MARIANA DE LIMA SANTOS

IDENTIFICATION, CHARACTERIZATION AND COMPARATIVE ANALYSIS OF PATTERN RECOGNITION RECEPTORS (PRR) AND NUCLEOTIDE BINDING SITE-LEUCINE RICH REPEAT (NBS-LRR) IN *Coffea spp.* GENOME

LAVRAS - MG
2021

MARIANA DE LIMA SANTOS

IDENTIFICATION, CHARACTERIZATION AND COMPARATIVE ANALYSIS OF PATTERN RECOGNITION RECEPTORS (PRR) AND NUCLEOTIDE BINDING SITE-LEUCINE RICH REPEAT (NBS-LRR) IN *Coffea spp.* GENOME

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Biotecnologia Vegetal, para a obtenção do título de Doutor.

Dr. Mário Lúcio Vilela de Resende
Orientador

LAVRAS – MG
2021

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Santos, Mariana de Lima.

Identification,characterization and comparative analysis of pattern recognition receptors (PRR)and nucleotide binding site-leucine rich repeat (NBS-LRR) in *Coffea* spp.Genome / Mariana de Lima Santos. - 2021.

123 p.

Orientador(a): Mario Lúcio Vilela de Resende.

Tese (doutorado) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. PRR. 2. NBS-LRR. 3. Coffea. I. de Resende, Mario Lúcio Vilela. II. Título.

MARIANA DE LIMA SANTOS

IDENTIFICATION, CHARACTERIZATION AND COMPARATIVE ANALYSIS OF PATTERN RECOGNITION RECEPTORS (PRR) AND NUCLEOTIDE BINDING SITE-LEUCINE RICH REPEAT (NBS-LRR) IN *Coffea spp.* GENOME

IDENTIFICAÇÃO, CARACTERIZAÇÃO E ANÁLISE COMPARATIVA DE RECEPTORES DE RECONHECIMENTO DE PADRÕES (PRR) E SÍTIO DE LIGAÇÃO DE NUCLEOTÍDEO-REPETIÇÕES RICAS EM LEUCINA (NBS-LRR) NO GENOMA DE *Coffea spp.*

APROVADA em 28 de outubro de 2021.

Dr. Mario Lúcio Vilela de Resende UFLA/INCT-Café
Dra. Leonor Guerra-Guimarães ULisboa/CIFC
Dra. Eveline Teixeira Caixeta EMBRAPA Café /UFV
Dr. Márcio Fernando Ribeiro de Resende Júnior UF
Dr. Jose Carlos Huguet-Tapia UF

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Biotecnologia Vegetal, área de concentração em Biotecnologia Vegetal, para a obtenção do título de Doutor.

Dr. Mário Lúcio Vilela de Resende
Orientador

LAVRAS – MG
2021

Para aqueles que tiveram medo, mas foram com medo mesmo,
Ofereço!

Para os meus avós, ***Maria Lindaura*** *e* ***Euclides*** *(in memoriam), que formaram a minha essência,*
Dedico!

AGRADECIMENTOS

À minha família, em especial ao meu pai Almir, minha mãe Cleunice, meu irmão João Luiz, minha irmãzinha Júlia, vovó Lindaura e vovô Euclides (*in memoriam*), que estão presentes em todas as minhas conquistas.

Ao meu "namorido" Antonio Carlos (Tony) por todo amor e companheirismo nessa longa jornada, desde a iniciação científica, sempre me incentivando na busca de novos desafios.

À minha gatinha Joana Bosca (Bosquinha), por alegrar todos os meus dias desde que chegou, sempre dormindo ao lado do computador (rs) enquanto eu trabalhava na tese.

A todos do laboratório de Fisiologia do Parasitismo, por me aceitarem como parte do grupo e pelo convívio e amizade, em especial para Dra. Deila (Deiloca, Deiloca querida!) por ser minha mãezona e amiga em lavras, por todas as conversas, idas ao supermercado, açaís, bolos e presentes de aniversário. Agradeço também a Bárbara, Matheus, Natália e Tharyn por toda ajuda e orientação para a condução do primeiro artigo da tese.

Ao meu orientador, Prof. Dr. Mário Lúcio por aceitar me orientar, confiar no meu trabalho e me proporcionar oportunidades para meu crescimento profissional.

À Dra. Sandra Mathioni por ser a idealizadora do projeto base do primeiro artigo da tese, o que foi essencial para a obtenção de recursos e desenvolvimento do artigo.

Ao Dr. Márcio Resende por ter me dado a oportunidade de passar um período do doutorado como Visiting Research na University of Florida, trabalhar com bioinformática e conhecer o Dr. Jose Carlos Huguet-Tapia e o Dr. Jeremy Brawner. Experiência que me trouxe enorme aprendizado profissional e principalmente pessoal.

Ao Dr. Jose Carlos Huguet-Tapia por me acompanhar durante o desenvolvimento do segundo artigo da tese, sempre disposto a tirar minhas dúvidas e me ensinar novas possibilidades.

Ao Dr. Gabriel, pela parceria desde o mestrado, parceria esta que me fez enxergar um gap que resultou no segundo artigo da tese.

À toda comunidade científica de bioinformática pelos vídeos e tutoriais, aos quais recorri por inúmeras vezes.

À Universidade Federal de Lavras e ao programa de Biotecnologia Vegetal, pela oportunidade de realização do doutorado.

À FAPEMIG e INCT do Café pelo apoio financeiro, uso de equipamentos científicos e utilização das cultivares de café para coleta de frutos e inóculo de ferrugem para a condução do primeiro artigo da tese.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Código de financiamento 001, pela bolsa de estudo concedida.

Muito Obrigada!

"*O que é muito difícil é você vencer a injustiça secular que dilacera o Brasil em dois países distintos: o país dos privilegiados e o país dos despossuídos*."

Ariano Suassuna

RESUMO

O café apresenta destaque no agronegócio mundial, especialmente no Brasil. Entretanto, a produção desta commodity tem sido afetada devido à ocorrência de diversas doenças, dentre elas a ferrugem, cujo agente etiológico é o fungo biotrófico *Hemileia vastatrix*. Uma estratégia promissora para o controle de doenças é o estudo dos receptores que desencadeiam a sinalização para o mecanismo de resistência em plantas. Portanto, os objetivos deste trabalho foram identificar e caracterizar receptores de reconhecimento de padrões (PRRs) no genoma de *Coffea arabica* e analisar a expressão gênica destes receptores em cultivares contrastantes de *C. arabica* inoculada com *H. vastatrix*. Além de identificar loci NLR (*nucleotide-binding site leucine-rich repeat* - NBS-LRR) em genomas de *C. arabica*, *C. canephora* e *C. eugenioides* usando o *NLR-annotator*; caracterizar a distribuição destes loci nos genomas de *Coffea spp.* e compreender a contribuição de *C. canephora* e *C. eugenioides* para o repertório de NLRs em *C. arabica*. Foram utilizadas abordagens baseadas no princípio de similaridade de sequência, conservação de motivos e domínios, análises filogenéticas, modulação da expressão gênica e análise de grupos ortólogos. Os resultados demostram que os PRRs candidatos em *C. arabica* (*Ca1-LYP*, *Ca2-LYP*, *Ca1-CERK1*, *Ca2-CERK1*, *Ca-LYK4*, *Ca1-LYK5* e *Ca2-LYK5*) apresentam alta similaridade com PRRs de referência usados: *Os-CEBiP*, *At-CERK1*, *At-LYK4* e *At-LYK5*. Os ectodomínios destes receptores apresentaram alta identidade ou similaridade com as sequências de referência, indicando conservação estrutural e funcional. Os PRRs candidatos são filogeneticamente relacionadas aos PRRs de referência (em Arabidopsis e arroz) e aqueles descritos em outras espécies de plantas. Todos os receptores candidatos tiveram sua expressão induzida após a inoculação com *H. vastatrix*, desde o primeiro tempo avaliado, às 6 horas pós-inoculação (hpi). Houve um aumento significativo às 24 hpi para a maioria dos receptores avaliados e uma supressão às 48 hpi. Um total de 1311 loci NLR não redundantes foram identificados em *C. arabica*, 927 em *C. canephora* e 1079 em *C. eugenioides*, dos quais 809, 562 e 695 são loci completos, respectivamente. O *NLR-annotator* apresentou alta sensibilidades e especificidades (acima de 99%) para identificar loci NLRs em café, além de aumentar a capacidade de detecção de NLR putativos nos genomas estudados. Os loci NLRs no café são distribuídos em todos os cromossomos e são organizados principalmente em clusters. O genoma de *C. arabica* apresenta um número menor de loci NLR quando comparado à soma dos genomas parentais (*C. canephora* e *C. eugenioide*s). Existem NLRs ortólogos (ortogrupos) compartilhados entre café, tomate, batata e NLRs de referência e aqueles que são compartilhados apenas entre espécies de café. A análise filogenética demonstrou NLRs ortólogos compartilhados entre *C. arabica* e os genomas parentais e aqueles que foram possivelmente perdidos. Os membros da família NLR no café são subdivididos em dois grupos principais: TIR-NLR (TNL) e não-TNL. Os não-TNLs parecem representar um importante repertório de genes de resistência em café. Esses resultados podem subsidiar estudos funcionais de PRRs e NLRs e contribuir para o uso destes receptores no melhoramento genético do café visando o desenvolvimento de cultivares resistentes.

Palavras-chave: Resistência de amplo espectro. Receptores de reconhecimento de padrões. Ferrugem do café. Coffea. Genes de resistência NBS-LRR.

ABSTRACT

Coffee stands out in world agribusiness, especially in Brazil. However, the production of this commodity has been affected due to the occurrence of several diseases, including rust, whose etiologic agent is the biotrophic fungus *Hemileia vastatrix*. A promising strategy for disease control is the identification and study of receptors that trigger the signaling for the resistance mechanism in plants. Therefore, the objectives of this work were to identify and characterize pattern recognition receptors (PRRs) in the *Coffea arabica* genome and analyze the gene expression of these receptors in contrasting cultivars of *C. arabica* inoculated with *H. vastatrix*. Besides identifying NLR loci (nucleotide-binding leucine-rich repeat site - NBS-LRR) in *C. arabica*, *C. canephora* and *C. eugenioides* genomes using the NLR-annotator; characterize the distribution of these loci in *Coffea spp*. and understand the contribution of *C. canephora* and *C. eugenioides* to the NLR repertoire of *C. arabica*. Approaches based on the principle of sequence similarity, motif and domain conservation, phylogenetic analysis, gene expression modulation and ortholog group analysis were used. The results demonstrate that the candidate PRRs in *C. arabica* (*Ca1-LYP*, *Ca2-LYP*, *Ca1-CERK1*, *Ca2-CERK1*, *Ca-LYK4*, *Ca1-LYK5* and *Ca2-LYK5*) have high similarity with the reference PRRs used: *Os-CEBiP*, *At-CERK1*, *At-LYK4* and *At-LYK5*. The ectodomains of these receptors showed high identity or similarity with the reference sequences, indicating structural and functional conservation. The candidate PRRs are phylogenetically related to reference PRRs (in Arabidopsis and rice) and those described in other plant species. All candidate receptors had their expression induced after the inoculation with *H. vastatrix*, since the first time of sampling at 6 hours post-inoculation (hpi). There was a significant increase at 24 hpi for most receptors evaluated and a suppression at 48 hpi. A total of 1311 non-redundant NLR loci were identified in *C. arabica*, 927 in *C. canephora* and 1079 in *C. eugenioides*, of which 809, 562 and 695 are complete loci, respectively. The NLR-annotator showed extremely high sensitivities and specificities (over 99%) for identifying NLR loci in coffee, besides to increasing the detection capability of putative NLRs in the studied genomes. The NLR loci in coffee are distributed among all chromosomes and are organized mostly in clusters. The *C. arabica* genome present a smaller number of NLR loci when compared to the sum of the parental genomes (*C. canephora* and *C. eugenioides*). There are orthologous NLRs (orthogroups) shared between coffee, tomato, potato and reference NLRs and those that are shared only between coffee species. Phylogenetic analysis demonstrated orthologs NLRs shared between *C. arabica* and the parental genomes and those that were possibly lost. The NLR family members in coffee are subdivided into two main groups: TIR-NLR (TNL) and non-TNL. The Non-TNLs seem to represent an important repertoire of resistance genes in coffee. These results can support functional studies of PRRs and NLRs and contribute to the use of these receptors in the coffee breeding, aiming at the development of resistant cultivars.

Keywords: Broad-spectrum resistance. Pattern Recognition Receptor. Coffee rust. Coffea. NBS-LRR Resistance Genes.

SUMÁRIO

PRIMEIRA PARTE

1 INTRODUÇÃO GERAL

O café pertence ao gênero *Coffea*, o qual é constituído por mais de 100 espécies botânicas (DAVIS *et al.*, 2006), entretanto as espécies mais cultivadas são *Coffea canephora e Coffea arabica*. *C. canephora* é diploide (2n = 2x = 22 cromossomos) enquanto *C. arabica* é tetraploide (2n = 4x = 44cromossomos), originada da hibridização natural entre *C. canephora* e *C. eugenidoides* (BAWIN *et al.*, 2020; DENOEUD *et al.*, 2014). O café é uma bebida apreciada em todo o mundo, o que movimenta economias locais e o mercado internacional. Para 2020/2021 o consumo mundial de café é estimado em 167 milhões de sacas de 60kg, o que representa um aumento de 1,9%, mesmo com as restrições impostas pela pandemia do Covid-19. Neste mesmo período, a estimativa é que o volume total de produção aumente 0,4%, passando de 168,94 milhões de sacas para 169,60 milhões (ICO, 2021).

Dentre os países produtores de café, o Brasil se destaca como maior produtor mundial, sendo também um dos maiores consumidores. Em 2020, ano de bienalidade positiva e safra recorde, a produção brasileira foi de aproximadamente 63,08 milhões de sacas (48,77 de *C. arabica* e 14,31 de *C. canephora*) e o consumo interno foi estimado em 21,2 milhões de sacas (ABIC, 2020; CONAB, 2020). Entre os meses de janeiro a abril de 2020, a exportação de café solúvel também teve um desempenho positivo no país. Oitenta e sete países compraram café solúvel do Brasil, o equivalente a 1,3 milhão de sacas, aumento de 7,3% em relação ao mesmo período do ano de 2019 (EMBRAPA, 2020). Neste mesmo ano, a receita bruta das lavouras de café brasileiras ficou em torno de R$25,00 bilhões, sendo que só o estado de Minas Gerais contribuiu com R$15,43 bilhões (EMBRAPA, 2020a). Neste estado foram produzidas aproximadamente 35 mil sacas de café beneficiadas, com um aumento de 41,1% em relação a 2019 (CONAB, 2020). Em 2021, ano de bienalidade negativa, as exportações seguiram em alta. Apesar da queda de produtividade, os preços internacionais atrativos mantiveram o estímulo às vendas externas. Entre janeiro e abril deste ano, o Brasil já exportou para mais de 120 países, o que corresponde a 15,8 milhões de sacas vendidas e receita de US$ 2 bilhões (CONAB, 2021).

Os dados expostos mostram que o café é uma das principais *commodities* mundiais, apresentando grande importância para o Brasil, especialmente para o estado de Minas Gerais. Entretanto, os custos de produção são afetados devido a ocorrência de várias doenças. Entre os exemplos de doenças frequentemente associadas à cultura cafeeira encontram-se a

cercosporiose, mancha aureolada, antracnose e a ferrugem alaranjada. Além disso, a *Coffee Berry Disease* (CBD), afeta severamente plantações no Quênia (GIMASE *et al.*, 2020) e é uma doença potencial para os cafés brasileiros. Esta doença tem sido alvo de estudos preventivos para evitar devastações das lavouras, assim como ocorreu com a chegada da ferrugem, que era inexistente no Brasil até a década de 70 (ALKIMIM *et al.*, 2017; MCCOOK; VANDERMEER, 2015; ZAMBOLIM; CAIXETA, 2021).

A ferrugem alaranjada, causada pelo agente etiológico *Hemileia vastatrix* é uma das principais doenças e que pode levar a uma queda na produtividade de até 50%, principalmente em *C. arabica,* (ZAMBOLIM, 2016). A infecção deste patógeno inicia-se na parte abaxial da folha com a adesão e germinação dos urediniósporos, alongamento do tubo germinativo até o estômato e formação do apressório. A hifa produzida pelo apressório se diferencia e forma a célula-mãe do haustório e haustórios primários nas células subsidiárias e adjacentes do estômato, onde, posteriormente, se diferenciam em uma vesícula em formato de âncora na câmara subestomática. Haustórios secundários são produzidos a partir dessa vesícula nas células do mesófilo em genótipos suscetíveis (SILVA *et al.*, 2002). Na face abaxial da folha infectada, a ferrugem é inicialmente caracterizada por pequenas manchas cloróticas, translúcidas de cor amarelo pálido e que em pouco tempo se expandem. Nesta região formam-se os urediniósporos, um caraterístico "pó" de cor laranja-amarelado, produzidos repetidamente durante o ciclo da cultura e que são os responsáveis pelo início e manutenção da doença nas lavouras (AVELINO *et al.*, 2015; RAMIRO *et al.*, 2009; ZAMBOLIM, 2016). O fenótipo resultante é a desfolha, em consequência de lesões foliares que afetam a fotossíntese, secagem dos ramos ou morte prematura dos ramos com frutos, antes da colheita, o que consequentemente leva a redução da produtividade (SILVA *et al.*, 2006; TALHINHAS *et al.*, 2017).

Para que a produção cafeeira continue avançando no Brasil e no mundo, é de grande importância o manejo adequado desta cultura, buscando estratégias de curto a longo prazo e que permitam lidar com problemas fitossanitários atuais e os que podem ocorrer futuramente. Uma maneira de construir estratégias eficientes é compreender os mecanismos vegetais de reconhecimento dos fitopatógenos. Neste sentido, a identificação e estudo dos receptores que desencadeiam a sinalização para o mecanismo de defesa em plantas, seja ele de amplo espectro ou específico, vêm sendo alvo de pesquisas em diversas culturas (ALMEIDA *et al.*, 2020; BARKA *et al.*, 2020; CHEN *et al.*, 2020; JUPE *et al.*, 2012; LOZANO *et al.*, 2015; WANG *et al.*, 2021).

A percepção dos fitopatógenos pelas plantas é didaticamente dividida em duas etapas de reconhecimento e sinalização (JONES; DANGL, 2006). A primeira etapa baseia-se no

reconhecimento de moléculas microbianas conservadas, denominadas de padrões moleculares associados a patógenos (PAMPs). Exemplos de PAMPS são a flagelina ou fator de elongação (EF-Tu) em bactérias, além da quitina presente na parede celular de fungos (FURUKAWA *et al.*, 2014; GÓMEZ-GÓMEZ; BOLLER, 2000; KUNZE, 2004; LIU *et al.*, 2012). Os PAMPs são reconhecidos por receptores de reconhecimento de padrões (PRRs) que ativam a imunidade disparada por PAMP (PTI- PAMP Triggered immunity) (COUTO; ZIPFEL, 2016). Estes receptores são proteínas de membranas e estão envolvidos em um mecanismo de defesa de amplo espectro, pois reconhecem grupos de organismos, não diferenciando espécies (RANF, 2017).

Os patógenos adaptados podem inibir essa primeira linha de defesa por meio da secreção de efetores específicos, codificados pelos genes de avirulência (avr), e que levam a suscetibilidade disparada por efetores (ETS - effector-triggered susceptibility). Em resposta a esta supressão, as proteínas de resistência vegetais, codificadas pelos genes de resistência (R), reconhecem direta ou indiretamente estes efetores, disparando assim a imunidade desencadeada por efetores (ETI- Effector - Triggered immunity) (BOYD *et al.*, 2013; JONES; DANGL, 2006; KOURELIS; VAN DER HOORN, 2018). Os genes R, em sua maioria, codificam proteínas intracelulares que pertencem à família NBS-LRR ou NLR (nucleotide binding site-leucine-rich repeat ou NOD-like receptors) e ativam um mecanismos de reconhecimento espécie-específica, geralmente mais forte e que caracteriza a segunda etapa de percepção (BENTHAM *et al.*, 2017; DANGL; JONES, 2001). Em um contexto evolutivo, isolados de patógenos são selecionados para perder ou ganhar novos efetores que suprimem a ETI. Em resposta, a seleção também favorece novos alelos de genes R que podem reconhecer efetores recém-adquiridos, resultando novamente em ETI (JONES; DANGL, 2006).

