (Table 4).

The size of the CDS and the organization of exons demonstrated that the genes encoding *LYK4* and *LYK5* candidate proteins in *C. arabica* do not have introns, and the coding sequences are the result of a single exon. In fact, when compared to *CERK1* or *CEBiP*, these receptors are closer to each other in phylogenetic analysis. These results (Fig 4) corroborates with others described in the literature [54,58] and shows a greater evolutionary relationship between these receptors. Homologs of the *At-LYK4* and *At-LYK5* in many plant species have no introns and the coding region is the result of a single exon [25,58–61]. Regarding the LysM receptors homologous

474 to *At-CERK1*, the CDS region mostly presents around 1800 bp with ten to twelve exons [29,54,62],
475 which is likewise with the size of the CDS and number of exons found for the *CERK1* candidate
476 sequences in coffee, except for the Scaffold 539.592, which presents a larger coding region, with
477 2511bp and 13 exons. However, this number of thirteen exons has also been found in *Ps-LYK9*, a
478 *CERK1-like* gene in peas, which is involved in the control of plant immunity and symbiosis
479 formation [62].

480 Regarding the genes LYPs (Receptor-like proteins or RLPs) such as *Os-CEBiP*, the number
481 of exons reported is more variable from two to six [23,58,63]. In *C. arabica*, Scaffold 1196.90 and
482 439.212 presented four and five, respectively. The structural pattern of genes, such as the
483 distribution of introns or exons in gene families, reinforces the ortholog identification between
484 sequences since these are almost conserved among all orthologous. Minor differences may be due
485 to evolutionary changes or errors in gene structure predictions [59].

486 Characterization of domains and motifs (LysM)

487 Proteins classified as LYKs (Receptor-like kinases or RLKs) are composed of lysin motifs
488 (LysM)-containing ectodomains, a transmembrane domain and an intracellular kinase. LYP
489 proteins (RLPs), on the other hand, present LysM ectodomain, but without intracellular kinase and
490 can be anchored to the plasma membrane by a transmembrane domain or GPI-anchor [58,64]. The
491 *At-CERK1*, *At-LYK4* and *At-LYK5* contain three extracellular LysM motifs, a transmembrane
492 domain and intracellular kinase, while *Os-CEBiP* has two extracellular LysM motifs and GPI
493 anchor [22–24]. The SMART and PredGPI analysis predicted that the amino acid sequences of the
494 PRRs studied in *C. arabica* present a signal peptide, extracellular LysM motifs, a transmembrane
495 domain, or a putative signal sequence for the GPI anchor, besides the presence or absence of

intracellular kinase. These characteristics differentiate them into LYKs (Ca1 and 2 *CERK1*, Ca1 and 2 *LYK5* and *Ca-LYK4*) and LYPs (Ca1 and 2 *LYP*) (Fig 2) and suggest that they all act as membrane receptors.

As a result of the organization of the domains, these proteins have different protein sizes. LYKs are generally larger than LYPs because they have an additional kinase domain. Protein sequences reported for these classes of receptors are around 500 or 600 and 300 or 400 aa respectively [23,58,65]. Candidate sequences in coffee have equivalent sizes, except for Scaffold 539.592 with 836aa, which may be a consequence of the size of the coding region.

The PRR extracellular region varies in plant with sizes from 35 to 50 aa [57,58]. These regions define the type of recognized PAMP and its binding affinity in addition to the interaction between receptors and coreceptors [66]. Differences in the chitin-binding properties between *At/Os-CERK1* ectodomains show variation in the performance of these receptors in Arabidopsis and rice. *At-CERK1*and *At-LYK5*, for instance, bind directly to chitin through their ectodomains containing LysM motifs with different affinities to the ligand, while *At-LYK4* appears to be a co-receptor [22,24,67]. In rice, *Os-CERK1* does not bind to chitooligosaccharides and the heterodimerization between *Os-CERK1* and *Os-CEBiP* is necessary for the innate immune response in this species [21,68]. Distinction in the role of these receptors suggests that plants use different chitin binding and signaling strategies [25,69].

In *C. arabica*, this region varied from 38 to 49 aa and the candidate sequences showed a high degree of identity and/or similarity with the reference LysM sequences used, indicating a conserved extracellular structure [54,56]. For *CERK1*, eight residues reported as important for chitin binding in Arabidopsis are present in the Scaffold 2193.164 and Scaffold 476.38 sequences (seven identical and one similar), suggesting that they can bind chitin. However, complementary

519 data are still needed to clarify which would be the primary receptor and co-receptor of the innate
520 immunity in this species, and further studies of chitin-receptor and receptor-receptor interaction
521 are required.

522 Joint phylogenetic analysis

523 PRRs are conserved in several plant species [59].This conservation indicates a fundamental
524 importance of the PAMP recognition system [26]. The joint phylogenetic analysis showed that the
525 sequences selected as candidates for *CERK1* in coffee, were highly related to *Md-CERK1*, *Md-*
526 *CERK1-2*, *Ps-LYK9*, *Mm-LYK2*, *Vv-LYK1-1*, *Vv-LYK1-2*, *Os-CERK1* and *At-CERK* (Fig 4). All of
527 these proteins have been described as being involved in the defense against fungal pathogens
528 [21,22,29–31,54,62], suggesting that the studied sequences also participate in the defense
529 responses against this group of phytopathogens. Among the species compared, tomato and grape
530 have greater evolutionary proximity to coffee. *Bti9* (*Sl-LYK1*), a *CERK1* homolog in tomato, which
531 grouped more closely to the Scaffold 2193.164 and 476.38 sequences (*Ca2-CERK1*) in this clade,
532 presents an identity of 58.6% with *At-CERK* [70]. Candidate sequences in coffee, however, showed
533 around 57% of identity (Table 3).

534 The *Bti9* (*Sl-LYK1*) in tomato interacts with *AvrPtoB*, effector in *Pseudomonas syringae*.
535 The kinase region of this protein is the target and this results in blocking the PTI signaling [70].
536 Despite being described as a bacterial effector target, the study by Zeng et al., 2012 [70] or later
537 reports by Xin and He, 2013 [71] did not describe the interaction of this protein with chitin or the
538 transcriptional profiles regarding the response to fungal pathogens. Nonetheless, *Bti9* is a
539 membrane receptor with extracellular LysM motifs and high homology to *At-CERK1*.
540 Furthermore, the *At/Os-CERK1,* besides playing a role as a receptor for fungal PAMPs, also

participates as a co-receptor for PRRs in bacterial recognition [53,72], which demonstrates the multiple functions of this receptor and turns it into a possible target of bacterial and fungal effectors that suppress PTI.

The Ca1 and 2 *LYK 4* and 5, clades II and III, were grouped to grape receptors *Vv-LYK4-1/2* and *Vv-LYK5-1* (Fig 4). These were shown to be highly expressed during infection by *Botrytis cinerea* in grapevine fruits [50]. The clustering of *Bd-LYK4* in this clade corroborates the results presented by Tombuloglu et al., 2019 [58] for this PRR described in the *Brachypodium* genome, which presented a greater phylogenetic relationship to *At-LYK5*. In clade IV, the Ca1 and 2 *LYP* grouped, in addition to other homologs, to *Mm-LYP1*. The *Mm-LYP1* is a receptor described in white mulberry, besides having a high affinity for chitin, it displays a significant increase in transcriptional profiles in fruits and leaves of mulberry infested with popcorn disease. The *Mm-LYP1* interacts with *Mm-LYK2*, a homolog of *At-CERK1*, present in clade I and grouped with the candidate sequences for *CERK1* in *C. arabica*. The *Mm-LYK2* does not have a high affinity for chitin, but it functions as a co-receptor with intracellular kinase for the PTI signaling [31]. Additionally, in this clade, the *Hv-CEBiP* in barley, has been described for recognizing chitin oligosaccharides derived from *Magnaporthe oryzae* [28] and *Mt-LYM2*, in *Medicago truncatula*, demonstrated specific binding to biotinylated N-acetylchitooctaose in a similar way to *CEBiP* in rice [23,63]. Thus, the receptors cited for the phylogenetic groupings of this study reinforces the possible role of candidate sequences in *C. arabica* as PAMP receptors.

Transcriptional response of candidate receptors in *C. arabica*

The PAMPS are defined as highly conserved molecules from microorganisms and, therefore, have an essential function in their survival or fitness [73,74]. It is suggested that since PAMPs are essential for the viability or lifestyle of microorganisms, it is less likely that they avoid

host immunity through mutation or deletion in these regions [15,75]. Chitin is a PAMP present in the fungal cell wall. Fragments of N-acetylquitooligosaccharides are released by the breakdown of this PAMP by plant chitinases during plant-fungus interactions. These fragments serve as elicitors for the innate immunity of plants by modifying the transcriptional levels of PRRs [23].

In this study, the expression increases were detected from 6 hpi, showing that all candidate PRR were stimulated after the inoculation of *H. vastatrix*. The highest averages of expression were observed at 24 hpi, for most receptors, followed by a decrease at 48 hpi (Fig 5). These results describe an initial stimulus with subsequent suppression. The experiments showed that at 24 hpi it is already possible to detect the penetration of the hypha produced by the appressorium of *H. vastatrix* in stomata of coffee leaves, both in resistant and susceptible genotypes and at 48 hpi the presence of haustoria is already observed [76–78]. In addition, a LRR receptor-like kinase described in this pathosystem has a peak expression at 24 hpi in compatible and incompatible interactions [79], thus suggesting that the signal exchange between the two organisms is already occurring in this period.

To inhibit PTI, some fungal pathogens secrete proteins containing LysM motifs that compete with plant receptors [80,81]. These proteins seem to impede the detection of chitin polymers or interfere with the functioning of essential molecules in the downstream signaling of basal immunity. It is assumed that the decrease in PRR expression in *C. arabica* leaves, observed at 48 hpi, may be related to the suppression of PTI signaling. Fungal effectors such as *Ecp6*, *ChELP1/2* bind to chitin oligosaccharides released by the action of chitinases and prevent their recognition by the host PRR [80,82], while effectors like *Avr4* protect chitin from fungal cell walls from degradation by host chitinase [83]. In addition, a study of the *H. vastatrix* secretome showed that effector candidates expressed in incompatible interaction (resistance) were more abundant

587 within 24 hours, suggesting that these pre-haustorial effectors could be involved in the attempt to
588 suppress PTI [84].

589 The expression results of the candidate receptors did not show difference in profiles
590 between the groups of resistant and susceptible cultivars. Despite the IP showing high levels of
591 expression at 24 hpi for the transcripts *Ca1-LYP*, *Ca2-LYK5* and *Ca-LYK4*, the susceptible cultivar
592 MN showed equivalent levels of expression for *Ca1-CERK1* and *Ca2-LYP* or MN and CV showed
593 comparable levels or even larger than the AR resistant cultivar for *Ca2-CERK1*, *Ca2-LYP*, *Ca1-*
594 *LYK5* and *Ca-LYK4* (Fig 5). This result was expected, since the basal immunity is characterized
595 by being broad-spectrum and non-specific [13,18]. The resistance of coffee to rust has been
596 reported as pre-haustorial [78,85], in which resistant genotypes cease the growth of the fungus
597 with mechanisms of pathogen recognition by resistance proteins. Thus, the difference between
598 resistant and susceptible cultivars is generally evidenced in studies of expression of genes involved
599 in pathogen-specific pathways and not in broad-spectrum receptors, such as PRRs [85].

600 Additionally, the recognition and signaling of PAMPs occurs when PRRs associate and act
601 as part of multiprotein immune complexes on the cell surface [86,87]. Although they share
602 common structural characteristics, these receptors are distinct in terms of recognized expression
603 patterns and epitopes [24,26,53,63]. This shows that the receptors roles appear to have evolved
604 independently in different groups of plants [26,72]. Therefore, considering that all candidate
605 receptors in coffee, described in this study, increased their expression from 6 hpi in all evaluated
606 cultivars, each one may have possible roles in the basal immunity of *C. arabica.*

607 # Conclusion

608 The results indicate that candidate sequences in *C. arabica* have protein domains and
609 motifs characteristic of fungal PRRs and are homologous to *At-CERK1*, *At-LYK4*, *At-LYK5* and

610 *Os-CEBiP*. Additionally, the expression of these genes was increased after the inoculation of *H.*
611 *vastatrix* at all times and cultivars evaluated. Therefore, this study presents an advance in the
612 understanding of the basal immunity of this species.

613 # Acknowledgments

614 The authors would like to acknowledge the M.Sc. Antonio Carlos da Mota Porto for his help with
615 the statistical analysis.

616

617 # References

618 1. Klemptner RL, Sherwood JS, Tugizimana F, Dubery IA, Piater LA. Ergosterol, an orphan
619 fungal microbe-associated molecular pattern (MAMP). Mol Plant Pathol. 2014;15: 747–
620 761. doi:10.1111/mpp.12127

621 2. Boyd LA, Ridout C, O'Sullivan DM, Leach JE, Leung H. Plant-pathogen interactions:
622 Disease resistance in modern agriculture. Trends Genet. 2013;29: 233–240.
623 doi:10.1016/j.tig.2012.10.011

624 3. Jones JDG, Dangl JL. The plant immune system. Nature. 2006;444: 323–329.
625 doi:10.1038/nature05286

626 4. Böhm H, Albert I, Fan L, Reinhard A, Nürnberger T. Immune receptor complexes at the
627 plant cell surface. Curr Opin Plant Biol. 2014;20: 47–54. doi:10.1016/j.pbi.2014.04.007

628 5. Kourelis J, Van Der Hoorn RA. Defended to the Nines: 25 years of Resistance Gene
629 Cloning Identifies Nine Mechanisms for R Protein Function. Plant Cell. 2018;30: 285–
630 299. doi:10.1105/tpc.17.00579

631 6. Tsuda K, Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and
632 effector-triggered immunity. Curr Opin Plant Biol. 2010;13: 459–465.
633 doi:10.1016/j.pbi.2010.04.006

634 7. Jeworutzki E, Roelfsema MRG, Anschütz U, Krol E, Elzenga JTM, Felix G, et al. Early
635 signaling through the Arabidopsis pattern recognition receptors FLS2 and EFR involves
636 Ca^{2+}-associated opening of plasma membrane anion channels. Plant J. 2010;62: 367–378.
637 doi:10.1111/j.1365-313X.2010.04155.x

638 8. Nomura H, Komori T, Uemura S, Kanda Y, Shimotani K, Nakai K, et al. Chloroplast-
639 mediated activation of plant immune signalling in Arabidopsis. Nat Commun. 2012;3:
640 910–926. doi:10.1038/ncomms1926