Entende-se que essa divisão tem um caráter didático para representar o tipo de moléculas reconhecidas e os tipos de receptores vegetais, além do modelo evolutivo cíclico em que planta e patógeno estão envolvidos. Isso é constatado com dados atuais que demonstram que ambos os tipos de reconhecimento ocorrem de forma dinâmica e contínua, convergindo em vias de sinalização para a resposta de defesa. Atualmente já se sabe que falhas na sinalização em PTI podem comprometer a robustez da ETI, especialmente na resposta de hipersensibilidade (HR), produção de ROS (espécies reativas de oxigênio) e ativação da cascata MAPK (mitogen-activated protein kinase). A PTI também co-regula múltiplas respostas em ETI, atuando de diferentes formas, a depender do tipo de NLR ativado (LU; TSUDA, 2021; YUAN *et al.*, 2021, 2021a). Descobertas recentes sugerem que a sinalização de ETI retroalimenta as respostas de PTI, formando um *continuum* que é necessário para uma resposta de defesa eficiente (NGOU

et al., 2021; THOMMA; NÜRNBERGER; JOOSTEN, 2011). Os novos dados também demonstram que ETI não desencadeia uma via imunológica separada, mas sim uma amplificação que depende da maquinaria de PTI para funcionar efetivamente (YUAN *et al.*, 2021a). Desta forma, estratégias que buscam a resistência a doenças com foco no aumento da capacidade do próprio sistema de defesa vegetal, agregando conhecimentos das respostas PTI e ETI parecem promissoras para obtenção de cultivares resistentes.

Um aspecto relevante da caracterização dos PRR entre as espécies vegetais é permitir a reengenharia do reconhecimento de PAMPs, o que pode ser um aspecto a ser explorado para o melhoramento genético (BENT; MACKEY, 2007; BOUTROT; ZIPFEL, 2017; LEE; WHITAKER; HUTTON, 2016). Além disso, o aumento da resistência de amplo espectro conferida por estes receptores, pode reduzir o impacto dos fitopatógenos, permitindo potencialmente uma resistência mais duradoura e sustentável no campo. As proteínas NLRs, por sua vez, ativam um segundo pico de resposta mais eficiente. A HR, por exemplo, é uma resposta importante na interação específica ativada por NLRs, especialmente contra patógenos biotróficos, como a ferrugem do café. Esta resposta caracteriza-se pela morte celular rápida e localizada, evitando que o patógeno se espalhe e colonize o tecido, além de poder desencadear uma resistência Sistêmica Adquirida (SAR) (LACOMBE *et al.*, 2010; ZIPFEL *et al.*, 2006).

Muitas ferramentas de bioinformática têm sido desenvolvidas ou aperfeiçoadas com o objetivo de identificar e caracterizar de forma mais precisa os receptores envolvidos na ativação da resposta de defesa em plantas. Estudos de expressão gênica em resposta a inoculação de fitopatógenos também auxiliam nesta caracterização (FERNANDEZ-GUTIERREZ; GUTIERREZ-GONZALEZ, 2021; REICHEL *et al.*, 2021; ZHOU *et al.*, 2018). *Pipelines* de bioinformática específicos para anotação de NLRs, por exemplo, estão sendo desenvolvidos com o objetivo de aumentar a capacidade de detecção de genes R desta família em genomas de plantas (KUSHWAHA *et al.*, 2016; STEUERNAGEL *et al.*, 2020; TODA *et al.*, 2020). Além disso, abordagens que utilizam o Modelo de Markov Oculto (Hidden Markov Models - HMM) para análise de domínios e motivos conservados por meio de bancos de dados como o Pfam (The protein families database), SMART (Simple Modular Architecture Research Tool) ou o CDD (NCBI Conserved Domain Database) auxiliam na caracterização das proteínas codificadas por estes receptores (CHEN *et al.*, 2021; INTURRISI *et al.*, 2020; ZHOU *et al.*, 2018). Agrupamentos de genes ortólogos, análises filogenéticas ou de BLAST (Basic Local Alignment Search Tool) também são úteis para classificar e indicar possíveis funções de genes R ou PRRs identificados nos genomas alvos (SEO *et al.*, 2016; TOMBULOGLU *et al.*, 2019).

No melhoramento genético do cafeeiro, até então, não se fez uso direcionado da resistência condicionada por PRRs. Uma das limitações para o uso destes receptores é a ausência de informações sobre o repertório de PRR no café, além da modulação da expressão gênica destes receptores em resposta a diferentes patógenos. Adicionalmente, estudos de identificação em todo o genoma e caracterização dos genes R da família NLR é escasso nesta cultura, sobretudo em *C. arabica*. Desta forma, o presente estudo teve como objetivos identificar receptores PRR no genoma de *C. arabica* e caracterizá-lo quanto a expressão gênica no patossistema ferrugem-cafeeiro. Além disso, identificar e caracterizar genes R da família NLR (ou NBS-LRR) nos genomas de *C. arabica*, *C. canephora* e *C. eugeniodes*, e por fim, de posse dos resultados, obter subsídios para a utilização de PRRs e NLRs nos programas de melhoramento do café.

REFERÊNCIAS

ABIC. **Evolução do consumo interno de café no Brasil**. 2020. Disponível em: <https://www.abic.com.br/estatisticas/indicadores-da-industria/>. Acesso em: 23 jun. 2021.

ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; PEREIRA, A. A.; OLIVEIRA, A. C. B. DE; ZAMBOLIM, L.; SAKIYAMA, N. S. Marker-assisted selection provides arabica coffee with genes from other Coffea species targeting on multiple resistance to rust and coffee berry disease. **Molecular Breeding**, v. 37, n. 1, p. 6, 2017.

ALMEIDA, D. P. DE; CASTRO, I. S. L.; MENDES, T. A. DE O.; ALVES, D. R.; BARKA, G. D.; BARREIROS, P. R. R. M.; ZAMBOLIM, L.; SAKIYAMA, N. S.; CAIXETA, E. T. Receptor-Like Kinase (RLK) as a candidate gene conferring resistance to *Hemileia vastatrix* in coffee. **Scientia Agricola**, v. 78, n. 6, p. 1–9, 2020.

AVELINO, J.; CRISTANCHO, M.; GEORGIOU, S.; IMBACH, P.; AGUILAR, L.; BORNEMANN, G.; LÄDERACH, P.; ANZUETO, F.; HRUSKA, A. J.; MORALES, C. The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. **Food Security**, v. 7, n. 2, p. 303–321, 2015.

BARKA, G. D.; CAIXETA, E. T.; FERREIRA, S. S.; ZAMBOLIM, L. In silico guided structural and functional analysis of genes with potential involvement in resistance to coffee leaf rust: A functional marker based approach. **PLoS ONE**, v. 15, n. 7, p. 1–22, 2020.

BAWIN, Y.; RUTTINK, T.; STAELENS, A.; HAEGEMAN, A.; STOFFELEN, P.; MWANGA MWANGA, J. C. I.; ROLDÁN-RUIZ, I.; HONNAY, O.; JANSSENS, S. B. Phylogenomic analysis clarifies the evolutionary origin of Coffea arabica. **Journal of Systematics and Evolution**, 2020.

BENT, A. F.; MACKEY, D. Elicitors, Effectors, and R Genes: The New Paradigm and a Lifetime Supply of Questions. **Annual Review of Phytopathology**, v. 45, n. 1, p. 399–436, 2007.

BENTHAM, A.; BURDETT, H.; ANDERSON, P. A.; WILLIAMS, S. J.; KOBE, B. Animal NLRs provide structural insights into plant NLR function. **Annals of Botany**, v. 119, n. 5, p. 689–702, 2017.

BOUTROT, F.; ZIPFEL, C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. **Annual Review of Phytopathology**, v. 55, n. 1, p. 257–286, 2017.

BOYD, L. A.; RIDOUT, C.; O'SULLIVAN, D. M.; LEACH, J. E.; LEUNG, H. Plant-pathogen interactions: Disease resistance in modern agriculture. **Trends in Genetics**, v. 29, n. 4, p. 233–240, 2013.

CHEN, Q.; DONG, C.; SUN, X.; ZHANG, Y.; DAI, H.; BAI, S. Overexpression of an apple LysM-containing protein gene, MdCERK1-2, confers improved resistance to the pathogenic fungus, *Alternaria alternata*, in Nicotiana benthamiana. **BMC Plant Biology**, v. 20, n. 1, p. 1–13, 2020.

CHEN, Z.; SHEN, Z.; ZHAO, D.; XU, L.; ZHANG, L.; ZOU, Q. Genome-wide analysis of lysm-containing gene family in wheat: Structural and phylogenetic analysis during development and defense. **Genes**, v. 12, n. 1, p. 1–18, 2021.

CONAB. **Safra Brasileira de Café - Boletim Café dezembro 2020**. 2020. Disponível em: <https://www.conab.gov.br/info-agro/safras/cafe>. Acesso em: 19 jun. 2021.

CONAB. **Safra Brasileira de Café - Boletim de Café Maio de 2021**. 2021. Disponível em: <https://www.conab.gov.br/info-agro/safras/cafe>. Acesso em: 22 jun. 2021.

COUTO, D.; ZIPFEL, C. Regulation of pattern recognition receptor signalling in plants. **Nature Reviews Immunology**, v. 16, n. 9, p. 537–552, 2016.

DANGL, J. L.; JONES, J. D. G. Plant pathogens and integrated defence responses to infection. **Nature**, v. 411, n. 6839, p. 826–833, 2001.

DAVIS, A. P.; GOVAERTS, R.; BRIDSON, D. M.; STOFFELEN, P. An annotated taxonomic of the genus coffea (Rubiaceae). **Botanical Journal of the Linnean Society**, v. 152, n. 4, p. 465–512, 2006.

DENOEUD, F. *et al.* The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. **Science**, v. 345, n. 6201, p. 1181–1184, 2014.

EMBRAPA. **Exportações de café solúvel do Brasil mantêm crescimento apesar da pandemia de COVID-19**. 2020. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/53048005/exportacoes-de-cafe-soluvel-do-brasil-mantem-crescimento-apesar-da-pandemia-de-covid-19>. Acesso em: 15 jun. 2021.

EMBRAPA. **Receita bruta da lavoura dos Cafés do Brasil prevista para 2020 tem aumento de 25% em comparação com 2019**. 2020a. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/50684374/receita-bruta-da-lavoura-dos-cafes-do-brasil-prevista-para-2020-tem-aumento-de-25-em-comparacao-com-2019>. Acesso em: 10 jun. 2021.

FERNANDEZ-GUTIERREZ, A.; GUTIERREZ-GONZALEZ, J. J. Bioinformatic-Based Approaches for Disease-Resistance Gene Discovery in Plants. **Agronomy**, v. 11, n. 11, p. 2259, 2021.

FURUKAWA, T.; INAGAKI, H.; TAKAI, R.; HIRAI, H.; CHE, F.-S. Two Distinct EF-Tu Epitopes Induce Immune Responses in Rice and *Arabidopsis*. **Molecular Plant-Microbe Interactions**, v. 27, n. 2, p. 113–124, 2014.

GIMASE, J. M.; THAGANA, W. M.; OMONDI, C. O.; CHESEREK, J. J.; GICHIMU, B. M.; GICHURU, E. K.; ZIYOMO, C.; SNELLER, C. H. Genome-Wide Association Study identify the genetic loci conferring resistance to Coffee Berry Disease (*Colletotrichum kahawae*) in *Coffea arabica* var. Rume Sudan. **Euphytica**, v. 216, n. 6, 2020.

GÓMEZ-GÓMEZ, L.; BOLLER, T. FLS2: An LRR Receptor-like Kinase Involved in the Perception of the Bacterial Elicitor Flagellin in Arabidopsis. **Molecular Cell**, v. 5, n. 6, p. 1003–1011, 2000.

ICO. **International Coffee Organization - What's New**. 2021. Disponível em: <https://www.ico.org/>. Acesso em: 22 jun. 2021.

INTURRISI, F.; BAYER, P. E.; YANG, H.; TIRNAZ, S.; EDWARDS, D.; BATLEY, J. Genome-wide identification and comparative analysis of resistance genes in *Brassica juncea*. **Molecular Breeding**, v. 40, n. 8, p. 1–14, 2020.

JONES, J. D. G.; DANGL, J. L. The plant immune system. **Nature**, v. 444, n. 7117, p. 323–329, 2006.

JUPE, F.; PRITCHARD, L.; ETHERINGTON, G. J.; MACKENZIE, K.; COCK, P. J. A.; WRIGHT, F.; SHARMA, S. K.; BOLSER, D.; BRYAN, G. J.; JONES, J. D. G.; HEIN, I. Identification and localisation of the NB-LRR gene family within the potato genome. **BMC Genomics**, v. 13, n. 1, p. 1–14, 2012.

KOURELIS, J.; HOORN, R. A. VAN DER. Defended to the Nines: 25 years of Resistance Gene Cloning Identifies Nine Mechanisms for R Protein Function. **The Plant Cell**, v. 30, n. 2, p. 285–299, 2018.

KUNZE, G. The N Terminus of Bacterial Elongation Factor Tu Elicits Innate Immunity in Arabidopsis Plants. **the Plant Cell Online**, v. 16, n. 12, p. 3496–3507, 2004.

KUSHWAHA, S. K.; CHAUHAN, P.; HEDLUND, K.; AHREN, D. NBSPred: A support vector machine-based high-throughput pipeline for plant resistance protein NBSLRR prediction. **Bioinformatics**, v. 32, n. 8, p. 1223–1225, 2016.

LACOMBE, S. *et al.* Interfamily transfer of a plant pattern-recognition receptor confers broad-spectrum bacterial resistance. **Nature Biotechnology**, v. 28, n. 4, p. 365–369, 2010.

LEE, S.; WHITAKER, V. M.; HUTTON, S. F. Potential applications of non-host resistance for crop improvement. **Frontiers in Plant Science**, v. 7, p. 1–6, 2016.

LIU, T. *et al.* Chitin-induced dimerization activates a plant immune receptor. **Science**, v. 336, n. 6085, p. 1160–1164, 2012.

LOZANO, R.; HAMBLIN, M. T.; PROCHNIK, S.; JANNINK, J. L. Identification and distribution of the NBS-LRR gene family in the Cassava genome. **BMC Genomics**, v. 16, n. 1, p. 1–14, 2015.

LU, Y.; TSUDA, K. Intimate Association of PRR- and NLR-Mediated Signaling in Plant Immunity. **Molecular Plant-Microbe Interactions**, v. 34, n. 1, p. 3–14, 2021.

MCCOOK, S.; VANDERMEER, J. The Big Rust and the Red Queen: Long-Term Perspectives on Coffee Rust Research. **Phytopathology**, v. 105, n. 9, p. 1164–1173, 2015.

NGOU, B. P. M.; AHN, H. K.; DING, P.; JONES, J. D. G. Mutual potentiation of plant immunity by cell-surface and intracellular receptors. **Nature**, v. 592, n. 7852, p. 110–115, 2021.

RAMIRO, D. A.; ESCOUTE, J.; PETITOT, A. S.; NICOLE, M.; MALUF, M. P.; FERNANDEZ, D. Biphasic haustorial differentiation of coffee rust (*Hemileia vastatrix* race II) associated with defence responses in resistant and susceptible coffee cultivars. **Plant Pathology**, v. 58, n. 5, p. 944–955, 2009.

RANF, S. Sensing of molecular patterns through cell surface immune receptors. **Current Opinion in Plant Biology**, v. 38, p. 68–77, 2017.

REICHEL, T.; RESENDE, M. L. V.; MONTEIRO, A. C. A.; FREITAS, N. C.; BOTELHO, D. M. S. Constitutive Defense Strategy of Coffee Under Field Conditions : A Comparative Assessment of Resistant and Susceptible Cultivars to Rust. **Molecular Biotechnology**, p. 1–15, 2021.

SEO, E.; KIM, S.; YEOM, S. I.; CHOI, D. Genome-wide comparative analyses reveal the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among solanaceae plants. **Frontiers in Plant Science**, v. 7, p. 1205, 2016.

SILVA, M. C.; NICOLE, M.; GUERRA-GUIMARÃES, L.; RODRIGUES, C. J.

Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. **Physiological and Molecular Plant Pathology**, v. 60, n. 4, p. 169–183, 2002.

SILVA, M. D. C.; VÁRZEA, V.; GUERRA-GUIMARÃES, L.; AZINHEIRA, H. G.; FERNANDEZ, D.; PETITOT, A. S.; BERTRAND, B.; LASHERMES, P.; NICOLE, M. Coffee resistance to the main diseases: Leaf rust and coffee berry disease. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 119–147, 2006.

STEUERNAGEL, B. *et al.* The NLR-annotator tool enables annotation of the intracellular immune receptor repertoire. **Plant Physiology**, v. 183, n. 2, p. 468–482, 2020.

TALHINHAS, P. *et al.* The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. **Molecular Plant Pathology**, v. 18, n. 8, p. 1039–1051, 2017.

THOMMA, B. P. H. J.; NÜRNBERGER, T.; JOOSTEN, M. H. A. J. Of PAMPs and Effectors: The Blurred PTI-ETI Dichotomy. **The Plant Cell**, v. 23, n. 1, p. 4–15, 2011.

TODA, N.; RUSTENHOLZ, C.; BAUD, A.; PASLIER, M. C. LE; AMSELEM, J.; MERDINOGLU, D.; FAIVRE-RAMPANT, P. NLGenomeSweeper: A tool for genome-wide NBS-LRR resistance gene identification. **Genes**, v. 11, n. 3, p. 333, 2020.

TOMBULOGLU, G.; TOMBULOGLU, H.; CEVIK, E.; SABIT, H. Genome-wide identification of Lysin-Motif Receptor-Like Kinase (LysM-RLK) gene family in Brachypodium distachyon and docking analysis of chitin/LYK binding. **Physiological and Molecular Plant Pathology**, v. 106, p. 217–225, 2019.

WANG, Z.; XU, F.; REN, H.; LU, G.; QUE, Y.; XU, L. Genome-Wide Characterization of NLRs in Saccharum spontaneum L. and Their Responses to Leaf Blight in Saccharum. **Agronomy**, v. 11, n. 1, p. 153, 2021.

YUAN, M.; JIANG, Z.; BI, G.; NOMURA, K.; LIU, M.; HE, S. Y.; ZHOU, J. M.; XIN, X. F. Pattern-recognition receptors are required for NLR-mediated plant immunity. **Nature**, p. 1–5, 2021.

YUAN, M.; NGOU, B. P. M.; DING, P.; XIN, X. F. PTI-ETI crosstalk: an integrative view of plant immunity. **Current Opinion in Plant Biology**, v. 62, p. 102030, 2021a.

ZAMBOLIM, L. Current status and management of coffee leaf rust in Brazil. **Tropical Plant Pathology**, v. 41, n. 1, p. 1–8, 2016.

ZAMBOLIM, L.; CAIXETA, E. T. An overview of physiological specialization ofcoffee leaf rust – new designation of Pathotypes. **International Journal of Current Research**, v. 13, n. 1, p. 15564–15575, 2021.

ZHOU, Z.; TIAN, Y.; CONG, P.; ZHU, Y. Functional characterization of an apple (Malus x domestica) LysM domain receptor encoding gene for its role in defense response. **Plant Science**, v. 269, p. 56–65, 2018.

ZIPFEL, C.; KUNZE, G.; CHINCHILLA, D.; CANIARD, A.; JONES, J. D. G.; BOLLER, T.; FELIX, G. Perception of the Bacterial PAMP EF-Tu by the Receptor EFR Restricts Agrobacterium-Mediated Transformation. **Cell**, v. 125, n. 4, p. 749–760, 2006.

SEGUNDA PARTE – ARTIGOS

ARTIGO 1

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

(Artigo redigido de acordo com as normas e submetido ao periódico PLOS ONE)

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)

Abstract

Pathogen-associated molecular patterns (PAMPs) are recognized by pattern recognition receptors (PRRs) localized on the host plasma membrane. 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, called 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 [6,11] 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 [12,13]. In addition, it leads to a durable and broad-spectrum resistance [14,15]. 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 [16].

Given that the PRRs are involved in a broad-spectrum and durable defense, currently they have been the target of studies aiming at a greater use in plant breeding [15,17]. 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 [17,18] 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) [19]. For fungi, well-described receptors are those that recognize chitin and have in common extracellular domains with lysin residues (Lys) [4,20], such as *CERK1* (chitin elicitor receptor kinase 1) [21], *CEBiP* (chitin elicitor binding protein) [22], *LYK4*, *LYK5* (LysM-containing receptor-like kinase 4 and 5) [23,24], *LYP4* and *LYP6* (LysM domain-containing protein 4 and 6) [25].

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* [20,21] and mutations in *At-LYK5* [23] 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 [14]. 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 [14,26].

Few non-model plants, such as barley [27], apple [28,29] and mulberry [30], had PRRs characterized. *Coffea arabica* is an important coffee species cultivated in countries such as Brazil, Vietnam, Colombia, and Indonesia and is consumed around the world [31]. 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 [32,33]. In Brazil, the biotrophic fungus *Hemileia vastatrix* Berk. & Br, the etiological agent of coffee rust, has caused damage since the 1970s [34,35]. In regions with favorable conditions for the pathogen, the decline in productivity can reach 50% [35]. 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 [32,36,37]. 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 [38] 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/) [39].

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 [40].

Considering the fact that *C. arabica* is an allotetraploid (2n = 4x = 44 chromosomes), originated from natural hybridization between *C. canephora* and *C. eugenioides* [41,42], the sequences selected as PRR candidates for the arabica coffee (variety Geisha from Phytozome) were also analyzed by BLASTp in the database of the NCBI (https://www.ncbi.nlm.nih.gov/) against the genome of *C. arabica*, Red Caturra cultivar (Cara_1.0, GenBank assembly accession: GCA_003713225.1). This genome was deposited after the beginning of this study and presents the scaffolds anchored to the chromosomes of each ancestral subgenomes. 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, added in sterile distilled water and the

suspension was calibrated at 1 x 10^6 urediniospores/ml. The viability of inoculum was verified by observing the spore germination in glass cavity slides. After preparing the suspension 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 [43]. 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

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

For each of the three biological samples, technical triplicates were used and for each plate an inter-assay sample was used to ensure the reproducibility of the technique. The relative quantification was calculated according to the formula by Pfaffl, 2001 [44]. Referring to the

data normalization, the expression stability of four reference genes was analyzed: protein 14-3-3 (*14-3-3*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), ribosomal protein 24S (*24S*) and factor elongation 1α (*EF1-α*) [45–48]. The efficiency correction of these genes in Cq values was performed by the GenEx Enterprise program (version 7.0) and the stability was verified by the RefFinder tool [49]. The two most stable genes were *14-3-3* and *GAPDH* (S3 Fig), which were used to normalize the transcription levels of the target genes. The samples with the lowest expression were used as calibrators. The MN 48 hpi was used as calibrator sample, except for the *Ca1-CERK1* (experiment 2), which was used the IP 48 hpi sample. The PCR amplification efficiencies and linear regression coefficients were determined using the LinRegPCR software version 2018.0 (Table 2) [50]. The average expression was obtained by the ratio of the sample inoculated with *H. vastatrix* compared to the average of the control treatment (without inoculation).

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

Gene	Target sequence	Primer	Amplicon length (bp)	Amplification efficiency	R^2
LYK4	612.376 and 952.320-1 (***Ca-LYK4***)	AAAGGCCACAAACAGATGCGACAG (F)	168	Exp1 - 1,855	0,929
		AGGTGGGATGGATCAGCTGCTAAG (R)		Exp2 - 1,866	0,961
LYK5	628.522 (***Ca1-LYK5***)	TTTGGTTCCTGCGGTATAGG (F)	112	Exp1 - 2,056	0,974
		TCTGGCAAAGCCCTGTAAAC (R)		Exp2 - 2,095	0,988
	1841.91 (***Ca2-LYK5***)	TTGCAGCATGCCACAGGTTCTTTC (F)	237	Exp1 - 1,920	0,961
		ATCACTCAGGCCACCTTTCTCTGC (R)		Exp2 - 1,898	0,952
CERK1	1805.113 and 539.592 (***Ca1-CERK***)	CGAGACATTAAGCCAGCTAAC (F)	139	Exp1 - 1,881	0,990
		GCATGTAACCGAAAGTACCC (R)		Exp2 - 1,887	0,965
	2193.164 and 476.38 (***Ca2-CERK***)	CAGTTCCAGTTAGCTGCTCCA (F)	83	Exp1 - 1,899	0,999
		GGAGAAGTTCCTTCAGCAACAC (R)		Exp2 - 1,885	0,992
***LYP* (*CEBiP*-like)**	439.212 (***Ca1-LYP***)	ACCACCGCCGATGTTCTGTTGC (F)	82	Exp1 - 1,898	0,992
		GAGGAACATCGAGAATAGCGCCGG (R)		Exp2 - 1,887	0,994
	1196.90 (***Ca2-LYP***)	TCCAGACCCTCTTCAACGTC (F)	121	Exp1 - 1,824	0,983
		CAGGCGAAAGGAATCTTGAG (R)		Exp2 - 1,829	0,997
14-3-3	SGN-U347734	TGTGCTCTTTAGCTTCCAAACG (F)	75	Exp1 - 1,983	0,943
		CTTCACGAGACATATTGTCTTACTCAAA (R)		Exp2 - 2,001	0,933

GAPDH	SGN-U356404	TTGAAGGGCGGTGCAAA (F)	59	Exp1- 2,007	0,993
		AACATGGGTGCATCCTTGCT (R)		Exp2 - 2,060	0,995
24S	GR986263.1	ACGGCATCAAAGGAGACAAT (F)	114	Exp1 - 1,893	0,998
		ATGCAGAACATCGATCACGA(R)		Exp2 - 1,902	0,994
EF1-α	GW466696.1	CTCTCTCGCCTCCTGTCTTC (F)	105	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 [45] and *24S* and *EF1-α* from Reichel 2021 [48]. 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, $\cap N(0, \sigma_e^2)$. Checks for outliers and of the assumptions of residuals from models were accomplished using diagnostic plots within the R software [51].

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 [51].

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 ≤ $10^{-5,}$ 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*) described in Arabidopsis and rice were included in a new analysis: *At-LYP1* (*At-CEBiP / LYM2*), *At-LYP2* (*LYM1*), *At-LYP3* (*LYM3*), *Os-LYP4* and *Os-LYP6* (Table 1).

Fig 1. Phylogenetic analysis of the selected sequences for *C. arabica* by comparison with the reference PRRs. (A) *CERK1*, (B) *LYK4*, (C) *CEBiP*, (D) *LYK5*, (E) *CEBiP* and reference proteins belonging to the *LYP* (*CEBiP*-like) group. The phylogenetic trees were constructed with complete amino acid sequence alignments using the Maximum Likelihood method with a bootstrap of 1000 replications. The cluster clade of candidate sequences for *C. arabica* and reference sequences are highlighted in blue.

The new phylogenetic analysis for *CEBiP* (Fig 1E) showed two distinct clades. The clade one formed by the sequences Scaffold 506.17 and 1856.2, *At-LYP2*, *At-LYP3*, *Os-LYP4* and *Os-LYP6*, and the clade two formed by *Os-CEBiP*, *At-LYP1*, Scaffold 439.212 and 1196.90.

As the Scaffold sequences 439.212 and 1196.90 showed greater similarity with the *Os-CEBiP* homologue in *A. thaliana* (*At-LYP1),* they were selected as candidate sequences for the *CEBiP-like* (Fig 1 and S1 Table). Moreover, the *At-LYP2* (*LYM1*) and *At-LYP3* (*LYM3*), belonging to clade one, are described in the literature for their ability to recognize the peptidoglycan, a bacterial PAMP [52]. These sequences formed the nearest clade to the Scaffold 506.17 and 1856.2 sequences, substantiating the choice of the two *C. arabica* sequences belonging to clade two. The *Os-LYP4* and *Os-LYP6* that play a dual role, recognizing peptidoglycan and chitin [25], were not evaluated in this study.

All the domains found in the coffee candidate sequences correspond to the characteristic 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* [53,54], eight ones displayed identity or similarity with the candidate sequences in *C. arabica*. For *Os-CEBiP*, from nine described [55], only three were present. In *At-LYK5*, only one of three described [23] 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 genome of *C. arabica*, *Caturra red cultivar*

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

The BLASTp analysis in the NCBI database against the genome of *C. arabica* (Red Caturra cultivar) showed that six candidate sequences for PRRs in *C. arabica* (variety Geisha) have greater percentage of identity with sequences belonging to the *C. eugenioides* subgenome and four showing greater identity with the *C. canephora* subgenome (Table 4). This analysis allowed us to identify that while each of the candidate sequences for *LYK5* and *LYP*, in addition to the two sets of sequences for *CERK1* (considering subclades I and II, Fig 1), had greater identity with sequences from each of the subgenomes, both candidate sequences for *LYK4* had greater identity with a sequence in the *C. eugenioides* subgenome.

Table 4. BLASTp analysis of candidate sequences in *C. arabica* Geisha) against *C. arabica* (Red Caturra).

C. arabica **(Phytozome - Variety Geisha)**		***C. arabica*** **(NCBI - Red Caturra cultivar)**			
Candidate sequence	**Query Cover**	**E-value**	**Identity (%)**	**ID***	**Chr**
CERK1-Scaffold_539.592	98%	0.0	97.73%	XP_027086837.1	9c
CERK1-Scaffold_1805.113	100%	0.0	97.73%	XP_027086837.1	9c
CERK1-Scaffold_2193.164	99%	0.0	100.00%	XP_027061585.1	5e
CERK1-Scaffold_476.38	100%	0.0	96.93%	XP_027061585.1	5e
LYK4-Scaffold_612.376	99%	0.0	97.52%	XP_027077444.1	7e
LYK4-Scaffold_952.320	99%	0.0	100.00%	XP_027077444.1	7e
LYK5-Scaffold 628.522	99%	0.0	100.00%	XP_027092883.1	10c
LYK5-Saffold 1841.91	99%	0.0	99.85%	XP_027090781.1	10e
LYP-Scaffold_439.212	94%	0.0	79.37%	XP_027089306.1	9e
LYP-Scaffold_1196.90	99%	0.0	100.00%	XP_027087432.1	9c

*GenBank National Center for Biotechnology Information (NCBI) sequence database. Chr: chromosome corresponding to the subgenomes of *C. arabica*, being the subgenome of *C. canephora* represented by the letter c and the subgenome of *C. eugenioides* represented by the letter e.

Primer design

The four sequences selected as candidates for *CERK1* in the *C. arabica* genome by phylogenetic analysis formed two distinct subclades (Fig 1A). The subclade I formed by the Scaffold 539.592 and Scaffold 1805.113 sequences and the subclade II formed by the Scaffold 2193.164 and Scaffold 476.38 sequences. The coding sequences (CDS) of subclade I showed an 71.33% identity, with the 1805.113 sequence presenting a smaller CDS (1815bp) and shared almost entirely with the Scaffold 539.592 sequence. The Scaffold 539.592 sequence, on the other hand, presents a larger CDS (2511 bp) with two regions that are not present in 1805.113

(S4 Fig). The Scaffold 2193.164 and 476.38 showed CDS of the same size (1860bp) and an identity 98.28% (S5 Fig). For the primer design in the gene expression analysis, the formation of these two subclades was considered, thus using a pair of primers for each of the formed subclades. They were named *Ca1-CERK1* and *Ca2-CERK1* respectively and are referred to as such in the gene expression analysis (Table 2).

Concerning the *LYK4* candidate sequences (Scaffold 612.376 and 952.320), a primer pair was also designed for both candidate sequences. These showed a 98.45% identity (S6 Fig) and were named as *Ca-LYK4*. Regarding the candidate sequences *LYK5* (Scaffold 628.522 and Scaffold 1841.91) and *LYP* (*CEBiP*-Like) (Scaffold 439,212 and 1196.90), a primer pair was designed for each sequence separately and they are referred to as *Ca1-LYK5*, *Ca2-LYK5*, *Ca1-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*. For 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. No significant difference was observed for 6 and 48 hpi. 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 *Ca2-LYP* and *Ca1-LYK5*, 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 expression occurred at 6 and 24 hpi, with no difference between these times. The analysis between cultivars showed that at 6 hpi, AR obtained the highest expression while IP presented the lowest expression. On the other hand, at 24 and 48 hpi, there were no differences between cultivars. However, it was found that 48 hpi was the time with the lowest average observed, within and between cultivars.

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

464 times between cultivars, there were differences only in 24 hpi, with AR and IP having the
465 highest expression.

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

473 Discussion

474 Fungal PRRs in the *C. arabica* genome

475 Understanding basal immunity has been the focus of several studies with the purpose of
476 identifying the mechanisms governing this line of defense, enabling its use as another tool in
477 the search for plant resistance to pathogens [17]. The description of the reference PRRs and
478 studies of the modulation of their gene expression in response to *H. vastatrix*, one of the most
479 devastating pathogens in coffee trees, presents an advance for understanding this crop basal
480 immunity. In the present study, fungal PRR candidate sequences well described in the literature
481 for model plants such as Arabidopsis and rice were studied in *C. arabica*. We observed that
482 there is more than one candidate sequence for each receptor studied, which may be the result
483 of the ploidy of this species or duplication of these receptors, a common mechanism in plant
484 genomes [56].

485 Each of the candidate sequences for *LYK5* and *LYP* (*CEBiP*-like) presented higher
486 percentages of identity with one of the *C. arabica* subgenomes. Therefore, it is possible to infer
487 that those genes may have come from of each of the parental genomes (Table 4). Referring to

LYK4, both candidate sequences showed greater identity with *C. eugenioides* subgenome, which can indicate duplication events. For *CERK1*, as two sequences had a higher percentage of identity with a sequence from *C. canephora* subgenome (subclade I), while the other two (subclade II) with a sequence from *C. eugenioides* subgenome, we can suggest that both events occurred in this case. Besides to having a gene from each of the subgenomes, a duplication event of these genes may also have occurred in *C. arabica* (Variety Geisha). However, differences in the quality of *C. arabica* genomes (Geisha and Caturra red) can also interfere with this conclusion.

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 [53,57] 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 [24,57–60]. For LysM receptors homologous to *At-CERK1*, the CDS region mostly presents around 1800 bp with ten to twelve exons [28,53,61], which is likewise with the size of the CDS and number of exons found for the *CERK1* candidate sequences in coffee, except for the Scaffold 539.592, which presents a larger coding region, with 2511bp and 13 exons. However, this number of thirteen exons has also been found in *Ps-LYK9*, a *CERK1-like* gene in peas, which is involved in the control of plant immunity and symbiosis formation [61].

Regarding the genes LYPs (Receptor-like proteins or RLPs) such as *Os-CEBiP*, the number of exons reported is more variable from two to six [22,57,62]. In *C. arabica*, Scaffold 1196.90 and 439.212 presented four and five, respectively. The structural pattern of genes, such as the distribution of introns or exons in gene families, reinforces the ortholog identification

between sequences since these are almost conserved among all orthologous. Minor differences may be due to evolutionary changes or errors in gene structure predictions [58].

Characterization of domains and motifs (LysM)

Proteins with LysM domain classified as LYKs (Receptor-like kinases or RLKs) are composed of lysin motifs (LysM)-containing ectodomains, a transmembrane domain and an intracellular kinase. LYP proteins (RLPs), on the other hand, present LysM ectodomain, but without intracellular kinase and can be anchored to the plasma membrane by a transmembrane domain or GPI-anchor [57,63]. The *At-CERK1*, *At-LYK4* and *At-LYK5* contain three extracellular LysM motifs, a transmembrane domain and intracellular kinase, while *Os-CEBiP* has two extracellular LysM motifs and GPI anchor [21–23]. The SMART and PredGPI analysis predicted that the amino acid sequences of the PRRs studied in *C. arabica* present a signal peptide, extracellular LysM motifs, a transmembrane 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 [22,57,64]. 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 [56,57]. These regions define the type of recognized PAMP and its binding affinity in addition to the interaction between receptors and coreceptors [65]. 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 [21,23,66]. 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 [20,67]. Distinction in the role of these receptors suggests that plants use different chitin binding and signaling strategies [24,68].

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 [53,55]. For *CERK1*, eight residues reported as important for chitin binding in Arabidopsis are present in the Scaffold 2193.164 and Scaffold 476.38 sequences (six identical and two similar), suggesting that they can bind chitin. However, complementary data are still needed to clarify which would be the primary receptor and co-receptor of the innate immunity in this species, and further studies of chitin-receptor and receptor-receptor interaction are required.

Joint phylogenetic analysis

PRRs are conserved in several plant species [58].This conservation indicates a fundamental importance of the PAMP recognition system [25]. The joint phylogenetic analysis showed that the sequences selected as candidates for *CERK1* in coffee, were highly related to *Md-CERK1*, *Md-CERK1-2*, *Ps-LYK9*, *Mm-LYK2*, *Vv-LYK1-1*, *Vv-LYK1-2*, *Os-CERK1* and *At-CERK* (Fig 4). All of these proteins have been described as being involved in the defense against fungal pathogens [20,21,28–30,53,61], suggesting that the studied sequences also participate in the defense responses against this group of phytopathogens. Among the species compared, tomato and grape have greater evolutionary proximity to coffee. *Bti9* (*Sl-LYK1*), a *CERK1* homolog in tomato, which grouped more closely to the Scaffold 2193.164 and 476.38 sequences

(*Ca2-CERK1*) in this clade, presents an identity of 58.6% with *At-CERK* [69]. Candidate sequences in coffee, however, showed around 57% of identity (Table 3).

The *Bti9* (*Sl-LYK1*) in tomato interacts with *AvrPtoB*, effector in *Pseudomonas syringae*. The kinase region of this protein is the target and this results in blocking the PTI signaling [69]. Despite being described as a bacterial effector target, the study by Zeng et al., 2012 [69] or later reports by Xin and He, 2013 [70] did not describe the interaction of this protein with chitin or the transcriptional profiles regarding the response to fungal pathogens. Nonetheless, *Bti9* is a membrane receptor with extracellular LysM motifs and high homology to *At-CERK1*. 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 [52,71], 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 [53]. The clustering of *Bd-LYK4* in this clade corroborates the results presented by Tombuloglu et al., 2019 [57] 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 [30]. Additionally, in this clade, the *Hv-CEBiP* in barley, has been described for recognizing chitin oligosaccharides derived from *Magnaporthe oryzae* [27] and

586 *Mt-LYM2*, in *Medicago truncatula*, demonstrated specific binding to biotinylated N-
587 acetylchitooctaose in a similar way to *CEBiP* in rice [22,62]. Thus, the receptors cited for the
588 phylogenetic groupings of this study reinforces the possible role of candidate sequences in *C.*
589 *arabica* as PAMP receptors.

590 Transcriptional response of candidate receptors in *C. arabica*

591 The PAMPS are defined as highly conserved molecules from microorganisms and,
592 therefore, have an essential function in their survival or fitness [72,73]. It is suggested that since
593 PAMPs are essential for the viability or lifestyle of microorganisms, it is less likely that they
594 avoid host immunity through mutation or deletion in these regions [14,74]. Chitin is a PAMP
595 present in the fungal cell wall. Fragments of N-acetylquitooligosaccharides are released by the
596 breakdown of this PAMP by plant chitinases during plant-fungus interactions. These fragments
597 serve as elicitors for the innate immunity of plants by modifying the transcriptional levels of
598 PRRs [22].