641 9. Macho AP, Zipfel C. Plant PRRs and the activation of innate immune signaling. Mol Cell.
642 2014;54: 263–272. doi:10.1016/j.molcel.2014.03.028

643 10. Ranf S, Eschen-Lippold L, Pecher P, Lee J, Scheel D. Interplay between calcium
644 signalling and early signalling elements during defence responses to microbe- or damage-
645 associated molecular patterns. Plant J. 2011;68: 100–113. doi:10.1111/j.1365-
646 313X.2011.04671.x

647 11. Thomma BPHJ, Nürnberger T, Joosten MHAJ. Of PAMPs and Effectors: The Blurred
648 PTI-ETI Dichotomy. Plant Cell. 2011;23: 4–15. doi:10.1105/tpc.110.082602

649 12. Tsuda K, Katagiri F. Comparing signaling mechanisms engaged in pattern-triggered and
650 effector-triggered immunity. Curr Opin Plant Biol. 2010;13: 459–465.
651 doi:10.1016/j.pbi.2010.04.006

13. Ranf S. Sensing of molecular patterns through cell surface immune receptors. Curr Opin Plant Biol. 2017;38: 68–77. doi:10.1016/j.pbi.2017.04.011

14. Hegenauer V, Fürst U, Kaiser B, Smoker M, Zipfel C, Felix G, et al. Detection of the plant parasite Cuscuta reflexa by a tomato cell surface receptor. Science. 2016;353: 478–481. doi:10.1126/science.aaf3919

15. Lacombe S, Rougon-Cardoso A, Sherwood E, Peeters N, Dahlbeck D, Van Esse HP, et al. Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. Nat Biotechnol. 2010;28: 365–369. doi:10.1038/nbt.1613

16. Lee S, Whitaker VM, Hutton SF. Potential applications of non-host resistance for crop improvement. Front Plant Sci. 2016;7: 1–6. doi:10.3389/fpls.2016.00997

17. Lee HA, Lee HY, Seo E, Lee J, Kim SB, Oh S, et al. Current understandings on plant nonhost resistance. Mol Plant-Microbe Interact. 2016;30: 5–15. doi:10.1094/MPMI-10-16-0213-CR

18. Boutrot F, Zipfel C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. Annu Rev Phytopathol. 2017;55: annurev-phyto-080614-120106. doi:10.1146/annurev-phyto-080614-120106

19. Gómez-Gómez L, Boller T. FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. Mol Cell. 2000;5: 1003–1011. doi:10.1016/S1097-2765(00)80265-8

20. Zipfel C, Kunze G, Chinchilla D, Caniard A, Jones JDG, Boller T, et al. Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated

673 Transformation. Cell. 2006;125: 749–760. doi:10.1016/j.cell.2006.03.037

674 21. Shimizu T, Nakano T, Takamizawa D, Desaki Y, Ishii-Minami N, Nishizawa Y, et al.
675 Two LysM receptor molecules, CEBiP and OsCERK1, cooperatively regulate chitin
676 elicitor signaling in rice. Plant J. 2010;64: 204–214. doi:10.1111/j.1365-
677 313X.2010.04324.x

678 22. Miya A, Albert P, Shinya T, Desaki Y, Ichimura K, Shirasu K, et al. CERK1, a LysM
679 receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. Proc Natl Acad
680 Sci. 2007;104: 19613–19618. doi:10.1073/pnas.0705147104

681 23. Kaku H, Nishizawa Y, Ishii-minami N, Akimoto-tomiyama C, Dohmae N, Takio K, et al.
682 Plant cells recognize chitin fragments for defense signaling through a plasma membrane
683 receptor. Proc Natl Acad Sci. 2006;103: 11086–11091. doi:/10.1073/pnas.0508882103

684 24. Cao Y, Liang Y, Tanaka K, Nguyen CT, Jedrzejczak RP, Joachimiak A, et al. The kinase
685 LYK5 is a major chitin receptor in Arabidopsis and forms a chitin-induced complex with
686 related kinase CERK1. Elife. 2014;3: 1–19. doi:10.7554/eLife.03766

687 25. Wan J, Tanaka K, Zhang X-C, Son GH, Brechenmacher L, Nguyen THN, et al. LYK4, a
688 lysin motif receptor-like kinase, is important for chitin signaling and plant innate
689 immunity in Arabidopsis. Plant Physiol. 2012;160: 396–406. doi:10.1104/pp.112.201699

690 26. Liu B, Li JF, Ao Y, Qu J, Li Z, Su J, et al. Lysin Motif-Containing Proteins LYP4 and
691 LYP6 Play Dual Roles in Peptidoglycan and Chitin Perception in Rice Innate Immunity.
692 Plant Cell. 2012;24: 3406–3419. doi:10.1105/tpc.112.102475

693 27. Zipfel C, Robatzek S, Navarro L, Oakeley E, al et. Bacterial disease resistance in

694 Arabidopsis through flagellin perception. Nature. 2004;428: 15–18.
695 doi:10.1038/nature02485

696 28. Tanaka S, Ichikawa A, Yamada K, Tsuji G, Nishiuchi T, Mori M, et al. HvCEBiP, a gene
697 homologous to rice chitin receptor CEBiP, contributes to basal resistance of barley to
698 *Magnaporthe oryzae*. BMC Plant Biol. 2010;10. doi:10.1186/1471-2229-10-288

699 29. Zhou Z, Tian Y, Cong P, Zhu Y. Functional characterization of an apple (*Malus x*
700 *domestica*) LysM domain receptor encoding gene for its role in defense response. Plant
701 Sci. 2018;269: 56–65. doi:10.1016/j.plantsci.2018.01.006

702 30. Chen Q, Dong C, Sun X, Zhang Y, Dai H, Bai S. Overexpression of an apple LysM-
703 containing protein gene, MdCERK1-2, confers improved resistance to the pathogenic
704 fungus, *Alternaria* alternata, in *Nicotiana benthamiana*. BMC Plant Biol. 2020;20: 1–13.
705 doi:10.1186/s12870-020-02361-z

706 31. Lv Z, Huang Y, Ma B, Xiang Z, He N. LysM1 in MmLYK2 is a motif required for the
707 interaction of MmLYP1 and MmLYK2 in the chitin signaling. Plant Cell Rep. 2018;37:
708 1101–1112. doi:10.1007/s00299-018-2295-4

709 32. ICO. Crop year production by country. International Coffee Organization. July 2020
710 [Cited 2020 july 28] . Available from: https://www.ico.org/

711 33. Cabral PGC, Maciel-Zambolim E, Oliveira SAS, Caixeta ET, Zambolim L. Genetic
712 diversity and structure of *Hemileia vastatrix* populations on Coffea spp. Plant Pathol.
713 2016;65: 196–204. doi:10.1111/ppa.12411

714 34. McCook S, Vandermeer J. The Big Rust and the Red Queen: Long-Term Perspectives on

715 Coffee Rust Research. Phytopathology. 2015;105: 1164–1173. doi:10.1094/PHYTO-04-
716 15-0085-RVW

717 35. Avelino J, Cristancho M, Georgiou S, Imbach P, Aguilar L, Bornemann G, et al. The
718 coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible
719 causes and proposed solutions. Food Secur. 2015;7: 303–321. doi:10.1007/s12571-015-
720 0446-9