599 In this study, the expression increases were detected from 6 hpi, showing that all
600 candidate PRR were stimulated after the inoculation of *H. vastatrix*. The highest averages of
601 expression were observed at 24 hpi, for most receptors, followed by a decrease at 48 hpi (Fig
602 5). These results describe an initial stimulus with subsequent suppression. The experiments
603 showed that at 24 hpi it is already possible to detect the penetration of the hypha produced by
604 the appressorium of *H. vastatrix* in stomata of coffee leaves, both in resistant and susceptible
605 genotypes and at 48 hpi the presence of haustoria is already observed [75–77]. In addition, a
606 LRR receptor-like kinase described in this pathosystem has a peak expression at 24 hpi in
607 compatible and incompatible interactions [78], thus suggesting that the signal exchange
608 between the two organisms is already occurring in this period.

609 To inhibit PTI, some fungal pathogens secrete proteins containing LysM motifs that
610 compete with plant receptors [79,80]. 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 [79,81], while effectors like *Avr4* protect chitin from fungal cell walls from degradation by host chitinase [82]. In addition, a study of the *H. vastatrix* secretome showed that effector candidates expressed in incompatible interaction (resistance) were more abundant within 24 hours, suggesting that these pre-haustorial effectors could be involved in the attempt to suppress PTI [83].

The expression results of the candidate receptors did not show difference in profiles between the groups of resistant and susceptible cultivars. Despite the IP showing high levels of expression at 24 hpi for the transcripts *Ca1-LYP*, *Ca2-LYK5* and *Ca-LYK4*, the susceptible cultivar MN showed equivalent levels of expression for *Ca1-CERK1* and *Ca2-LYP* or MN and CV showed comparable levels or even larger than the AR resistant cultivar for *Ca2-CERK1*, *Ca2-LYP*, *Ca1-LYK5* and *Ca-LYK4* (Fig 5). This result was expected, since the basal immunity is characterized by being broad-spectrum and non-specific [12,17]. The resistance of coffee to rust has been reported as pre-haustorial [77,84], in which resistant genotypes cease the growth of the fungus with mechanisms of pathogen recognition by resistance proteins. Thus, the difference between resistant and susceptible cultivars is generally evidenced in studies of expression of genes involved in pathogen-specific pathways and not in broad-spectrum receptors, such as PRRs [84].

Additionally, the recognition and signaling of PAMPs occurs when PRRs associate and act as part of multiprotein immune complexes on the cell surface [85,86]. Although they share common structural characteristics, these receptors are distinct in terms of recognized expression patterns and epitopes [23,25,52,62]. This shows that the receptors roles appear to have evolved

independently in different groups of plants [25,71]. Therefore, considering that all candidate receptors in coffee, described in this study, increased their expression from 6 hpi in all evaluated cultivars, each one may have possible roles in the basal immunity of *C. arabica.*

Conclusion

The results indicate that candidate sequences in *C. arabica* have protein domains and motifs characteristic of fungal PRRs and are homologous to *At-CERK1, At-LYK4, At-LYK5* and *Os-CEBiP*. Additionally, the expression of these genes was increased after the inoculation of *H. vastatrix* at all times and cultivars evaluated. Therefore, this study presents an advance in the understanding of the basal immunity of this species.

Acknowledgments

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

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

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

2. Boyd LA, Ridout C, O'Sullivan DM, Leach JE, Leung H. Plant-pathogen interactions: Disease resistance in modern agriculture. Trends Genet. 2013;29: 233–240.

658 doi:10.1016/j.tig.2012.10.011

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

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

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

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

670 7. Jeworutzki E, Roelfsema MRG, Anschütz U, Krol E, Elzenga JTM, Felix G, et al.
671 Early signaling through the Arabidopsis pattern recognition receptors FLS2 and EFR
672 involves Ca2+-associated opening of plasma membrane anion channels. Plant J.
673 2010;62: 367–378. doi:10.1111/j.1365-313X.2010.04155.x

674 8. Nomura H, Komori T, Uemura S, Kanda Y, Shimotani K, Nakai K, et al. Chloroplast-
675 mediated activation of plant immune signalling in Arabidopsis. Nat Commun. 2012;3:
676 910–926. doi:10.1038/ncomms1926

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

679 10. Ranf S, Eschen-Lippold L, Pecher P, Lee J, Scheel D. Interplay between calcium
680 signalling and early signalling elements during defence responses to microbe- or

damage-associated molecular patterns. Plant J. 2011;68: 100–113. doi:10.1111/j.1365-313X.2011.04671.x

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

12. 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

13. 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

14. 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

15. 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

16. 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

17. Boutrot F, Zipfel C. Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. Annu Rev Phytopathol. 2017;55: 257–286. doi:10.1146/annurev-phyto-080614-120106

18. 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

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

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

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

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

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

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

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

728 26. Zipfel C, Robatzek S, Navarro L, Oakeley E, al et. Bacterial disease resistance in
729 Arabidopsis through flagellin perception. Nature. 2004;428: 15–18.
730 doi:10.1038/nature02485

731 27. Tanaka S, Ichikawa A, Yamada K, Tsuji G, Nishiuchi T, Mori M, et al. HvCEBiP, a
732 gene homologous to rice chitin receptor CEBiP, contributes to basal resistance of
733 barley to *Magnaporthe oryzae*. BMC Plant Biol. 2010;10. doi:10.1186/1471-2229-10-
734 288

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

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

742 30. Lv Z, Huang Y, Ma B, Xiang Z, He N. LysM1 in MmLYK2 is a motif required for the
743 interaction of MmLYP1 and MmLYK2 in the chitin signaling. Plant Cell Rep.
744 2018;37: 1101–1112. doi:10.1007/s00299-018-2295-4

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

747 32. Cabral PGC, Maciel-Zambolim E, Oliveira SAS, Caixeta ET, Zambolim L. Genetic
748 diversity and structure of *Hemileia vastatrix* populations on *Coffea* spp. Plant Pathol.
749 2016;65: 196–204. doi:10.1111/ppa.12411

750 33. McCook S, Vandermeer J. The Big Rust and the Red Queen: Long-Term Perspectives

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

753 34. Avelino J, Cristancho M, Georgiou S, Imbach P, Aguilar L, Bornemann G, et al. The
754 coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible
755 causes and proposed solutions. Food Secur. 2015;7: 303–321. doi:10.1007/s12571-015-
756 0446-9

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

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

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

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

768 39. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence
769 alignment, interactive sequence choice and visualization. Brief Bioinform. 2018;20:
770 1160–1166. doi:10.1093/bib/bbx108

771 40. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis
772 program for Windows 95/98/NT. Nucleic Acids Symp Ser. 1999;41: 95–98.
773 doi:10.1039/c7qi00394c

774 41. Lashermes P, Combes MC, Robert J, Trouslot P, D'Hont A, Anthony F, et al.
775 Molecular characterization and origin of the Coffea arabica L. genome. Mol Gen
776 Genet. 1999;261: 259–266. doi:10.1007/s004380050965

777 42. Tran HTM, Ramaraj T, Furtado A, Lee LS, Henry RJ. Use of a draft genome of coffee
778 (*Coffea arabica*) to identify SNPs associated with caffeine content. Plant Biotechnol J.
779 2018;9: 1756–1766. doi:10.1111/pbi.12912

780 43. Monteiro ACA, de Resende MLV, Valente TCT, Ribeiro Junior PM, Pereira VF, da
781 Costa JR, et al. Manganese phosphite in coffee defence against *Hemileia vastatrix* , the
782 coffee rust fungus: biochemical and molecular analyses. J Phytopathol. 2016;164:
783 1043–1053. doi:10.1111/jph.12525

784 44. Pfaffl MW. A new mathematical model for relative quantification in real-time RT–
785 PCR. Mon Not R Astron Soc. 2001;29: e45–e45. doi:10.1111/j.1365-
786 2966.2012.21196.x

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

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

794 47. Fernandes-Brum CN, Garcia B de O, Moreira RO, Ságio SA, Barreto HG, Lima AA, et
795 al. A panel of the most suitable reference genes for RT-qPCR expression studies of
796 coffee: screening their stability under different conditions. Tree Genet Genomes.
797 2017;13: 1–13. doi:10.1007/s11295-017-1213-1

48. Reichel, T, Resende, MLV.; Monteiro, ACA.; Freitas, NC.; Botelho, DMS. Constitutive Defense Strategy of Coffee Under Field Conditions : A Comparative Assessment of Resistant and Susceptible Cultivars to Rust. Mol Biotechnol. 2021; 1–15. doi:10.1007/s12033-021-00405-9

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

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

51. R Core Team. 2020 [cited 2020 August 30] in: The R Project for Statistical Computing [Internet]. Vienna:The R Foundation for Statistical Computing. Availablefrom: https://www.r-project.org/

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

53. Brulé D, Villano C, Davies LJ, Trdá L, Claverie J, Héloir MC, et al. The grapevine (*Vitis vinifera*) LysM receptor kinases VvLYK1-1 and VvLYK1-2 mediate chitooligosaccharide-triggered immunity. Plant Biotechnol J. 2019;17: 812–825. doi:10.1111/pbi.13017

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

821 55. Liu S, Wang J, Han Z, Gong X, Zhang H, Chai J. Molecular Mechanism for Fungal
822 Cell Wall Recognition by Rice Chitin Receptor OsCEBiP. Structure. 2016;24: 1192–
823 1200. doi:10.1016/j.str.2016.04.014

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

826 57. Tombuloglu G, Tombuloglu H, Cevik E, Sabit H. Genome-wide identification of
827 Lysin-Motif Receptor-Like Kinase (LysM-RLK) gene family in *Brachypodium*
828 *distachyon* and docking analysis of chitin/LYK binding. Physiol Mol Plant Pathol.
829 2019;106: 217–225. doi:10.1016/j.pmpp.2019.03.002

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

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

836 60. Zhang XC, Wu X, Findley S, Wan J, Libault M, Nguyen HT, et al. Molecular evolution
837 of lysin motif-type receptor-like kinases in plants. Plant Physiol. 2007;144: 623–636.
838 doi:10.1104/pp.107.097097

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

843 62. Fliegmann J, Uhlenbroich S, Shinya T, Martinez Y, Lefebvre B, Shibuya N, et al.

844 Biochemical and phylogenetic analysis of CEBiP-like LysM domain-containing
845 extracellular proteins in higher plants. Plant Physiol Biochem. 2011;49: 709–720.
846 doi:10.1016/j.plaphy.2011.04.004

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

849 64. Nazarian-Firouzabadi F, Joshi S, Xue H, Kushalappa AC. Genome-wide in silico
850 identification of LysM-RLK genes in potato (*Solanum tuberosum L.*). Mol Biol Rep.
851 2019;46: 5005–5017. doi:10.1007/s11033-019-04951-z

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

855 66. Shinya T, Motoyama N, Ikeda A, Wada M, Kamiya K, Hayafune M, et al. Functional
856 Characterization of CEBiP and CERK1 Homologs in Arabidopsis and Rice Reveals the
857 Presence of Different Chitin Receptor Systems in Plants. Plant Cell Physiol. 2012;53:
858 1696–1706. doi:10.1093/pcp/pcs113

859 67. Ao Y, Li Z, Feng D, Xiong F, Liu J, Li JF, et al. OsCERK1 and OsRLCK176 play
860 important roles in peptidoglycan and chitin signaling in rice innate immunity. Plant J.
861 2014;80: 1072–1084. doi:10.1111/tpj.12710

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

865 69. Zeng L, Velásquez AC, Munkvold KR, Zhang J, Martin GB. A tomato LysM receptor-
866 like kinase promotes immunity and its kinase activity is inhibited by AvrPtoB. Plant J.

2012;69: 92–103. doi:10.1038/jid.2014.371

70. 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

71. 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

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

73. 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

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

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

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

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

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

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

80. Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, et 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

81. 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

82. 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

83. 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

84. 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

85. 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

86. 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 Genomics Workbench software. Gray bars show the conservation level of the positions; red letters, the different nucleotides; and red dashes, the gaps. Identity: 98,28%.

S6 Fig. Alignments of CDS from candidate sequences to *LYK4* (Scaffold_612.376 and Scaffold_952.320) in *C. arabica.* 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: 98,45%.

S1 Table. BLASTp analysis of the PRR reference sequences against the *C. arabica* genome in Phytozome.

Figures

Fig1.

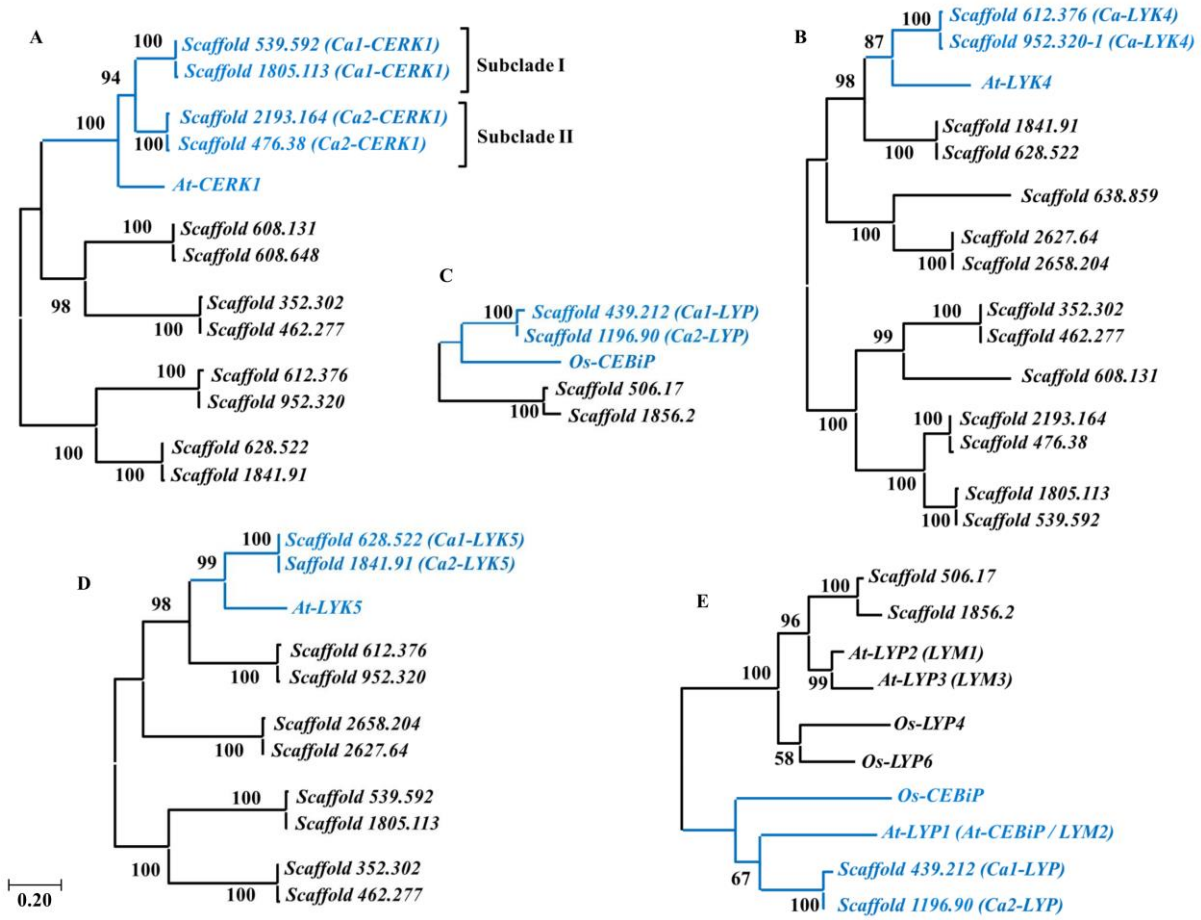


Fig2

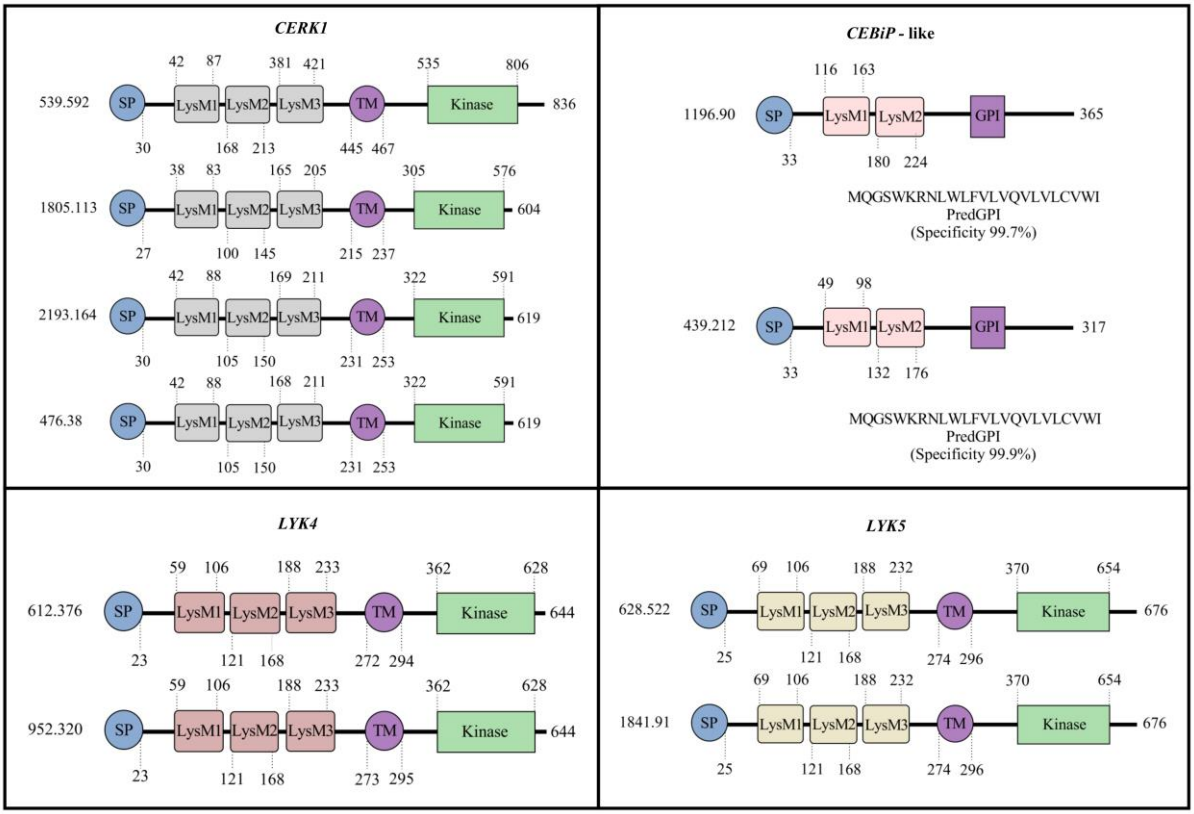

CERK1
539.592
SP
LysM1
LysM2
LysM3
TM
Kinase
836
42 87 381 421 535 806
30 168 213 445 467
1805.113
604
38 83 165 205 305 576
27 100 145 215 237
2193.164
619
42 88 169 211 322 591
30 105 150 231 253
476.38
619
42 88 168 211 322 591
30 105 150 231 253
CEBiP - like
1196.90
GPI
365
116 163
33 180 224
MQGSWKRNLWLFVLVQVLVLCVWI
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

Fig3.