721 36. Zambolim L. Current status and management of coffee leaf rust in Brazil. Trop Plant
722 Pathol. 2016;41: 1–8. doi:10.1007/s40858-016-0065-9

723 37. Periyannan S, Milne RJ, Figueroa M, Lagudah ES, Dodds PN. An overview of genetic
724 rust resistance: From broad to specific mechanisms. PLoS Pathog. 2017;13: 1–6.
725 doi:10.1371/journal.ppat.1006380

726 38. Talhinhas P, Batista D, Diniz I, Vieira A, Silva DN, Loureiro A, et al. The coffee leaf rust
727 pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. Mol Plant
728 Pathol. 2017;18: 1039–1051. doi:10.1111/mpp.12512

729 39. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary
730 genetics analysis across computing platforms. Mol Biol Evol. 2018;35: 1547–1549.
731 doi:10.1093/molbev/msy096

732 40. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence
733 alignment, interactive sequence choice and visualization. Brief Bioinform. 2018;20: 1160–
734 1166. doi:10.1093/bib/bbx108

735 41. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis

program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41: 95–98. doi:10.1039/c7qi00394c

42. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, et al. Molecular characterization and origin of the *Coffea arabica* L. genome. Mol Gen Genet. 1999;261: 259–266.

43. Tran HTM, Ramaraj T, Furtado A, Lee LS, Henry RJ. Use of a draft genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content. Plant Biotechnol J. 2018;9: 1756–1766. doi:10.1111/pbi.12912

44. Monteiro ACA, de Resende MLV, Valente TCT, Ribeiro Junior PM, Pereira VF, da Costa JR, et al. Manganese phosphite in coffee defence against *Hemileia vastatrix* , the coffee rust fungus: biochemical and molecular analyses. J Phytopathol. 2016;164: 1043–1053. doi:10.1111/jph.12525

45. Pfaffl MW. A new mathematical model for relative quantification in real-time RT–PCR. Mon Not R Astron Soc. 2001;29: e45–e45. doi:10.1111/j.1365-2966.2012.21196.x

46. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG. Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions. BMC Mol Biol. 2009;10: 1–11. doi:10.1186/1471-2199-10-1

47. De Carvalho K, Bespalhok Filho JC, Dos Santos TB, De Souza SGH, Vieira LGE, Pereira LFP, et al. Nitrogen starvation, salt and heat stress in coffee (*Coffea arabica* L.): Identification and validation of new genes for qPCR normalization. Mol Biotechnol. 2013;53: 315–325. doi:10.1007/s12033-012-9529-4

757 48. Fernandes-Brum CN, Garcia B de O, Moreira RO, Ságio SA, Barreto HG, Lima AA, et al.
758 A panel of the most suitable reference genes for RT-qPCR expression studies of coffee:
759 screening their stability under different conditions. Tree Genet Genomes. 2017;13: 1–13.
760 doi:10.1007/s11295-017-1213-1

761 49. Reichel T, Resende MLV, Monteiro ACA, Freitas NC, Botelho, DMS. Constitutive
762 defense strategy of coffee under field conditions: a comparative assessment of resistant
763 and susceptible cultivars to rust. Mol. Biotechnol. 2021. doi:10.1007/s12033-021-00405-9

764 50. Xie F, Xiao P, Chen D, Xu L, Zhang B. miRDeepFinder: A miRNA analysis tool for deep
765 sequencing of plant small RNAs. Plant Mol Biol. 2012;80: 75–84. doi:10.1007/s11103-
766 012-9885-2

767 51. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den hoff MJB, et al.
768 Amplification efficiency: Linking baseline and bias in the analysis of quantitative PCR
769 data. Nucleic Acids Res. 2009;37: e45–e45. doi:10.1093/nar/gkp045

770 52. R Core Team. 2020 [cited 2020 August 30] in: The R Project for Statistical Computing
771 [Internet]. Vienna: The R Foundation for Statistical Computing. Available from:
772 https://www.r-project.org/

773 53. Willmann R, Lajunen HM, Erbs G, Newman MA, Kolb D, Tsuda K, et al. Arabidopsis
774 lysin-motif proteins LYM1 LYM3 CERK1 mediate bacterial peptidoglycan sensing and
775 immunity to bacterial infection. Proc Natl Acad Sci U S A. 2011;108: 19824–19829.
776 doi:10.1073/pnas.1112862108

777 54. Brulé D, Villano C, Davies LJ, Trdá L, Claverie J, Héloir MC, et al. The grapevine (*Vitis*
778 *vinifera*) LysM receptor kinases VvLYK1-1 and VvLYK1-2 mediate

779 chitooligosaccharide-triggered immunity. Plant Biotechnol J. 2019;17: 812–825.
780 doi:10.1111/pbi.13017

781 55. Liu T, Liu Z, Song C, Hu Y, Han Z, She J, et al. Chitin-induced dimerization activates a
782 plant immune receptor. Science. 2012;336: 1160–1164. doi:10.1126/science.1218867

783 56. Liu S, Wang J, Han Z, Gong X, Zhang H, Chai J. Molecular Mechanism for Fungal Cell
784 Wall Recognition by Rice Chitin Receptor OsCEBiP. Structure. 2016;24: 1192–1200.
785 doi:10.1016/j.str.2016.04.014

786 57. Zhang XC, Cannon SB, Stacey G. Evolutionary genomics of LysM genes in land plants.
787 BMC Evol Biol. 2009;9: 1–13. doi:10.1186/1471-2148-9-183

788 58. Tombuloglu G, Tombuloglu H, Cevik E, Sabit H. Genome-wide identification of Lysin-
789 Motif Receptor-Like Kinase (LysM-RLK) gene family in *Brachypodium distachyon* and
790 docking analysis of chitin/LYK binding. Physiol Mol Plant Pathol. 2019;106: 217–225.
791 doi:10.1016/j.pmpp.2019.03.002

792 59. Buendia L, Girardin A, Wang T, Cottret L, Lefebvre B. LysM receptor-like kinase and
793 lysM receptor-like protein families: An update on phylogeny and functional
794 characterization. Front Plant Sci. 2018;871: 1–25. doi:10.3389/fpls.2018.01531

795 60. Lohmann GV, Shimoda Y, Nielsen MW, Jorgensen FG, Grossmann C, Sandal N, et al.
796 Evolution and regulation of the *Lotus japonicus* LysM receptor gene family. Mol Plant-
797 Microbe Interact. 2010;23: 510–521. doi:10.1094/MPMI-23-4-0510

798 61. Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, et al. Molecular evolution of
799 lysin motif-type receptor-like kinases in plants. Plant Physiol. 2007;144: 623–636.

doi:10.1104/pp.107.097097

62. Leppyanen I V, Shakhnazarova VY, Shtark OY, Vishnevskaya NA, Tikhonovich IA, Dolgikh EA. Receptor-like kinase LYK9 in *Pisum sativum* L. Is the CERK1-like receptor that controls both plant immunity and AM symbiosis development. Int J Mol Sci. 2018;19. doi:10.3390/ijms19010008

63. Fliegmann J, Uhlenbroich S, Shinya T, Martinez Y, Lefebvre B, Shibuya N, et al. Biochemical and phylogenetic analysis of CEBiP-like LysM domain-containing extracellular proteins in higher plants. Plant Physiol Biochem. 2011;49: 709–720. doi:10.1016/j.plaphy.2011.04.004

64. Monaghan J, Zipfel C. Plant pattern recognition receptor complexes at the plasma membrane. Curr Opin Plant Biol. 2012;15: 349–357. doi:10.1016/j.pbi.2012.05.006

65. Nazarian-Firouzabadi F, Joshi S, Xue H, Kushalappa AC. Genome-wide in silico identification of LysM-RLK genes in potato (<i>Solanum tuberosum<i/> L.). Mol Biol Rep. 2019;46: 5005–5017. doi:10.1007/s11033-019-04951-z

66. Xue DX, Li CL, Xie ZP, Staehelin C, Napier R. LYK4 is a component of a tripartite chitin receptor complex in *Arabidopsis thaliana*. J Exp Bot. 2019;70: 5507–5516. doi:10.1093/jxb/erz313

67. Shinya T, Motoyama N, Ikeda A, Wada M, Kamiya K, Hayafune M, et al. Functional Characterization of CEBiP and CERK1 Homologs in Arabidopsis and Rice Reveals the Presence of Different Chitin Receptor Systems in Plants. Plant Cell Physiol. 2012;53: 1696–1706. doi:10.1093/pcp/pcs113

68. Ao Y, Li Z, Feng D, Xiong F, Liu J, Li JF, et al. OsCERK1 and OsRLCK176 play important roles in peptidoglycan and chitin signaling in rice innate immunity. Plant J. 2014;80: 1072–1084. doi:10.1111/tpj.12710

69. Kaku H, Shibuya N. Molecular mechanisms of chitin recognition and immune signaling by LysM-receptors. Physiol Mol Plant Pathol. 2016;95: 60–65. doi:10.1016/j.pmpp.2016.02.003

70. Zeng L, Velásquez AC, Munkvold KR, Zhang J, Martin GB. A tomato LysM receptor-like kinase promotes immunity and its kinase activity is inhibited by AvrPtoB. Plant J. 2012;69: 92–103. doi:10.1038/jid.2014.371

71. Xin XF, He SY. *Pseudomonas syringae* pv. tomato DC3000: A model pathogen for probing disease susceptibility and hormone signaling in plants. Annu Rev Phytopathol. 2013;51: 473–498. doi:10.1146/annurev-phyto-082712-102321

72. Desaki Y, Kohari M, Shibuya N, Kaku H. MAMP-triggered plant immunity mediated by the LysM-receptor kinase CERK1. J Gen Plant Pathol. 2019;85: 1–11. doi:10.1007/s10327-018-0828-x

73. Albert M. Peptides as triggers of plant defence. J Exp Bot. 2013;64: 5269–5279. doi:10.1093/jxb/ert275

74. Naito K, Taguchi F, Suzuki T, Inagaki Y, Toyoda K, Shiraishi T, et al. Amino Acid Sequence of Bacterial Microbe-Associated Molecular Pattern flg22 Is Required for Virulence. Mol Plant-Microbe Interact. 2008;21: 1165–1174. doi:10.1094/MPMI-21-9-1165

842 75. Dickman MB, Fluhr R. Centrality of host cell death in plant-microbe interactions. Annu
843 Rev Phytopathol. 2013;51: 543–570. doi:10.1146/annurev-phyto-081211-173027

844 76. Ramiro DA, Escoute J, Petitot AS, Nicole M, Maluf MP, Fernandez D. Biphasic
845 haustorial differentiation of coffee rust (*Hemileia vastatrix* race II) associated with
846 defence responses in resistant and susceptible coffee cultivars. Plant Pathol. 2009;58:
847 944–955. doi:10.1111/j.1365-3059.2009.02122.x

848 77. Silva MC, Nicole M, Guerra-Guimarães L, Rodrigues CJ. Hypersensitive cell death and
849 post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in
850 resistant coffee leaves. Physiol Mol Plant Pathol. 2002;60: 169–183.
851 doi:10.1006/pmpp.2002.0389

852 78. Silva MC, Guerra-Guimarães L, Loureiro A, Nicole MR. Involvement of peroxidases in
853 the coffee resistance to orange rust (*Hemileia vastatrix*). Physiol Mol Plant Pathol.
854 2008;72: 29–38. doi:10.1016/j.pmpp.2008.04.004

855 79. Almeida DP De, Castro ISL, Mendes TA de O, Alves DR, Barka GD, Barreiros PRRM, et
856 al. Receptor-like kinase (RLK) as a candidate gene conferring resistance to Hemileia
857 vastatrix in coffee. Sci Agric. 2020;78: 1–9. doi:10.1590/1678-992x-2020-0023

858 80. Takahara H, Hacquard S, Kombrink A, Hughes HB, Halder V, Robin GP, et al.
859 Colletotrichum higginsianum extracellular LysM proteins play dual roles in appressorial
860 function and suppression of chitin-triggered plant immunity. New Phytol. 2016;211:
861 1323–1337. doi:10.1111/nph.13994

862 81. Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, et
863 al. Analysis of two in planta expressed LysM effector homologs from the fungus

mycosphaerella graminicola reveals novel functional properties and varying contributions to virulence on wheat. Plant Physiol. 2011;156: 756–769. doi:10.1104/pp.111.176347

82. De Jonge R, Van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, et al. Conserved Fungal LysM Effector Ecp6 Prevents Chitin-Triggered Immunity in Plants. Science. 2010;329: 953–955. doi:10.1126/science.1190859

83. Van Den Burg HA, Harrison SJ, Joosten MHAJ, Vervoort J, De Wit PJGM. *Cladosporium fulvum* Avr4 protects fungal cell walls against hydrolysis by plant chitinases accumulating during infection. Mol Plant-Microbe Interact. 2006;19: 1420–1430. doi:10.1094/MPMI-19-1420

84. Porto BN, Caixeta ET, Mathioni SM, Vidigal PMP, Zambolim L, Zambolim EM, et al. Genome sequencing and transcript analysis of *Hemileia vastatrix* reveal expression dynamics of candidate effectors dependent on host compatibility. PLoS One. 2019;14: 1–23. doi:10.1371/journal.pone.0215598

85. Ganesh D, Petitot AS, Silva MC, Alary R, Lecouls AC, Fernandez D. Monitoring of the early molecular resistance responses of coffee (*Coffea arabica* L.) to the rust fungus (Hemileia vastatrix) using real-time quantitative RT-PCR. Plant Sci. 2006;170: 1045–1051. doi:10.1016/j.plantsci.2005.12.009

86. Sun Y, Li L, Macho AP, Han Z, Hu Z, Zipfel C, et al. Structural basis for flg22-induced activation of the Arabidopsis FLS2-BAK1 immune complex. Science. 2013;342: 624–628. doi:10.1126/science.1243825

87. Tang D, Wang G, Zhou J-M. Receptor Kinases in Plant-Pathogen Interactions: More Than Pattern Recognition. Plant Cell. 2017;29: 618–637. doi:10.1105/tpc.16.00891

Supporting information

S1 Fig. Germination of *H. vastatrix* spores observed by optical microscope after 48 hours of inoculum preparation.

S2 Fig. Symptoms and signs of *H. vastatrix* in *C. arabica* seedlings.

(A, B, C, D) Cultivar Mundo Novo IAC 367-4, (E, F) Catuaí Vermelho. (A) abaxial face 20 days after inoculation of the pathogen, (B) adaxial face 20 days after inoculation, (C, E) abaxial face 35 days after inoculation, (D) adaxial face 35 days after inoculation.