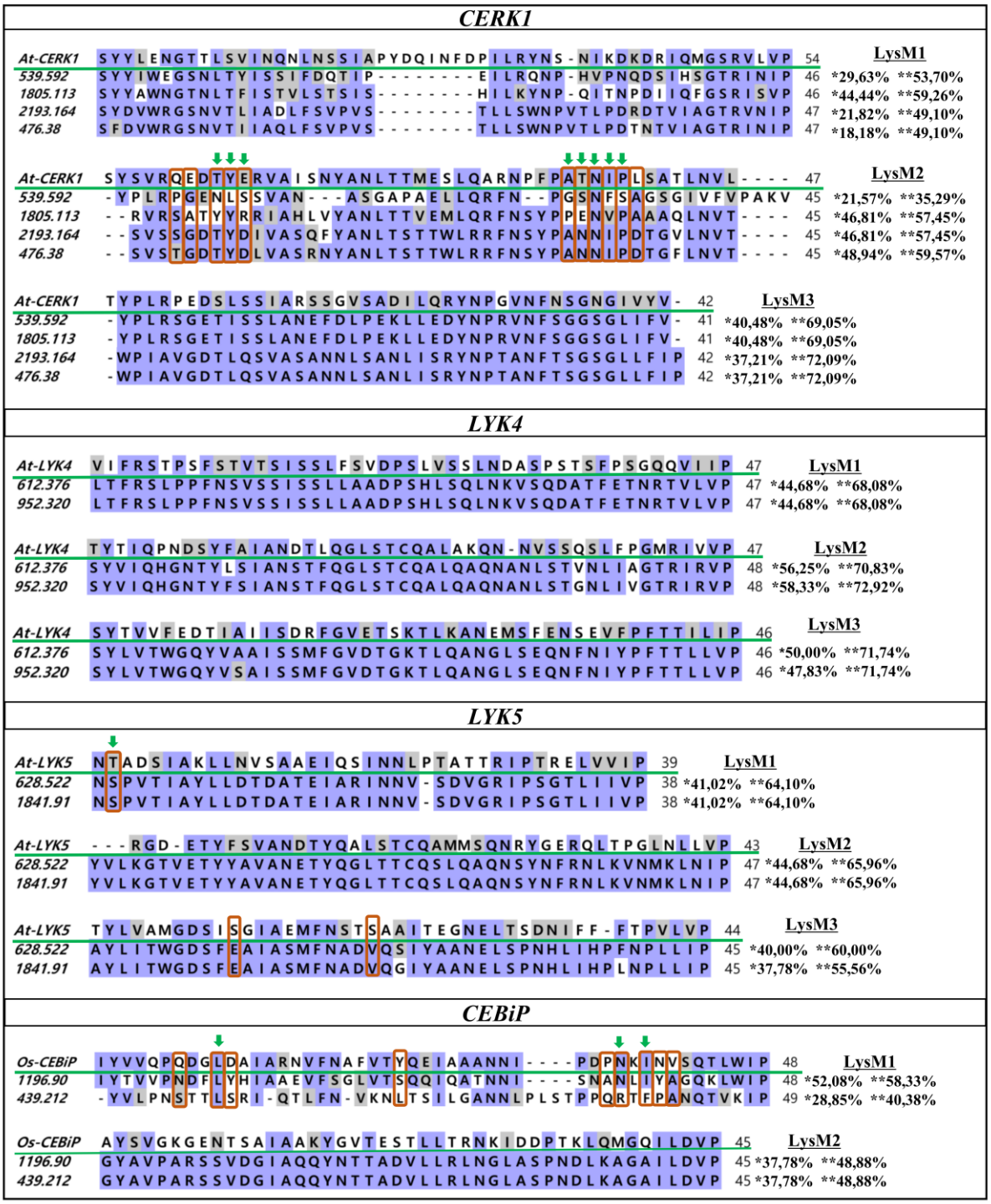

CERK1
At-CERK1 SYYLENGTTLSVINQNLNSSIAPYDQINFDPILRYNS-NIKDKDRIQMGSRVLVP 54
539.592 SYYIWEGSNLTYISSIFDQTIP--------EILRQNP-HVPNQDSIHSGTRINIP 46
1805.113 SYYAWNGTNLTFISTVLSTSIS--------HILKYNP-QITNPDIIQFGSRISVP 46
2193.164 SYDVWRGSNVTLIADLFSVPVS--------TLLSWNPVTLPDRDTVIAGTRVNIP 47
476.38 SFDVWRGSNVTIIAQLFSVPVS--------TLLSWNPVTLPDTNTVIAGTRINIP 47
LysM1
*29,63% **53,70%
*44,44% **59,26%
*21,82% **49,10%
*18,18% **49,10%
At-CERK1 SYSVRQEDTYERVAISNYANLTTMESLQARNPFPATNIPLSATLNVL---- 47
539.592 -YPLRPGENLSSVAN---ASGAPAELLQRFN--PGSNFSAGSGIVFVPAKV 45
1805.113 --RVRSATYYRRIAHLVYANLTTVEMLQRFNSYPPENVPAAAQLNVT---- 45
2193.164 --SVSSGDTYDIVASQFYANLTSTTWLRRFNSYPANNIPDTGVLNVT---- 45
476.38 --SVSTGDTYDLVASRNYANLTSTTWLRRFNSYPANNIPDTGFLNVT---- 45
LysM2
*21,57% **35,29%
*46,81% **57,45%
*46,81% **57,45%
*48,94% **59,57%
At-CERK1 TYPLRPEDSLSSIARSSGVSADILQRYNPGVNFNSGNGIVYV- 42
539.592 -YPLRSGETISSLANEFDLPEKLLEDYNPRVNFSGGSGLIFV- 41
1805.113 -YPLRSGETISSLANEFDLPEKLLEDYNPRVNFSGGSGLIFV- 41
2193.164 -WPIAVGDTLQSVASANNLSANLISRYNPTANFTSGSGLLFIP 42
476.38 -WPIAVGDTLQSVASANNLSANLISRYNPTANFTSGSGLLFIP 42
LysM3
*40,48% **69,05%
*40,48% **69,05%
*37,21% **72,09%
*37,21% **72,09%
LYK4
At-LYK4 VIFRSTPSFSTVTSISSLFSVDPSLVSSLNDASPSTSFPSGQQVIIP 47
612.376 LTFRSLPPFNSVSSISSLLAADPSHLSQLNKVSQDATFETNRTVLVP 47
952.320 LTFRSLPPFNSVSSISSLLAADPSHLSQLNKVSQDATFETNRTVLVP 47
LysM1
*44,68% **68,08%
*44,68% **68,08%
At-LYK4 TYTIQPNDSYFAIANDTLQGLSTCQALAKQN-NVSSQSLFPGMRIVVP 47
612.376 SYVIQHGNTYLSIANSTFQGLSTCQALQAQNANLSTVNLIAGTRIRVP 48
952.320 SYVIQHGNTYFSIANSTFQGLSTCQALQAQNANLSTGNLIVGTRIRVP 48
LysM2
*56,25% **70,83%
*58,33% **72,92%
At-LYK4 SYTVVFEDTIAIISDRFGVETSKTLKANEMSFENSEVFPFTTILIP 46
612.376 SYLVTWGQYVAAISSMFGVDTGKTLQANGLSEQNFNIYPFTTLLVP 46
952.320 SYLVTWGQYVSAISSMFGVDTGKTLQANGLSEQNFNIYPFTTLLVP 46
LysM3
*50,00% **71,74%
*47,83% **71,74%
LYK5
At-LYK5 NTADSIAKLLNVSAAEIQSINNLPTATTRIPTRELVVIP 39
628.522 NSPVTIAYLLDTDATEIARINNV-SDVGRIPSGTLIIVP 38
1841.91 NSPVTIAYLLDTDATEIARINNV-SDVGRIPSGTLIIVP 38
LysM1
*41,02% **64,10%
*41,02% **64,10%
At-LYK5 ---RGD-ETYFSVANDTYQALSTCQAMMSQNRYGERQLTPGLNLLVP 43
628.522 YVLKGTVETYYAVANETYQGLTTCQSLQAQNSYNFRNLKVNMKLNIP 47
1841.91 YVLKGTVETYYAVANETYQGLTTCQSLQAQNSYNFRNLKVNMKLNIP 47
LysM2
*44,68% **65,96%
*44,68% **65,96%
At-LYK5 TYLVAMGDSISGIAEMFNSTSAAITEGNELTSDNIFF-FTPVLVP 44
628.522 AYLITWGDSFEAIASMFNADVQSIYAANELSPNHLIHPFNPLLIP 45
1841.91 AYLITWGDSFEAIASMFNADVQGIYAANELSPNHLIHPLNPLLIP 45
LysM3
*40,00% **60,00%
*37,78% **55,56%
CEBiP
Os-CEBiP IYVVQPQDGLDAIARNVFNAFVTYQEIAAANNI----PDPNKINVSQTLWIP 48
1196.90 IYTVVPNDFLYHIAAEVFSGLVTSQQIQATNNI----SNANLIYAGQKLWIP 48
439.212 -YVLPNSTTLSRI-QTLFN-VKNLTSILGANNLPLSTPPQRTFPANQTVKIP 49
LysM1
*52,08% **58,33%
*28,85% **40,38%
Os-CEBiP AYSVGKGENTSAIAAKYGVTESTLLTRNKIDDPTKLQMGQILDVP 45
1196.90 GYAVPARSSVDGIAQQYNTTADVLLRLNGLASPNDLKAGAILDVP 45
439.212 GYAVPARSSVDGIAQQYNTTADVLLRLNGLASPNDLKAGAILDVP 45
LysM2
*37,78% **48,88%
*37,78% **48,88%

Fig4.

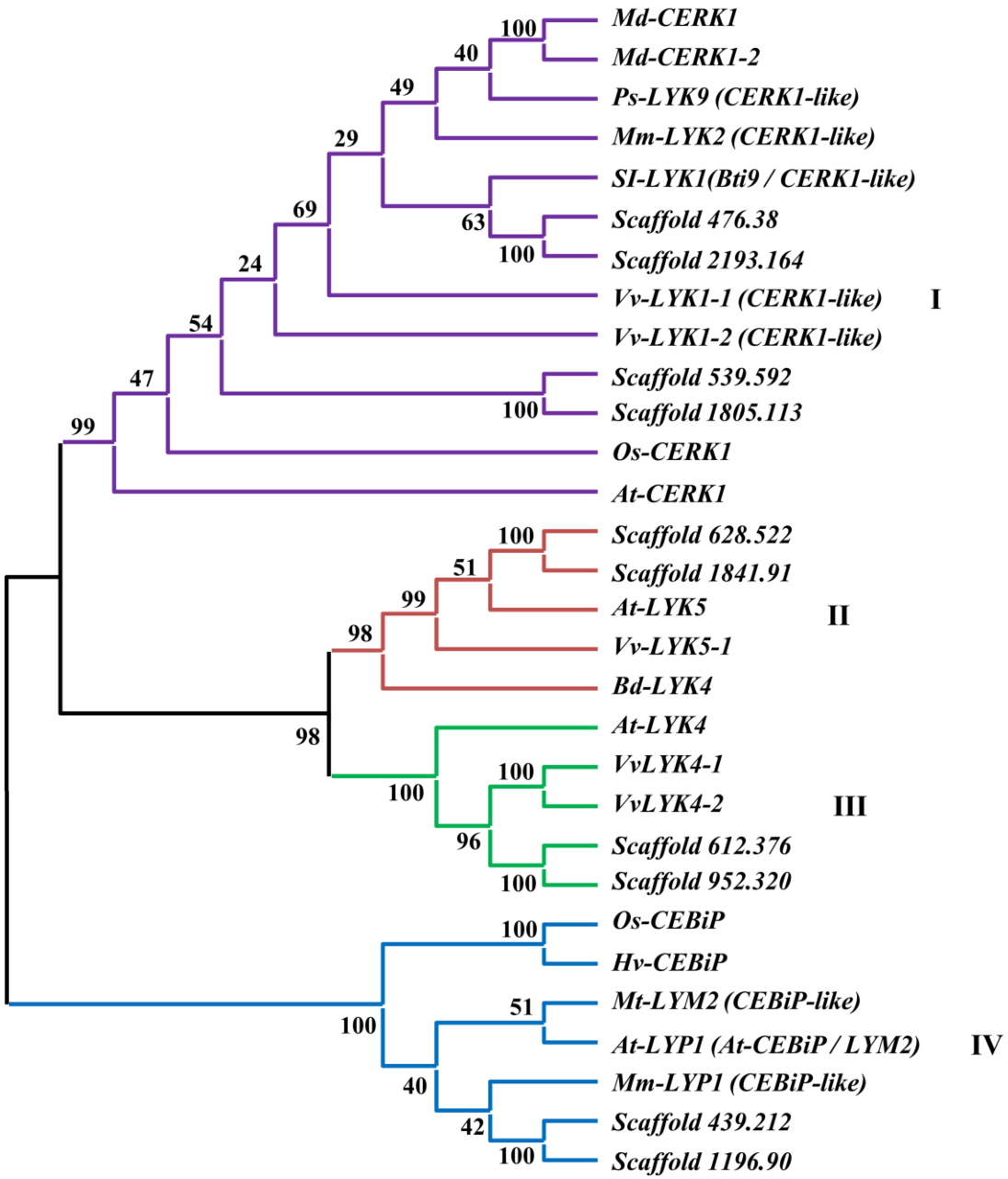
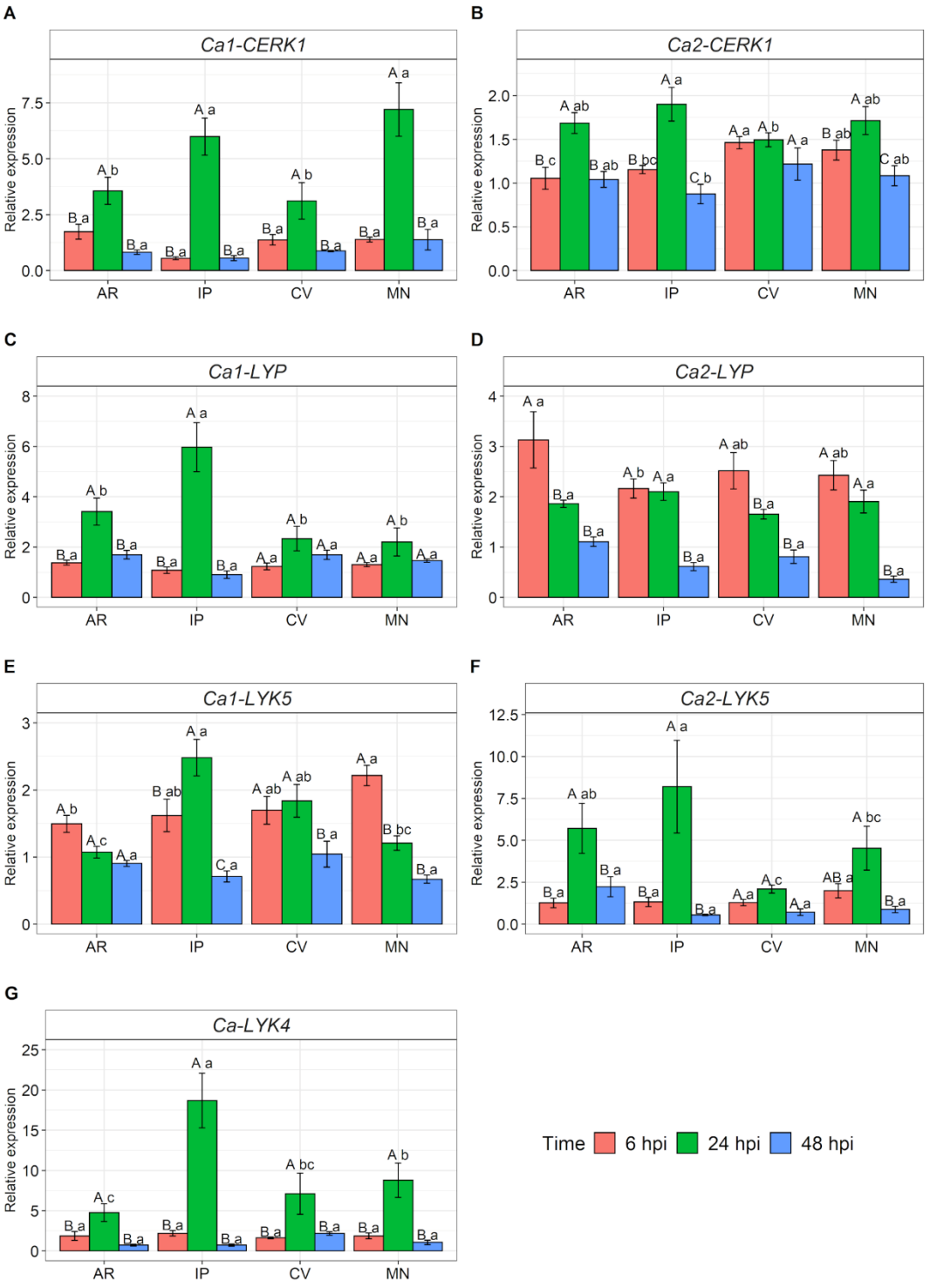
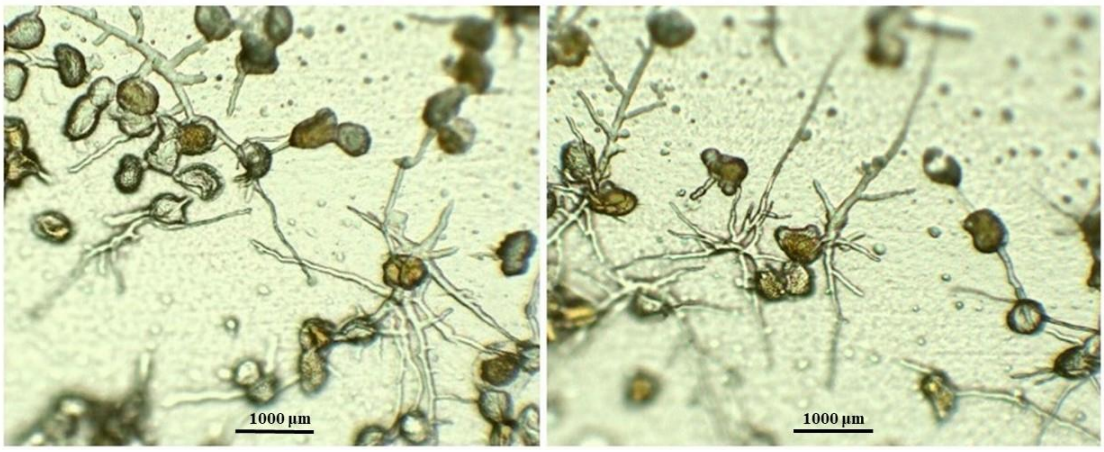

Md-CERK1
Md-CERK1-2
Ps-LYK9 (CERK1-like)
Mm-LYK2 (CERK1-like)
SI-LYK1(Bti9 / CERK1-like)
Scaffold 476.38
Scaffold 2193.164
Vv-LYK1-1 (CERK1-like)
Vv-LYK1-2 (CERK1-like)
Scaffold 539.592
Scaffold 1805.113
Os-CERK1
At-CERK1
Scaffold 628.522
Scaffold 1841.91
At-LYK5
Vv-LYK5-1
Bd-LYK4
At-LYK4
VvLYK4-1
VvLYK4-2
Scaffold 612.376
Scaffold 952.320
Os-CEBiP
Hv-CEBiP
Mt-LYM2 (CEBiP-like)
At-LYP1 (At-CEBiP / LYM2)
Mm-LYP1 (CEBiP-like)
Scaffold 439.212
Scaffold 1196.90
I
II
III
IV

Fig5.