S3 Fig. Stability ranking of the reference genes *14-3-3*, *GAPDH*, *EF1a* and *24S* obtained by RefFinder tool.

(A) Experiments 1, (B) Experiment 2. GM: Geometric mean of the weights from algorithms Delta-Ct, BestKeeper, NormFinder e geNorm.

S4 Fig. Alignments of CDS from candidate sequences to *CERK1* (*Ca1-CERK1* Scaffold_539.592 e Scaffold_1805.113). The alignments were obtained by CLC Genomics Workbench software. Gray bars show the conservation level of the positions; red letters, the different nucleotides; and red dashes, the gaps. Identity: 71, 33%.

S5 Fig. Alignments of CDS from candidate sequences to *CERK1* (*Ca2-CERK1* (Scaffold_2193.164 e Scaffold_476.38) in *C. arabica.* The alignments were obtained by CLC

907 Genomics Workbench software. Gray bars show the conservation level of the positions; red letters,
908 the different nucleotides; and red dashes, the gaps. Identity: 98,28%.
909
910 **S6 Fig. Alignments of CDS from candidate sequences to *LYK4* (Scaffold_612.376 e**
911 **Scaffold_952.320) in *C. arabica.*** The alignments were obtained by CLC Genomics Workbench
912 software. Gray bars show the conservation level of the positions; red letters, the different
913 nucleotides; and red dashes, the gaps. Identity: 98,45%.
914 **S1 Table. BLASTp analysis of the PRR reference sequences against the *C. arabica* genome**
915 **in Phytozome.**

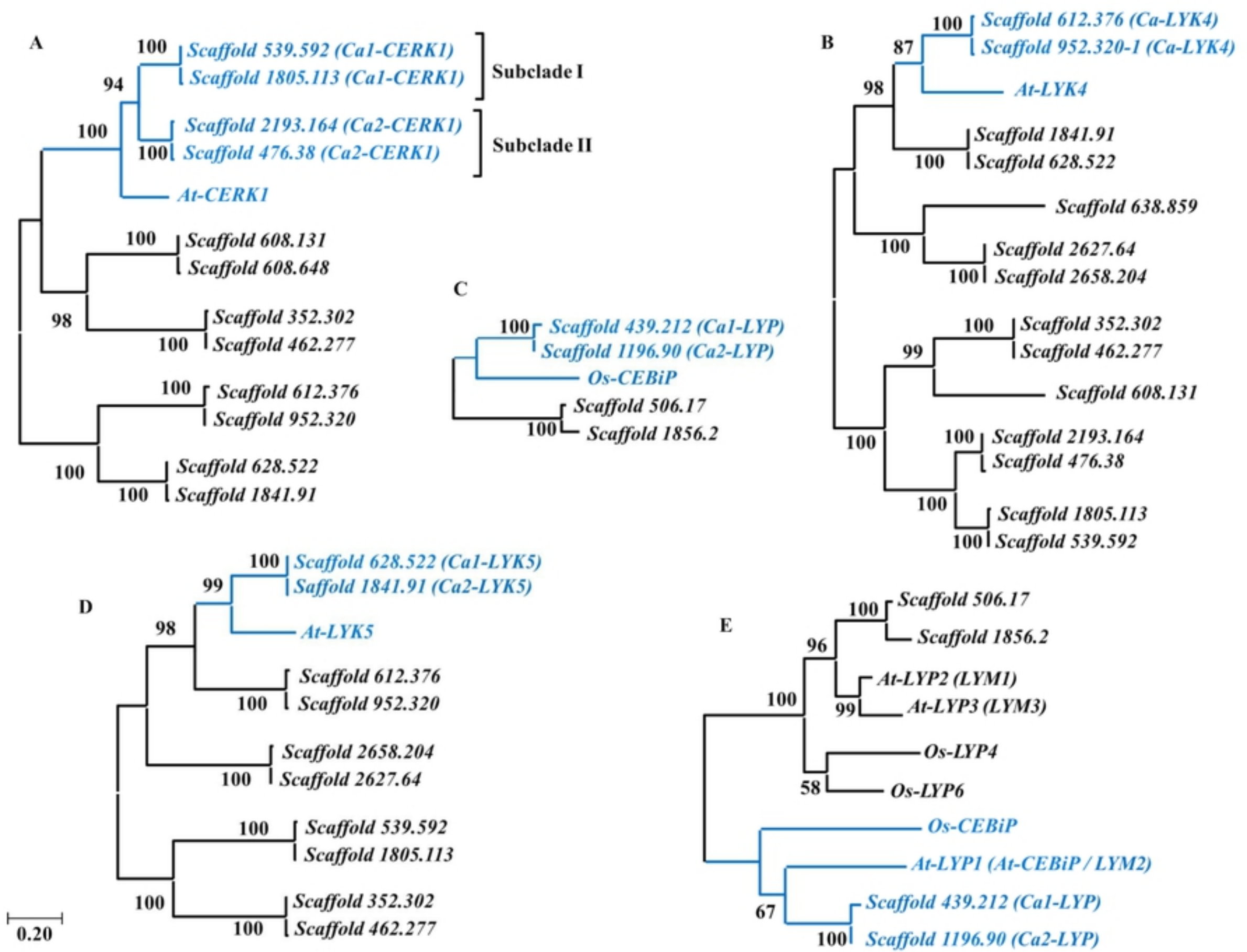
A
100
Scaffold 539.592 (Ca1-CERK1)
Scaffold 1805.113 (Ca1-CERK1)
Subclade I
94
100
Scaffold 2193.164 (Ca2-CERK1)
100
Scaffold 476.38 (Ca2-CERK1)
Subclade II
At-CERK1
100
Scaffold 608.131
Scaffold 608.648
98
Scaffold 352.302
100
Scaffold 462.277
100
Scaffold 612.376
Scaffold 952.320
100
Scaffold 628.522
100
Scaffold 1841.91
B
100
Scaffold 612.376 (Ca-LYK4)
87
Scaffold 952.320-1 (Ca-LYK4)
98
At-LYK4
Scaffold 1841.91
100
Scaffold 628.522
Scaffold 638.859
100
Scaffold 2627.64
100
Scaffold 2658.204
100
Scaffold 352.302
99
Scaffold 462.277
Scaffold 608.131
100
Scaffold 2193.164
Scaffold 476.38
100
Scaffold 1805.113
100
Scaffold 539.592
C
100
Scaffold 439.212 (Ca1-LYP)
Scaffold 1196.90 (Ca2-LYP)
Os-CEBiP
Scaffold 506.17
100
Scaffold 1856.2
D
100
Scaffold 628.522 (Ca1-LYK5)
99
Saffold 1841.91 (Ca2-LYK5)
98
At-LYK5
Scaffold 612.376
100
Scaffold 952.320
Scaffold 2658.204
100
Scaffold 2627.64
100
Scaffold 539.592
Scaffold 1805.113
100
Scaffold 352.302
100
Scaffold 462.277
0.20
E
100
Scaffold 506.17
96
Scaffold 1856.2
At-LYP2 (LYM1)
100
99
At-LYP3 (LYM3)
Os-LYP4
58
Os-LYP6
Os-CEBiP
At-LYP1 (At-CEBiP / LYM2)
67
Scaffold 439.212 (Ca1-LYP)
100
Scaffold 1196.90 (Ca2-LYP)