A
Ca1-CERK1
B
Ca2-CERK1
C
Ca1-LYP
D
Ca2-LYP
E
Ca1-LYK5
F
Ca2-LYK5
G
Ca-LYK4
Relative expression
AR
IP
CV
MN
Time
6 hpi
24 hpi
48 hpi

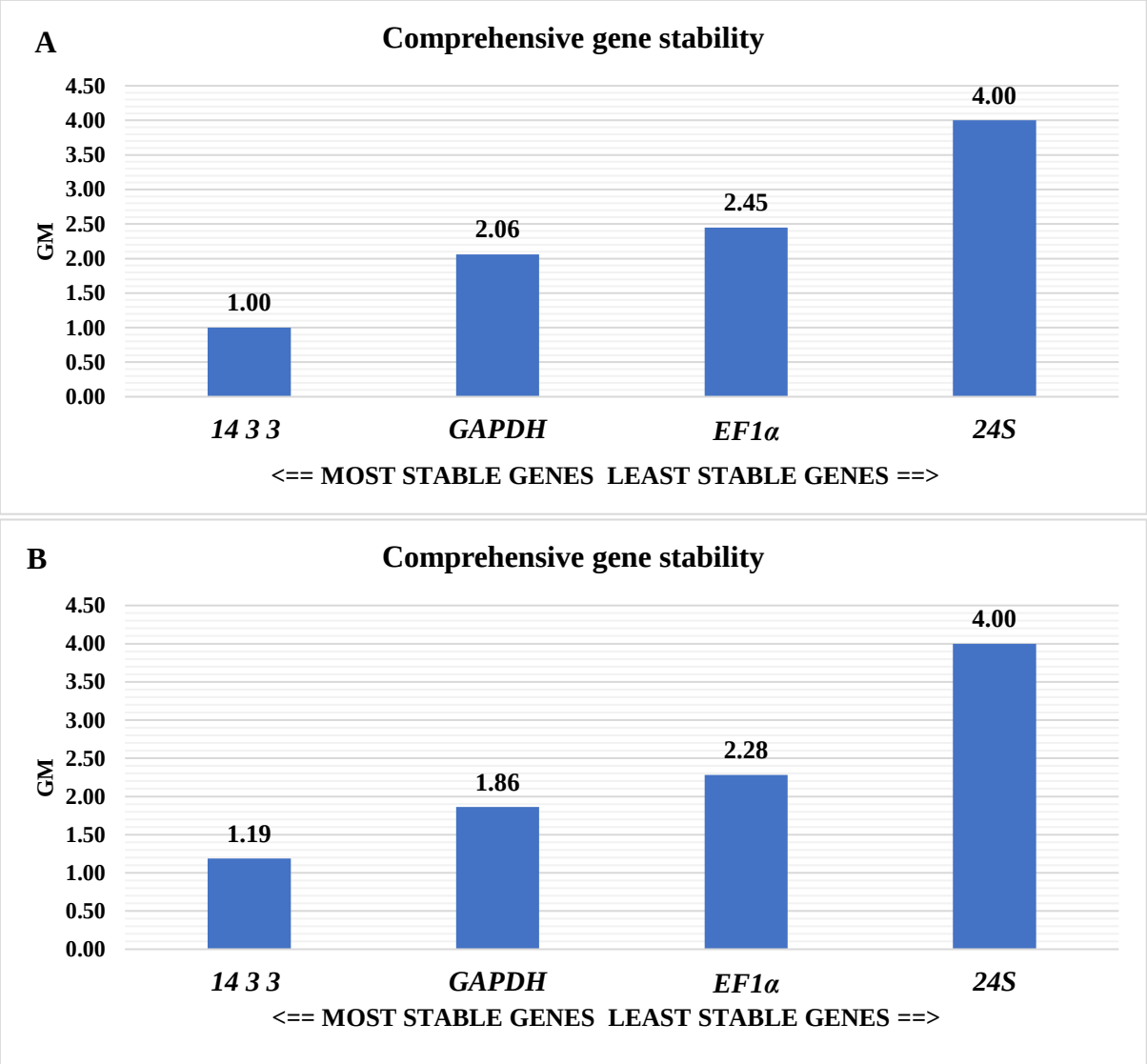
Supporting information – Figures and Table

S1fig.

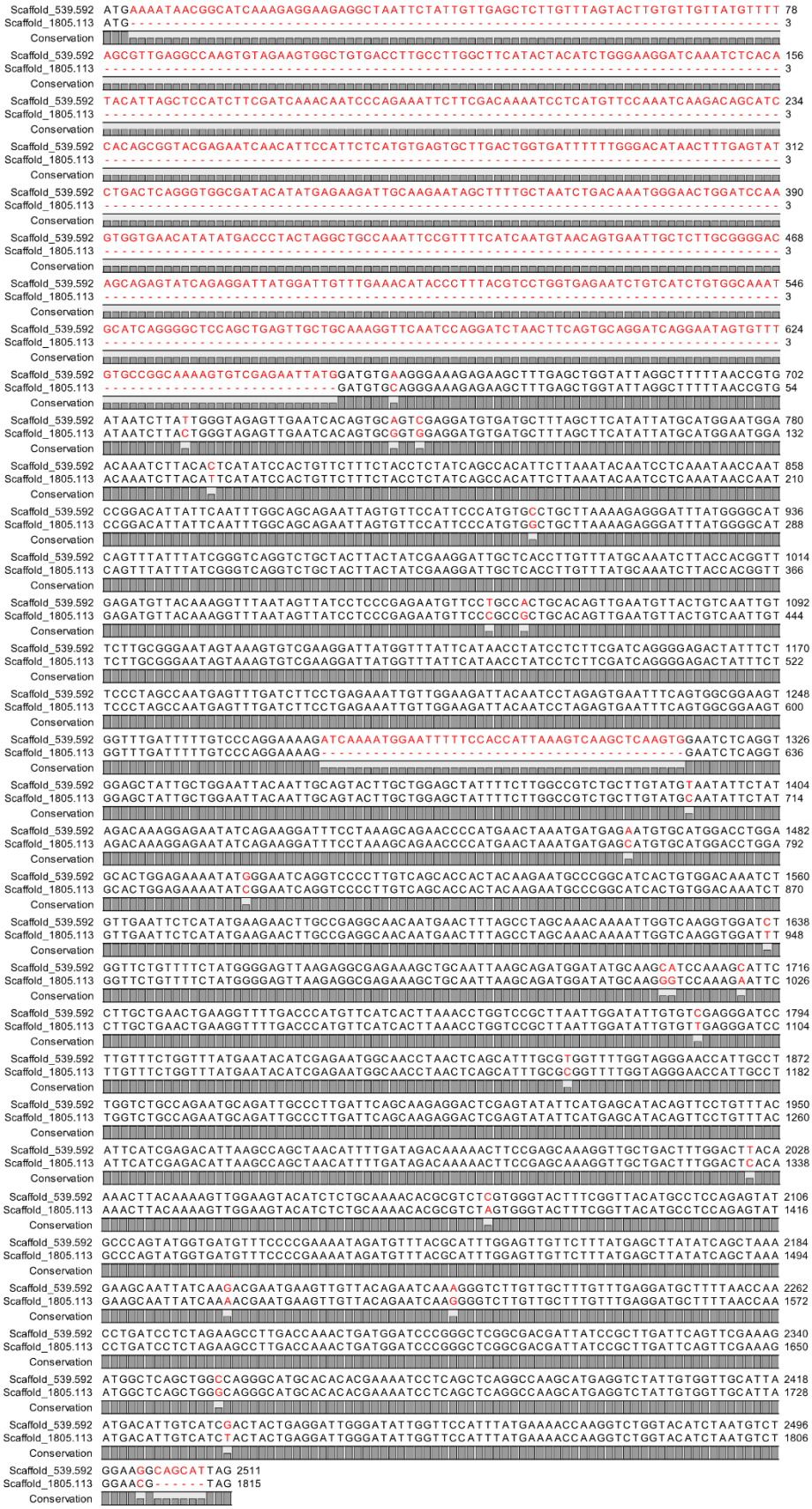


S2fig.

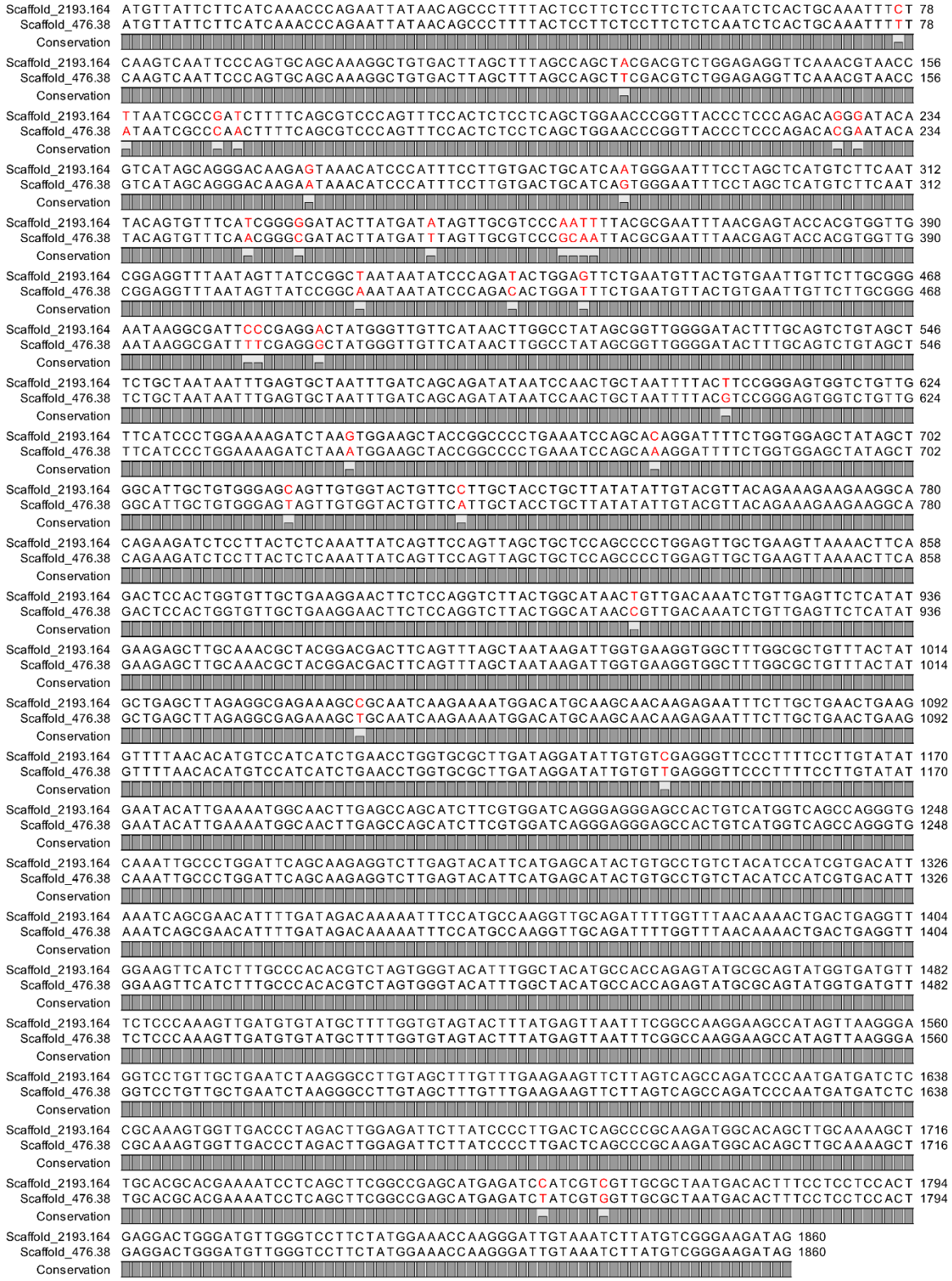
S3fig.


A
Comprehensive gene stability
GM
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1.00
2.06
2.45
4.00
14 3 3
GAPDH
EF1α
24S
<== MOST STABLE GENES LEAST STABLE GENES ==>
B
Comprehensive gene stability
GM
4.50
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00
1.19
1.86
2.28
4.00
14 3 3
GAPDH
EF1α
24S
<== MOST STABLE GENES LEAST STABLE GENES ==>

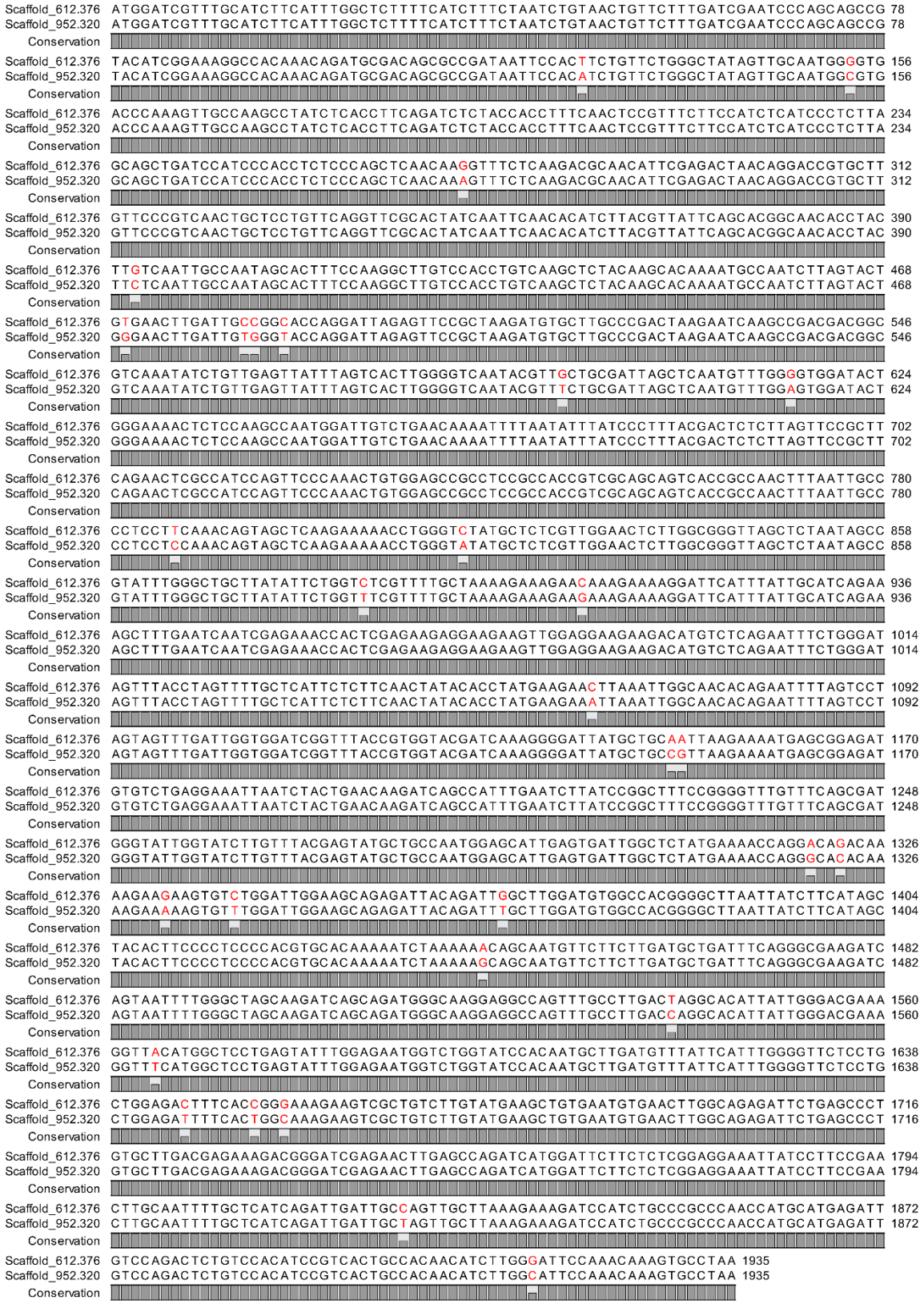
S4fig.


Scaffold_539.592 ATGAAAATAACGGCATCAAAGAGGAAGAGGCTAATTCTATTGTTGAGCTCTTGTTTAGTACTTGTGTTGTTATGTTTT 78
Scaffold_1805.113 ATG 3
Conservation
Scaffold_539.592 AGCGTTGAGGCCAAGTGTAGAAGTGGCTGTGACCTTGCCTTGGCTTCATACTACATCTGGGAAGGATCAAATCTCACA 156
Scaffold_1805.113 3
Conservation
Scaffold_539.592 TACATTAGCTCCATCTTCGATCAAACAATCCCAGAAATTCTTCGACAAAATCCTCATGTTCCAAATCAAGACAGCATC 234
Scaffold_1805.113 3
Conservation
Scaffold_539.592 CACAGCGGTACGAGAATCAACATTCCATTCTCATGTGAGTGCTTGACTGGTGATTTTTTGGGACATAACTTTGAGTAT 312
Scaffold_1805.113 3
Conservation
Scaffold_539.592 CTGACTCAGGGTGGCGATACATATGAGAAGATTGCAAGAATAGCTTTTGCTAATCTGACAAATGGGAACTGGATCCAA 390
Scaffold_1805.113 3
Conservation
Scaffold_539.592 GTGGTGAACATATATGACCCTACTAGGCTGCCAAATTCCGTTTTCATCAATGTAACAGTGAATTGCTCTTGCGGGGAC 468
Scaffold_1805.113 3
Conservation
Scaffold_539.592 AGCAGAGTATCAGAGGATTATGGATTGTTTGAAACATACCCTTTACGTCCTGGTGAGAATCTGTCATCTGTGGCAAAT 546
Scaffold_1805.113 3
Conservation
Scaffold_539.592 GCATCAGGGGCTCCAGCTGAGTTGCTGCAAAGGTTCAATCCAGGATCTAACTTCAGTGCAGGATCAGGAATAGTGTTT 624
Scaffold_1805.113 3
Conservation
Scaffold_539.592 GTGCCGGCAAAAGTGTCGAGAATTATGGATGTGAAGGGAAAGAGAAGCTTTGAGCTGGTATTAGGCTTTTTAACCGTG 702
Scaffold_1805.113 GATGTGCAGGGAAAGAGAAGCTTTGAGCTGGTATTAGGCTTTTTAACCGTG 54
Conservation
Scaffold_539.592 ATAATCTTATTGGGTAGAGTTGAATCACAGTGCAGTCGAGGATGTGATGCTTTAGCTTCATATTATGCATGGAATGGA 780
Scaffold_1805.113 ATAATCTTACTGGGTAGAGTTGAATCACAGTGCGGTGGAGGATGTGATGCTTTAGCTTCATATTATGCATGGAATGGA 132
Conservation
Scaffold_539.592 ACAAATCTTACACTCATATCCACTGTTCTTTCTACCTCTATCAGCCACATTCTTAAATACAATCCTCAAATAACCAAT 858
Scaffold_1805.113 ACAAATCTTACATTCATATCCACTGTTCTTTCTACCTCTATCAGCCACATTCTTAAATACAATCCTCAAATAACCAAT 210
Conservation
Scaffold_539.592 CCGGACATTATTCAATTTGGCAGCAGAATTAGTGTTCCATTCCCATGTGCCTGCTTAAAAGAGGGATTTATGGGGCAT 936
Scaffold_1805.113 CCGGACATTATTCAATTTGGCAGCAGAATTAGTGTTCCATTCCCATGTGGCTGCTTAAAAGAGGGATTTATGGGGCAT 288
Conservation
Scaffold_539.592 CAGTTTATTTATCGGGTCAGGTCTGCTACTTACTATCGAAGGATTGCTCACCTTGTTTATGCAAATCTTACCACGGTT 1014
Scaffold_1805.113 CAGTTTATTTATCGGGTCAGGTCTGCTACTTACTATCGAAGGATTGCTCACCTTGTTTATGCAAATCTTACCACGGTT 366
Conservation
Scaffold_539.592 GAGATGTTACAAAGGTTTAATAGTTATCCTCCCGAGAATGTTCCTGCCACTGCACAGTTGAATGTTACTGTCAATTGT 1092
Scaffold_1805.113 GAGATGTTACAAAGGTTTAATAGTTATCCTCCCGAGAATGTTCCCGCCGCTGCACAGTTGAATGTTACTGTCAATTGT 444
Conservation
Scaffold_539.592 TCTTGCGGGAATAGTAAAGTGTCGAAGGATTATGGTTTATTCATAACCTATCCTCTTCGATCAGGGGAGACTATTTCT 1170
Scaffold_1805.113 TCTTGCGGGAATAGTAAAGTGTCGAAGGATTATGGTTTATTCATAACCTATCCTCTTCGATCAGGGGAGACTATTTCT 522
Conservation
Scaffold_539.592 TCCCTAGCCAATGAGTTTGATCTTCCTGAGAAATTGTTGGAAGATTACAATCCTAGAGTGAATTTCAGTGGCGGAAGT 1248
Scaffold_1805.113 TCCCTAGCCAATGAGTTTGATCTTCCTGAGAAATTGTTGGAAGATTACAATCCTAGAGTGAATTTCAGTGGCGGAAGT 600
Conservation
Scaffold_539.592 GGTTTGATTTTTGTCCCAGGAAAAGATCAAAATGGAATTTTTCCACCATTAAAGTCAAGCTCAAGTGGAATCTCAGGT 1326
Scaffold_1805.113 GGTTTGATTTTTGTCCCAGGAAAAG GAATCTCAGGT 636
Conservation
Scaffold_539.592 GGAGCTATTGCTGGAATTACAATTGCAGTACTTGCTGGAGCTATTTTCTTGGCCGTCTGCTTGTATGTAATATTCTAT 1404
Scaffold_1805.113 GGAGCTATTGCTGGAATTACAATTGCAGTACTTGCTGGAGCTATTTTCTTGGCCGTCTGCTTGTATGCAATATTCTAT 714
Conservation
Scaffold_539.592 AGACAAAGGAGAATATCAGAAGGATTTCCTAAAGCAGAACCCCATGAACTAAATGATGAGAATGTGCATGGACCTGGA 1482
Scaffold_1805.113 AGACAAAGGAGAATATCAGAAGGATTTCCTAAAGCAGAACCCCATGAACTAAATGATGAGCATGTGCATGGACCTGGA 792
Conservation
Scaffold_539.592 GCACTGGAGAAAATATGGGAATCAGGTCCCCTTGTCAGCACCACTACAAGAATGCCCGGCATCACTGTGGACAAATCT 1560
Scaffold_1805.113 GCACTGGAGAAAATATCGGAATCAGGTCCCCTTGTCAGCACCACTACAAGAATGCCCGGCATCACTGTGGACAAATCT 870
Conservation
Scaffold_539.592 GTTGAATTCTCATATGAAGAACTTGCCGAGGCAACAATGAACTTTAGCCTAGCAAACAAAATTGGTCAAGGTGGATCT 1638
Scaffold_1805.113 GTTGAATTCTCATATGAAGAACTTGCCGAGGCAACAATGAACTTTAGCCTAGCAAACAAAATTGGTCAAGGTGGATTT 948
Conservation
Scaffold_539.592 GGTTCTGTTTTCTATGGGGAGTTAAGAGGCGAGAAAGCTGCAATTAAGCAGATGGATATGCAAGCATCCAAAGCATTC 1716
Scaffold_1805.113 GGTTCTGTTTTCTATGGGGAGTTAAGAGGCGAGAAAGCTGCAATTAAGCAGATGGATATGCAAGGGTCCAAAGAATTC 1026
Conservation
Scaffold_539.592 CTTGCTGAACTGAAGGTTTTGACCCATGTTCATCACTTAAACCTGGTCCGCTTAATTGGATATTGTGTCGAGGGATCC 1794
Scaffold_1805.113 CTTGCTGAACTGAAGGTTTTGACCCATGTTCATCACTTAAACCTGGTCCGCTTAATTGGATATTGTGTTGAGGGATCC 1104
Conservation
Scaffold_539.592 TTGTTTCTGGTTTATGAATACATCGAGAATGGCAACCTAACTCAGCATTTGCGTGGTTTTGGTAGGGAACCATTGCCT 1872
Scaffold_1805.113 TTGTTTCTGGTTTATGAATACATCGAGAATGGCAACCTAACTCAGCATTTGCGCGGTTTTGGTAGGGAACCATTGCCT 1182
Conservation
Scaffold_539.592 TGGTCTGCCAGAATGCAGATTGCCCTTGATTCAGCAAGAGGACTCGAGTATATTCATGAGCATACAGTTCCTGTTTAC 1950
Scaffold_1805.113 TGGTCTGCCAGAATGCAGATTGCCCTTGATTCAGCAAGAGGACTCGAGTATATTCATGAGCATACAGTTCCTGTTTAC 1260
Conservation
Scaffold_539.592 ATTCATCGAGACATTAAGCCAGCTAACATTTTGATAGACAAAAACTTCCGAGCAAAGGTTGCTGACTTTGGACTTACA 2028
Scaffold_1805.113 ATTCATCGAGACATTAAGCCAGCTAACATTTTGATAGACAAAAACTTCCGAGCAAAGGTTGCTGACTTTGGACTCACA 1338
Conservation
Scaffold_539.592 AAACTTACAAAAGTTGGAAGTACATCTCTGCAAAACACGCGTCTCGTGGGTACTTTCGGTTACATGCCTCCAGAGTAT 2106
Scaffold_1805.113 AAACTTACAAAAGTTGGAAGTACATCTCTGCAAAACACGCGTCTAGTGGGTACTTTCGGTTACATGCCTCCAGAGTAT 1416
Conservation
Scaffold_539.592 GCCCAGTATGGTGATGTTTCCCCGAAAATAGATGTTTACGCATTTGGAGTTGTTCTTTATGAGCTTATATCAGCTAAA 2184
Scaffold_1805.113 GCCCAGTATGGTGATGTTTCCCCGAAAATAGATGTTTACGCATTTGGAGTTGTTCTTTATGAGCTTATATCAGCTAAA 1494
Conservation
Scaffold_539.592 GAAGCAATTATCAAGACGAATGAAGTTGTTACAGAATCAAAGGGTCTTGTTGCTTTGTTTGAGGATGCTTTTAACCAA 2262
Scaffold_1805.113 GAAGCAATTATCAAAACGAATGAAGTTGTTACAGAATCAAGGGGTCTTGTTGCTTTGTTTGAGGATGCTTTTAACCAA 1572
Conservation
Scaffold_539.592 CCTGATCCTCTAGAAGCCTTGACCAAACTGATGGATCCCGGGCTCGGCGACGATTATCCGCTTGATTCAGTTCGAAAG 2340
Scaffold_1805.113 CCTGATCCTCTAGAAGCCTTGACCAAACTGATGGATCCCGGGCTCGGCGACGATTATCCGCTTGATTCAGTTCGAAAG 1650
Conservation
Scaffold_539.592 ATGGCTCAGCTGGCCAGGGCATGCACACACGAAAATCCTCAGCTCAGGCCAAGCATGAGGTCTATTGTGGTTGCATTA 2418
Scaffold_1805.113 ATGGCTCAGCTGGCCAGGGCATGCACACACGAAAATCCTCAGCTCAGGCCAAGCATGAGGTCTATTGTGGTTGCATTA 1728
Conservation
Scaffold_539.592 ATGACATTGTCATCGACTACTGAGGATTGGGATATTGGTTCCATTTATGAAAACCAAGGTCTGGTACATCTAATGTCT 2496
Scaffold_1805.113 ATGACATTGTCATCTACTACTGAGGATTGGGATATTGGTTCCATTTATGAAAACCAAGGTCTGGTACATCTAATGTCT 1806
Conservation
Scaffold_539.592 GGAAGGCAGCATTAG 2511
Scaffold_1805.113 GGAACG------TAG 1815
Conservation

S5fig.


Scaffold_2193.164 ATGTTATTCTTCATCAAACCCAGAATTATAACAGCCCTTTTACTCCTTCTCCTTCTCTCAATCTCACTGCAAATTTCT 78
Scaffold_476.38 ATGTTATTCTTCATCAAACCCAGAATTATAACAGCCCTTTTACTCCTTCTCCTTCTCTCAATCTCACTGCAAATTTTT 78
Conservation
Scaffold_2193.164 CAAGTCAATTCCCAGTGCAGCAAAGGCTGTGACTTAGCTTTAGCCAGCTACGACGTCTGGAGAGGTTCAAACGTAACC 156
Scaffold_476.38 CAAGTCAATTCCCAGTGCAGCAAAGGCTGTGACTTAGCTTTAGCCAGCTTCGACGTCTGGAGAGGTTCAAACGTAACC 156
Conservation
Scaffold_2193.164 TTAATCGCCGATCTTTTCAGCGTCCCAGTTTCCACTCTCCTCAGCTGGAACCCGGTTACCCTCCCAGACAGGGATACA 234
Scaffold_476.38 ATAATCGCCCAACTTTTCAGCGTCCCAGTTTCCACTCTCCTCAGCTGGAACCCGGTTACCCTCCCAGACACGAATACA 234
Conservation
Scaffold_2193.164 GTCATAGCAGGGACAAGAGTAAACATCCCATTTCCTTGTGACTGCATCAATGGGAATTTCCTAGCTCATGTCTTCAAT 312
Scaffold_476.38 GTCATAGCAGGGACAAGAATAAACATCCCATTTCCTTGTGACTGCATCAGTGGGAATTTCCTAGCTCATGTCTTCAAT 312
Conservation
Scaffold_2193.164 TACAGTGTTTCATCGGGGGATACTTATGATATAGTTGCGTCCCAATTTTACGCGAATTTAACGAGTACCACGTGGTTG 390
Scaffold_476.38 TACAGTGTTTCAACGGGCGATACTTATGATTTAGTTGCGTCCCGCAATTACGCGAATTTAACGAGTACCACGTGGTTG 390
Conservation
Scaffold_2193.164 CGGAGGTTTAATAGTTATCCGGCTAATAATATCCCAGATACTGGAGTTCTGAATGTTACTGTGAATTGTTCTTGCGGG 468
Scaffold_476.38 CGGAGGTTTAATAGTTATCCGGCAAATAATATCCCAGACACTGGATTTCTGAATGTTACTGTGAATTGTTCTTGCGGG 468
Conservation
Scaffold_2193.164 AATAAGGCGATTCCCGAGGACTATGGGTTGTTCATAACTTGGCCTATAGCGGTTGGGGATACTTTGCAGTCTGTAGCT 546
Scaffold_476.38 AATAAGGCGATTTTCGAGGGCTATGGGTTGTTCATAACTTGGCCTATAGCGGTTGGGGATACTTTGCAGTCTGTAGCT 546
Conservation
Scaffold_2193.164 TCTGCTAATAATTTGAGTGCTAATTTGATCAGCAGATATAATCCAACTGCTAATTTTACTTCCGGGAGTGGTCTGTTG 624
Scaffold_476.38 TCTGCTAATAATTTGAGTGCTAATTTGATCAGCAGATATAATCCAACTGCTAATTTTACGTCCGGGAGTGGTCTGTTG 624
Conservation
Scaffold_2193.164 TTCATCCCTGGAAAAGATCTAAGTGGAAGCTACCGGCCCCTGAAATCCAGCACAGGATTTTCTGGTGGAGCTATAGCT 702
Scaffold_476.38 TTCATCCCTGGAAAAGATCTAAATGGAAGCTACCGGCCCCTGAAATCCAGCAAAGGATTTTCTGGTGGAGCTATAGCT 702
Conservation
Scaffold_2193.164 GGCATTGCTGTGGGAGCAGTTGTGGTACTGTTCCTTGCTACCTGCTTATATATTGTACGTTACAGAAAGAAGAAGGCA 780
Scaffold_476.38 GGCATTGCTGTGGGAGTAGTTGTGGTACTGTTCATTGCTACCTGCTTATATATTGTACGTTACAGAAAGAAGAAGGCA 780
Conservation
Scaffold_2193.164 CAGAAGATCTCCTTACTCTCAAATTATCAGTTCCAGTTAGCTGCTCCAGCCCCTGGAGTTGCTGAAGTTAAAACTTCA 858
Scaffold_476.38 CAGAAGATCTCCTTACTCTCAAATTATCAGTTCCAGTTAGCTGCTCCAGCCCCTGGAGTTGCTGAAGTTAAAACTTCA 858
Conservation
Scaffold_2193.164 GACTCCACTGGTGTTGCTGAAGGAACTTCTCCAGGTCTTACTGGCATAACTGTTGACAAATCTGTTGAGTTCTCATAT 936
Scaffold_476.38 GACTCCACTGGTGTTGCTGAAGGAACTTCTCCAGGTCTTACTGGCATAACCGTTGACAAATCTGTTGAGTTCTCATAT 936
Conservation
Scaffold_2193.164 GAAGAGCTTGCAAACGCTACGGACGACTTCAGTTTAGCTAATAAGATTGGTGAAGGTGGCTTTGGCGCTGTTTACTAT 1014
Scaffold_476.38 GAAGAGCTTGCAAACGCTACGGACGACTTCAGTTTAGCTAATAAGATTGGTGAAGGTGGCTTTGGCGCTGTTTACTAT 1014
Conservation
Scaffold_2193.164 GCTGAGCTTAGAGGCGAGAAAGCCGCAATCAAGAAAATGGACATGCAAGCAACAAGAGAATTTCTTGCTGAACTGAAG 1092
Scaffold_476.38 GCTGAGCTTAGAGGCGAGAAAGCTGCAATCAAGAAAATGGACATGCAAGCAACAAGAGAATTTCTTGCTGAACTGAAG 1092
Conservation
Scaffold_2193.164 GTTTTAACACATGTCCATCATCTGAACCTGGTGCGCTTGATAGGATATTGTGTCGAGGGTTCCCTTTTCCTTGTATAT 1170
Scaffold_476.38 GTTTTAACACATGTCCATCATCTGAACCTGGTGCGCTTGATAGGATATTGTGTTGAGGGTTCCCTTTTCCTTGTATAT 1170
Conservation
Scaffold_2193.164 GAATACATTGAAAATGGCAACTTGAGCCAGCATCTTCGTGGATCAGGGAGGGAGCCACTGTCATGGTCAGCCAGGGTG 1248
Scaffold_476.38 GAATACATTGAAAATGGCAACTTGAGCCAGCATCTTCGTGGATCAGGGAGGGAGCCACTGTCATGGTCAGCCAGGGTG 1248
Conservation
Scaffold_2193.164 CAAATTGCCCTGGATTCAGCAAGAGGTCTTGAGTACATTCATGAGCATACTGTGCCTGTCTACATCCATCGTGACATT 1326
Scaffold_476.38 CAAATTGCCCTGGATTCAGCAAGAGGTCTTGAGTACATTCATGAGCATACTGTGCCTGTCTACATCCATCGTGACATT 1326
Conservation
Scaffold_2193.164 AAATCAGCGAACATTTTGATAGACAAAAATTTCCATGCCAAGGTTGCAGATTTTGGTTTAACAAAACTGACTGAGGTT 1404
Scaffold_476.38 AAATCAGCGAACATTTTGATAGACAAAAATTTCCATGCCAAGGTTGCAGATTTTGGTTTAACAAAACTGACTGAGGTT 1404
Conservation
Scaffold_2193.164 GGAAGTTCATCTTTGCCCACACGTCTAGTGGGTACATTTGGCTACATGCCACCAGAGTATGCGCAGTATGGTGATGTT 1482
Scaffold_476.38 GGAAGTTCATCTTTGCCCACACGTCTAGTGGGTACATTTGGCTACATGCCACCAGAGTATGCGCAGTATGGTGATGTT 1482
Conservation
Scaffold_2193.164 TCTCCCAAAGTTGATGTGTATGCTTTTGGTGTAGTACTTTATGAGTTAATTTCGGCCAAGGAAGCCATAGTTAAGGGA 1560
Scaffold_476.38 TCTCCCAAAGTTGATGTGTATGCTTTTGGTGTAGTACTTTATGAGTTAATTTCGGCCAAGGAAGCCATAGTTAAGGGA 1560
Conservation
Scaffold_2193.164 GGTCCTGTTGCTGAATCTAAGGGCCTTGTAGCTTTGTTTGAAGAAGTTCTTAGTCAGCCAGATCCCAATGATGATCTC 1638
Scaffold_476.38 GGTCCTGTTGCTGAATCTAAGGGCCTTGTAGCTTTGTTTGAAGAAGTTCTTAGTCAGCCAGATCCCAATGATGATCTC 1638
Conservation
Scaffold_2193.164 CGCAAAGTGGTTGACCCTAGACTTGGAGATTCTTATCCCCTTGACTCAGCCCGCAAGATGGCACAGCTTGCAAAAGCT 1716
Scaffold_476.38 CGCAAAGTGGTTGACCCTAGACTTGGAGATTCTTATCCCCTTGACTCAGCCCGCAAGATGGCACAGCTTGCAAAAGCT 1716
Conservation
Scaffold_2193.164 TGCACGCACGAAAATCCTCAGCTTCGGCCGAGCATGAGATCCATCGTCGTTGCGCTAATGACACTTTCCTCCTCCACT 1794
Scaffold_476.38 TGCACGCACGAAAATCCTCAGCTTCGGCCGAGCATGAGATCTATCGTGGTTGCGCTAATGACACTTTCCTCCTCCACT 1794
Conservation
Scaffold_2193.164 GAGGACTGGGATGTTGGGTCCTTCTATGGAAACCAAGGGATTGTAAATCTTATGTCGGGAAGATAG 1860
Scaffold_476.38 GAGGACTGGGATGTTGGGTCCTTCTATGGAAACCAAGGGATTGTAAATCTTATGTCGGGAAGATAG 1860
Conservation

S6fig.