Figure 1

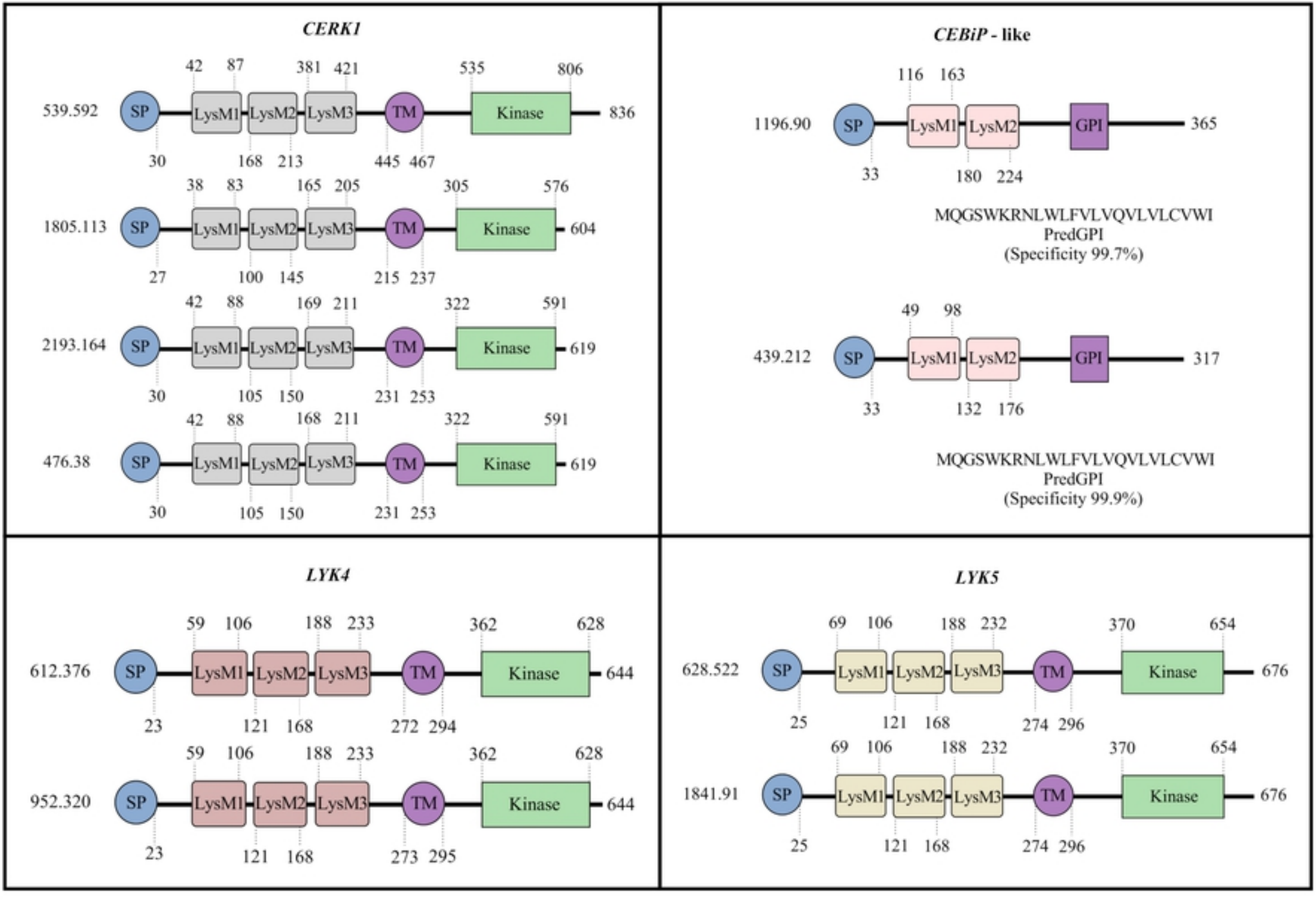
CERK1
539.592
SP
LysM1
LysM2
LysM3
TM
Kinase
836
42 87 381 421 535 806
30 168 213 445 467
1805.113
604
38 83 165 205 305 576
27 100 145 215 237
2193.164
619
42 88 169 211 322 591
30 105 150 231 253
476.38
619
42 88 168 211 322 591
30 105 150 231 253
CEBiP - like
1196.90
GPI
365
116 163
33 180 224
MQGSWKRNLWLFVLVQVLVLCVWI
PredGPI
(Specificity 99.7%)
439.212
317
49 98
33 132 176
MQGSWKRNLWLFVLVQVLVLCVWI
PredGPI
(Specificity 99.9%)
LYK4
612.376
644
59 106 188 233 362 628
23 121 168 272 294
952.320
644
59 106 188 233 362 628
23 121 168 273 295
LYK5
628.522
676
69 106 188 232 370 654
25 121 168 274 296
1841.91
676
69 106 188 232 370 654
25 121 168 274 296

Figure 2

CERK1

			LysM1	
At-CERK1	SYYLENGTTLSVINQNLNSSIAPYDQINFDPILRYNS-NIKDKDRIQMGSRVLVP	54		
539.592	SYYIWEGSNLTYISSIFDQTIP---------EILRQNP-HVPNQDSIHSGTRINIP	46	*29,63%	**53,70%
1805.113	SYYAWNGTNLTFISTVLSTSIS---------HILKYNP-QITNPDIIQFGSRISVP	46	*44,44%	**59,26%
2193.164	SYDVWRGSNVTLIADLFSVPVS---------TLLSWNPVTLPDRDTVIAGTRVNIP	47	*21,82%	**49,10%
476.38	SFDVWRGSNVTIIAQLFSVPVS---------TLLSWNPVTLPDTNTVIAGTRINIP	47	*18,18%	**49,10%

			LysM2	
At-CERK1	SYSVRQEDTYERVAISNYANLTTMESLQARNPFPATNIPLSATLNVL----	47		
539.592	-YPLRPGENLSSVAN---ASGAPAELLQRFN--PGSNFSAGSGIVFVPAKV	45	*21,57%	**35,29%
1805.113	--RVRSATYYRRIAHLVYANLTTVEMLQRFNSYPPENVPAAAQLNVT----	45	*46,81%	**57,45%
2193.164	--SVSSGDTYDIVASQFYANLTSTTWLRRFNSYPANNIPDTGVLNVT----	45	*46,81%	**57,45%
476.38	--SVSTGDTYDLVASRNYANLTSTTWLRRFNSYPANNIPDTGFLNVT----	45	*48,94%	**59,57%

			LysM3	
At-CERK1	TYPLRPEDSLSSIARSSGVSADILQRYNPGVNFNSGNGIVYV-	42		
539.592	-YPLRSGETISSLANEFDLPEKLLEDYNPRVNFSGGSGLIFV-	41	*40,48%	**69,05%
1805.113	-YPLRSGETISSLANEFDLPEKLLEDYNPRVNFSGGSGLIFV-	41	*40,48%	**69,05%
2193.164	-WPIAVGDTLQSVASANNLSANLISRYNPTANFTSGSGLLFIP	42	*37,21%	**72,09%
476.38	-WPIAVGDTLQSVASANNLSANLISRYNPTANFTSGSGLLFIP	42	*37,21%	**72,09%

LYK4

			LysM1	
At-LYK4	VIFRSTPSFSTVTSISSLFSVDPSLVSSLNDASPSTSFPSGQQVIIP	47		
612.376	LTFRSLPPFNSVSSISSLLAADPSHLSQLNKVSQDATFETNRTVLVP	47	*44,68%	**68,08%
952.320	LTFRSLPPFNSVSSISSLLAADPSHLSQLNKVSQDATFETNRTVLVP	47	*44,68%	**68,08%

			LysM2	
At-LYK4	TYTIQPNDSYFAIANDTLQGLSTCQALAKQN-NVSSQSLFPGMRIVVP	47		
612.376	SYVIQHGNTYLSIANSTFQGLSTCQALQAQNANLSTVNLIAGTRIRVP	48	*56,25%	**70,83%
952.320	SYVIQHGNTYFSIANSTFQGLSTCQALQAQNANLSTGNLIVGTRIRVP	48	*58,33%	**72,92%

			LysM3	
At-LYK4	SYTVVFEDTIAIISDRFGVETSKTLKANEMSFENSEVFPFTTILIP	46		
612.376	SYLVTWGQYVAAISSMFGVDTGKTLQANGLSEQNFNIYPFTTLLVP	46	*50,00%	**71,74%
952.320	SYLVTWGQYVSAISSMFGVDTGKTLQANGLSEQNFNIYPFTTLLVP	46	*47,83%	**71,74%

LYK5

			LysM1	
At-LYK5	NTADSIAKLLNVSAAEIQSINNLPTATTRIPTRELVVIP	39		
628.522	NSPVTIAYLLDTDATEIARINNV-SDVGRIPSGTLIIVP	38	*41,02%	**64,10%
1841.91	NSPVTIAYLLDTDATEIARINNV-SDVGRIPSGTLIIVP	38	*41,02%	**64,10%

			LysM2	
At-LYK5	---RGD-ETYFSVANDTYQALSTCQAMMSQNRYGERQLTPGLNLLVP	43		
628.522	YVLKGTVETYYAVANETYQGLTTCQSLQAQNSYNFRNLKVNMKLNIP	47	*44,68%	**65,96%
1841.91	YVLKGTVETYYAVANETYQGLTTCQSLQAQNSYNFRNLKVNMKLNIP	47	*44,68%	**65,96%

			LysM3	
At-LYK5	TYLVAMGDSISGIAEMFNSTSAAITEGNELTSDNIFF-FTPVLVP	44		
628.522	AYLITWGDSFEAIASMFNADVQSIYAANELSPNHLIHPFNPLLIP	45	*40,00%	**60,00%
1841.91	AYLITWGDSFEAIASMFNADVQGIYAANELSPNHLIHPLNPLLIP	45	*37,78%	**55,56%

CEBiP

			LysM1	
Os-CEBiP	IYVVQPQDGLDAIARNVFNAFVTYQEIAAANNI----PDPNKINVSQTLWIP	48		
1196.90	IYTVVPNDFLYHIAAEVFSGLVTSQQIQATNNI----SNANLIYAGQKLWIP	48	*52,08%	**58,33%
439.212	-YVLPNSTTLSRI-QTLFN-VKNLTSILGANNLPLSTPPQRTFPANQTVKIP	49	*28,85%	**40,38%

			LysM2	
Os-CEBiP	AYSVGKGENTSAIAAKYGVTESTLLTRNKIDDPTKLQMGQILDVP	45		
1196.90	GYAVPARSSVDGIAQQYNTTADVLLRLNGLASPNDLKAGAILDVP	45	*37,78%	**48,88%
439.212	GYAVPARSSVDGIAQQYNTTADVLLRLNGLASPNDLKAGAILDVP	45	*37,78%	**48,88%

Figure 3

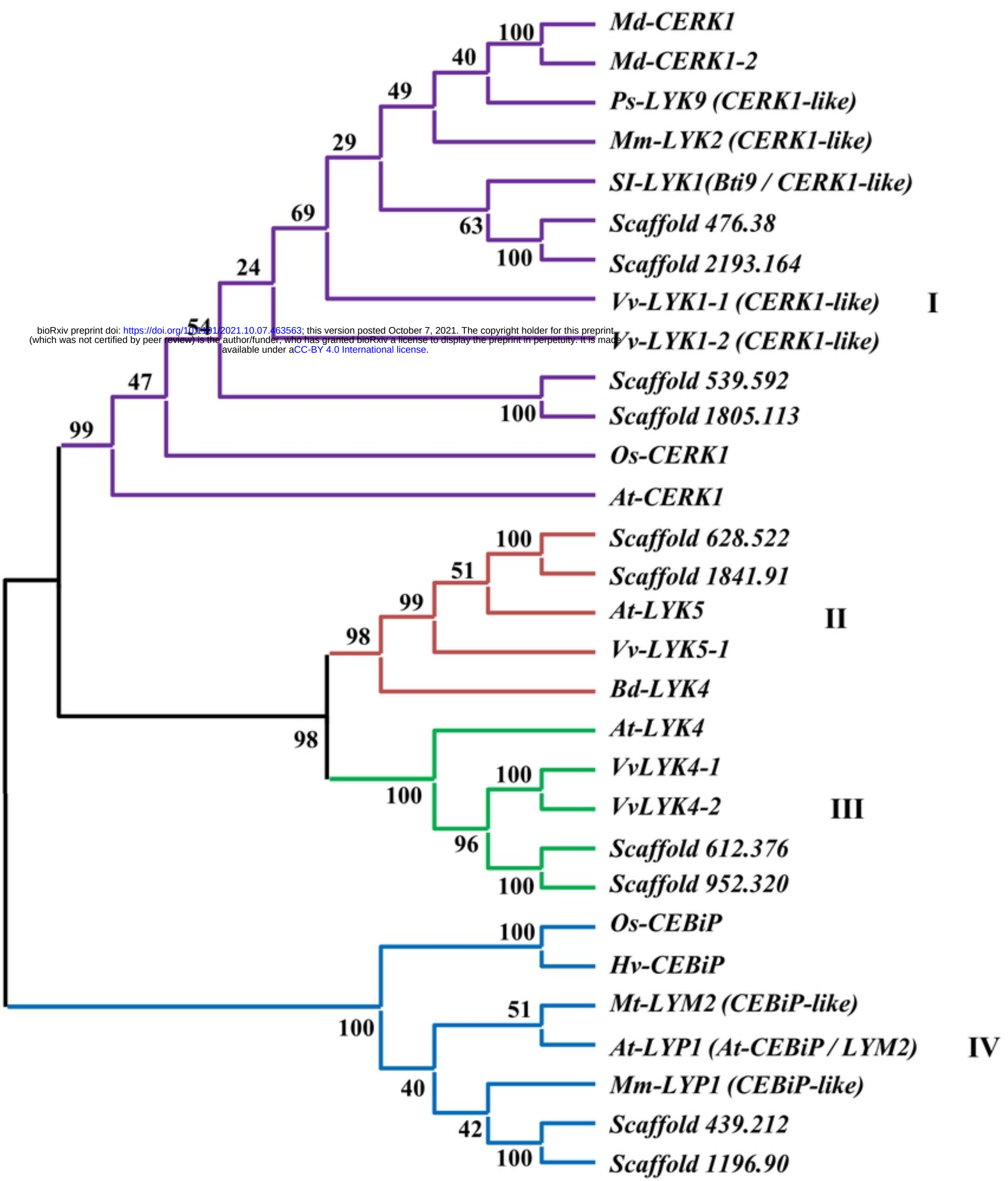

Figure 4

Figure 5