Scaffold_612.376 ATGGATCGTTTGCATCTTCATTTGGCTCTTTTCATCTTTCTAATCTGTAACTGTTCTTTGATCGAATCCCAGCAGCCG 78
Scaffold_952.320 ATGGATCGTTTGCATCTTCATTTGGCTCTTTTCATCTTTCTAATCTGTAACTGTTCTTTGATCGAATCCCAGCAGCCG 78
Conservation
Scaffold_612.376 TACATCGGAAAGGCCACAAACAGATGCGACAGCGCCGATAATTCCACTTCTGTTCTGGGCTATAGTTGCAATGGGGTG 156
Scaffold_952.320 TACATCGGAAAGGCCACAAACAGATGCGACAGCGCCGATAATTCCACATCTGTTCTGGGCTATAGTTGCAATGGCGTG 156
Conservation
Scaffold_612.376 ACCCAAAGTTGCCAAGCCTATCTCACCTTCAGATCTCTACCACCTTTCAACTCCGTTTCTTCCATCTCATCCCTCTTA 234
Scaffold_952.320 ACCCAAAGTTGCCAAGCCTATCTCACCTTCAGATCTCTACCACCTTTCAACTCCGTTTCTTCCATCTCATCCCTCTTA 234
Conservation
Scaffold_612.376 GCAGCTGATCCATCCCACCTCTCCCAGCTCAACAAGGTTTCTCAAGACGCAACATTCGAGACTAACAGGACCGTGCTT 312
Scaffold_952.320 GCAGCTGATCCATCCCACCTCTCCCAGCTCAACAAAGTTTCTCAAGACGCAACATTCGAGACTAACAGGACCGTGCTT 312
Conservation
Scaffold_612.376 GTTCCCGTCAACTGCTCCTGTTCAGGTTCGCACTATCAATTCAACACATCTTACGTTATTCAGCACGGCAACACCTAC 390
Scaffold_952.320 GTTCCCGTCAACTGCTCCTGTTCAGGTTCGCACTATCAATTCAACACATCTTACGTTATTCAGCACGGCAACACCTAC 390
Conservation
Scaffold_612.376 TTGTCAATTGCCAATAGCACTTTCCAAGGCTTGTCCACCTGTCAAGCTCTACAAGCACAAAATGCCAATCTTAGTACT 468
Scaffold_952.320 TTCTCAATTGCCAATAGCACTTTCCAAGGCTTGTCCACCTGTCAAGCTCTACAAGCACAAAATGCCAATCTTAGTACT 468
Conservation
Scaffold_612.376 GTGAACTTGATTGCCGGCACCAGGATTAGAGTTCCGCTAAGATGTGCTTGCCCGACTAAGAATCAAGCCGACGACGGC 546
Scaffold_952.320 GGGAACTTGATTGTGGGTACCAGGATTAGAGTTCCGCTAAGATGTGCTTGCCCGACTAAGAATCAAGCCGACGACGGC 546
Conservation
Scaffold_612.376 GTCAAATATCTGTTGAGTTATTTAGTCACTTGGGGTCAATACGTTGCTGCGATTAGCTCAATGTTTGGGGTGGATACT 624
Scaffold_952.320 GTCAAATATCTGTTGAGTTATTTAGTCACTTGGGGTCAATACGTTTCTGCGATTAGCTCAATGTTTGGAGTGGATACT 624
Conservation
Scaffold_612.376 GGGAAAACTCTCCAAGCCAATGGATTGTCTGAACAAAATTTTAATATTTATCCCTTTACGACTCTCTTAGTTCCGCTT 702
Scaffold_952.320 GGGAAAACTCTCCAAGCCAATGGATTGTCTGAACAAAATTTTAATATTTATCCCTTTACGACTCTCTTAGTTCCGCTT 702
Conservation
Scaffold_612.376 CAGAACTCGCCATCCAGTTCCCAAACTGTGGAGCCGCCTCCGCCACCGTCGCAGCAGTCACCGCCAACTTTAATTGCC 780
Scaffold_952.320 CAGAACTCGCCATCCAGTTCCCAAACTGTGGAGCCGCCTCCGCCACCGTCGCAGCAGTCACCGCCAACTTTAATTGCC 780
Conservation
Scaffold_612.376 CCTCCTTCAAACAGTAGCTCAAGAAAAACCTGGGTCTATGCTCTCGTTGGAACTCTTGGCGGGTTAGCTCTAATAGCC 858
Scaffold_952.320 CCTCCTCCAAACAGTAGCTCAAGAAAAACCTGGGTATATGCTCTCGTTGGAACTCTTGGCGGGTTAGCTCTAATAGCC 858
Conservation
Scaffold_612.376 GTATTTGGGCTGCTTATATTCTGGTCTCGTTTTGCTAAAAGAAAGAACAAAGAAAAGGATTCATTTATTGCATCAGAA 936
Scaffold_952.320 GTATTTGGGCTGCTTATATTCTGGTTTCGTTTTGCTAAAAGAAAGAAGAAAGAAAAGGATTCATTTATTGCATCAGAA 936
Conservation
Scaffold_612.376 AGCTTTGAATCAATCGAGAAACCACTCGAGAAGAGGAAGAAGTTGGAGGAAGAAGACATGTCTCAGAATTTCTGGGAT 1014
Scaffold_952.320 AGCTTTGAATCAATCGAGAAACCACTCGAGAAGAGGAAGAAGTTGGAGGAAGAAGACATGTCTCAGAATTTCTGGGAT 1014
Conservation
Scaffold_612.376 AGTTTACCTAGTTTTGCTCATTCTCTTCAACTATACACCTATGAAGAACTTAAATTGGCAACACAGAATTTTAGTCCT 1092
Scaffold_952.320 AGTTTACCTAGTTTTGCTCATTCTCTTCAACTATACACCTATGAAGAATTAAATTGGCAACACAGAATTTTAGTCCT 1092
Conservation
Scaffold_612.376 AGTAGTTTGATTGGTGGATCGGTTTACCGTGGTACGATCAAAGGGGATTATGCTGCAATTAAGAAAATGAGCGGAGAT 1170
Scaffold_952.320 AGTAGTTTGATTGGTGGATCGGTTTACCGTGGTACGATCAAAGGGGATTATGCTGCCGTTAAGAAAATGAGCGGAGAT 1170
Conservation
Scaffold_612.376 GTGTCTGAGGAAATTAATCTACTGAACAAGATCAGCCATTTGAATCTTATCCGGCTTTCCGGGGTTTGTTTCAGCGAT 1248
Scaffold_952.320 GTGTCTGAGGAAATTAATCTACTGAACAAGATCAGCCATTTGAATCTTATCCGGCTTTCCGGGGTTTGTTTCAGCGAT 1248
Conservation
Scaffold_612.376 GGGTATTGGTATCTTGTTTACGAGTATGCTGCCAATGGAGCATTGAGTGATTGGCTCTATGAAAACCAGGACAGACAA 1326
Scaffold_952.320 GGGTATTGGTATCTTGTTTACGAGTATGCTGCCAATGGAGCATTGAGTGATTGGCTCTATGAAAACCAGGGCACACAA 1326
Conservation
Scaffold_612.376 AAGAAGAAGTGTCTGGATTGGAAGCAGAGATTACAGATTGGCTTGGATGTGGCCACGGGGCTTAATTATCTTCATAGC 1404
Scaffold_952.320 AAGAAAAAGTGTTTGGATTGGAAGCAGAGATTACAGATTTGCTTGGATGTGGCCACGGGGCTTAATTATCTTCATAGC 1404
Conservation
Scaffold_612.376 TACACTTCCCCTCCCCACGTGCACAAAAATCTAAAAAACAGCAATGTTCTTCTTGATGCTGATTTCAGGGCGAAGATC 1482
Scaffold_952.320 TACACTTCCCCTCCCCACGTGCACAAAAATCTAAAAAGCAGCAATGTTCTTCTTGATGCTGATTTCAGGGCGAAGATC 1482
Conservation
Scaffold_612.376 AGTAATTTTGGGCTAGCAAGATCAGCAGATGGGCAAGGAGGCCAGTTTGCCTTGACTAGGCACATTATTGGGACGAAA 1560
Scaffold_952.320 AGTAATTTTGGGCTAGCAAGATCAGCAGATGGGCAAGGAGGCCAGTTTGCCTTGACCAGGCACATTATTGGGACGAAA 1560
Conservation
Scaffold_612.376 GGTTACATGGCTCCTGAGTATTTGGAGAATGGTCTGGTATCCACAATGCTTGATGTTTATTCATTTGGGGTTCTCCTG 1638
Scaffold_952.320 GGTTTCATGGCTCCTGAGTATTTGGAGAATGGTCTGGTATCCACAATGCTTGATGTTTATTCATTTGGGGTTCTCCTG 1638
Conservation
Scaffold_612.376 CTGGAGACTTTCACCGGGAAAGAAGTCGCTGTCTTGTATGAAGCTGTGAATGTGAACTTGGCAGAGATTCTGAGCCCT 1716
Scaffold_952.320 CTGGAGATTTTCACTGGCAAAGAAGTCGCTGTCTTGTATGAAGCTGTGAATGTGAACTTGGCAGAGATTCTGAGCCCT 1716
Conservation
Scaffold_612.376 GTGCTTGACGAGAAAGACGGGATCGAGAACTTGAGCCAGATCATGGATTCTTCTCTCGGAGGAAATTATCCTTCCGAA 1794
Scaffold_952.320 GTGCTTGACGAGAAAGACGGGATCGAGAACTTGAGCCAGATCATGGATTCTTCTCTCGGAGGAAATTATCCTTCCGAA 1794
Conservation
Scaffold_612.376 CTTGCAATTTTGCTCATCAGATTGATTGCCAGTTGCTTAAAGAAAGATCCATCTGCCCGCCCAACCATGCATGAGATT 1872
Scaffold_952.320 CTTGCAATTTTGCTCATCAGATTGATTGCTAGTTGCTTAAAGAAAGATCCATCTGCCCGCCCAACCATGCATGAGATT 1872
Conservation
Scaffold_612.376 GTCCAGACTCTGTCCACATCCGTCACTGCCACAACATCTTGGGATTCCAAACAAAGTGCCTAA 1935
Scaffold_952.320 GTCCAGACTCTGTCCACATCCGTCACTGCCACAACATCTTGGCATTCCAAACAAAGTGCCTAA 1935
Conservation

S1table.

Query id	Subject id	% Identity	Alignment length	Mismatches	Gap opens	Q. start	Q. end	S. start	S. end	e-value	Bit score
Os-CEBiP	evm.model.Scaffold_1196.90 (1 of 2) PTHR33734:SF5 - LYSM DOMAIN-CONTAINING GPI-ANCHORED PROTEIN 2	42.258	310	162	7	37	331	40	347	3.99E-62	206
Os-CEBiP	evm.model.Scaffold_506.17 (1 of 3) PTHR23354:SF75 - LYSM DOMAIN-CONTAINING GPI-ANCHORED PROTEIN 1	35.622	233	138	7	92	315	93	322	1.59E-31	124
Os-CEBiP	evm.model.Scaffold_1856.2 (1 of 3) PTHR23354:SF75 - LYSM DOMAIN-CONTAINING GPI-ANCHORED PROTEIN 1	34.764	233	140	7	92	315	98	327	1.14E-30	122
Os-CEBiP	evm.model.Scaffold_439.212 (1 of 2) PTHR33734:SF5 - LYSM DOMAIN-CONTAINING GPI-ANCHORED PROTEIN 2	35.385	260	132	11	86	316	26	278	6.13E-27	110
At-LYK5	evm.model.Scaffold_952.320 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	40.819	659	354	16	6	651	4	639	2.49E-139	426
At-LYK5	evm.model.Scaffold_612.376 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	40.516	659	356	16	6	651	4	639	1.71E-138	424
At-LYK5	evm.model.Scaffold_1841.91 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	58.631	336	115	5	349	662	339	672	2.43E-128	399
At-LYK5	evm.model.Scaffold_628.522 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	58.036	336	117	5	349	662	339	672	1.18E-127	397
At-LYK5	evm.model.Scaffold_1805.113 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	37.458	299	176	5	360	651	293	587	8.55E-62	219

At-LYK5	evm.model.Scaffold_539.592 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	37.793	299	175	5	360	651	523	817	1.07E-61	222
At-LYK5	evm.model.Scaffold_2627.64 (1 of 4) PF00069//PF01476 - Protein kinase domain (Pkinase) // LysM domain (LysM)	41.275	298	160	7	355	644	328	618	1.82E-57	207
At-LYK5	evm.model.Scaffold_2658.204 (1 of 4) PF00069//PF01476 - Protein kinase domain (Pkinase) // LysM domain (LysM)	40.94	298	161	7	355	644	328	618	6.30E-57	206
At-LYK5	evm.model.Scaffold_352.302 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	35.35	314	185	6	354	651	336	647	1.49E-55	202
At-LYK5	evm.model.Scaffold_462.277 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	34.713	314	187	6	354	651	302	613	6.84E-55	200
At-CERK1	evm.model.Scaffold_539.592 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	56.109	622	258	7	2	617	222	834	0	657
At-CERK1	evm.model.Scaffold_476.38 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	57.261	606	242	9	18	617	25	619	0	655
At-CERK1	evm.model.Scaffold_2193.164 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	57.546	603	239	9	21	617	28	619	0	652
At-CERK1	evm.model.Scaffold_1805.113 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	55.145	622	248	7	2	616	6	603	0	645
At-CERK1	evm.model.Scaffold_608.131 (1 of 40) K04733 - interleukin-1 receptor-associated kinase 4 (IRAK4)	36.102	626	342	17	7	606	13	606	5.17E-100	321

At-CERK1	evm.model.Scaffold_608.648 (1 of 2414) 2.7.11.1 - Non-specific serine/threonine protein kinase / Threonine-specific protein kinase	35.229	633	353	17	7	617	13	610	6.72E-95	308
At-CERK1	evm.model.Scaffold_352.302 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	44.377	329	168	7	304	617	336	664	5.28E-82	275
At-CERK1	evm.model.Scaffold_462.277 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	44.073	329	169	7	304	617	302	630	5.30E-82	274
At-CERK1	evm.model.Scaffold_612.376 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	39.203	301	163	8	310	601	350	639	2.88E-55	201
At-CERK1	evm.model.Scaffold_952.320 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	38.333	300	165	8	310	600	350	638	3.77E-54	197
At-CERK1	evm.model.Scaffold_628.522 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	33.639	327	181	7	303	601	343	661	3.90E-51	189
At-CERK1	evm.model.Scaffold_628.522 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	35.606	132	79	5	66	195	85	212	4.18E-12	70,1
At-CERK1	evm.model.Scaffold_1841.91 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	33.639	327	181	7	303	601	343	661	1.29E-50	187
At-CERK1	evm.model.Scaffold_1841.91 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	35.938	128	76	5	66	191	85	208	6.44E-12	69,3

At-LYK4	evm.model.Scaffold_952.320 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	46.154	624	292	10	22	608	23	639	7.01E-153	459
At-LYK4	evm.model.Scaffold_612.376 (1 of 6) PTHR27001:SF174 - LYSM DOMAIN RECEPTOR-LIKE KINASE 4	46.154	624	292	10	22	608	23	639	1.57E-152	458
At-LYK4	evm.model.Scaffold_1841.91 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	39.506	324	145	5	336	610	342	663	1.92E-56	204
At-LYK4	evm.model.Scaffold_628.522 (1 of 2) PTHR27001:SF183 - PROTEIN LYK5	39.198	324	146	5	336	610	342	663	2.44E-56	204
At-LYK4	evm.model.Scaffold_638.859 (1 of 2414) 2.7.11.1 - Non-specific serine/threonine protein kinase / Threonine-specific protein kinase	29.412	612	379	15	32	601	24	624	6.42E-54	196
At-LYK4	evm.model.Scaffold_2627.64 (1 of 4) PF00069//PF01476 - Protein kinase domain (Pkinase) // LysM domain (LysM)	37.898	314	162	9	318	601	308	618	3.50E-41	159
At-LYK4	evm.model.Scaffold_2658.204 (1 of 4) PF00069//PF01476 - Protein kinase domain (Pkinase) // LysM domain (LysM)	37.261	314	164	9	318	601	308	618	2.06E-40	156
At-LYK4	evm.model.Scaffold_2193.164 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	33.226	310	177	7	332	612	298	606	5.40E-37	146
At-LYK4	evm.model.Scaffold_476.38 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	33.226	310	177	7	332	612	298	606	5.82E-37	146
At-LYK4	evm.model.Scaffold_608.131 (1 of 40) K04733 - interleukin-1 receptor-associated kinase 4 (IRAK4)	36.226	265	136	7	378	610	337	600	4.31E-35	140

At-LYK4	evm.model.Scaffold_352.302 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	34.496	258	136	5	386	610	392	649	1.20E-34	140
At-LYK4	evm.model.Scaffold_462.277 (1 of 2) PTHR27001:SF102 - LYSM DOMAIN RECEPTOR-LIKE KINASE 3	34.109	258	137	5	386	610	358	615	2.71E-34	138
At-LYK4	evm.model.Scaffold_539.592 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	31.935	310	179	8	333	611	512	820	2.29E-32	134
At-LYK4	evm.model.Scaffold_1805.113 (1 of 4) K13429 - chitin elicitor receptor kinase 1 (CERK1)	31.613	310	180	7	333	611	282	590	2.84E-31	129

ARTIGO 2

(Artigo redigido de acordo com as normas do periódico Frontiers in Plant Science)

Genome-wide identification, characterization, and comparative analysis of NLR resistance genes in *Coffea spp.*

Mariana de Lima Santos [1*], Mário Lúcio Vilela de Resende [1*], Gabriel Sérgio Costa Alves[2], Jeremy Todd Brawner[3], Jose Carlos Huguet-Tapia[3], Márcio Fernando Ribeiro de Resende Júnior[4]

[1]Laboratorio de Fisiologia do Parasitismo, Faculdade de Ciências Agrárias, Departamento de Fitopatologia, Universidade Federal de Lavras, Lavras, Minas Gerais, Brazil

[2]Laboratório de Processos Biológicos e Produtos Biotecnológicos, Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Brasília, Distrito Federal, Brazil

[3]Institute of Food and Agricultural Sciences, Department of Plant Pathology, University of Florida, Gainesville, Florida, FL, United States

[4]Sweet Corn Laboratory, Institute of Food and Agricultural Sciences, Horticultural Sciences Department, University of Florida, Gainesville, Florida, FL, United States

*** Correspondence:**
Corresponding Authors
santos-ml@outlook.com
mlucio@ufla.br

Keywords: resistance genes, Coffea, nucleotide-binding site leucine-rich repeat, Genome-wide, NLR-annotator.

Abstract

The largest family of disease resistance genes in plants are nucleotide-binding site leucine-rich repeat genes (NLRs). The products of these genes are responsible for recognizing avirulence proteins (Avr) of phytopathogens and triggering specific defense responses. Identifying NLRs in plant genomes with standard gene annotation software is challenging due to their multidomain nature, sequence diversity, and clustered genomic distribution. We present the results of a genome-wide scan and comparative analysis of NLR loci in three coffee species (*Coffea arabica*, *Coffea canephora*, and *Coffea eugenioides*). A total of 1311 non-redundant NLR loci were identified in *C. arabica*, 927 in *C. canephora*, and 1079 in *C. eugenioides*, of which 809, 562, and 695 are loci completes, respectively. The NLR-annotator, tool used in this study, showed sensitivities and specificities extremely high (over 99%) in the coffee genomes and increased the detection capability of putative NLRs. The NLRs loci in coffee are distributed among all chromosomes and are organized mostly in clusters. The *C. arabica* genome present a smaller number of NLR loci when compared to the sum of the parental genomes (*C. canephora*, and *C. eugenioides*). There are orthologs NLRs (orthogroups) shared between coffee, tomato, potato, and reference NLRs and those that are shared only among coffee species, which gives us clues about the functionality and evolutionary history of these groups. Phylogenetic analysis demonstrated orthologs NLRs

shared between *C. arabica* and the parental genomes and those that were possibly lost. The NLR family members in coffee are subdivided into two main groups: TIR-NLR (TNL) and non-TNL. The non-TNLs seem to represent an important repertoire of resistance genes in coffee. These results will support functional studies and contribute to a more precise use of these genes for breeding disease-resistant coffee cultivars.

Introdution

Coffee is a globally important agricultural commodity that plays a significant economic role in producing and consuming countries (Krishnan, 2017). The genus Coffea consists of more than 100 botanical species (Davis et al., 2006), however, the most cultivated species are *Coffea canephora* and *Coffea arabica*. *C. canephora* is diploid (2n = 2x = 22 chromosomes) (Denoeud et al., 2014), while *C. arabica* is a allotetraploid (2n = 4x = 44 chromosomes) (Tran et al., 2018) originated from natural hybridization between *C. canephora* and *C. eugenidoides* (Lashermes et al., 1999; Bawin et al., 2020). Among the more than 50 coffee-producing countries, Brazil, Vietnam, Colombia, and Indonesia are major producers, with Brazil being the largest producer by volume. Currently, coffee diseases are the main factor affecting productivity (Cerda et al., 2017). Examples of diseases associated with coffee include cercosporiosis (*Cercospora coffeicola*), bacterial blight (*Pseudomonas syringae* pv. *garcae*), anthracnose (*Colletotrichum coffeanum*), root-knot nematodes (*Meloidogyne spp*.), coffee berry disease - CBD (*Colletotrichum kahawae*), and coffee leaf rust - CLR (*Hemileia vastatrix*) (Cabral et al., 2016; Krishnan, 2017). CLR is one of the most devastating diseases found in coffee and is present in all world regions where coffee is grown (McCook and Vandermeer, 2015; Cabral et al., 2016). Currently, 95% of *C. arabica* varieties cultivated in Brazil are susceptible to CLR due to the emergence of variants of the pathogen. (Cabral et al., 2016). Given the increasing problem of plant pathogens in coffee production, a greater understanding of the set of receptors regulating the plant immune system of coffee is needed.

Throughout evolution, plants have developed sophisticated systems to defend themselves from pathogens. The plant immune system involves both broad-spectrum and specific recognition of pathogens. Broad-spectrum recognition is related to the detection of pathogen-associated molecular patterns (PAMP), such as fungal chitin or bacterial flagella, by pattern recognition receptors (PRR) that are anchored to the plasma membrane and trigger the PAMP-triggered immunity (PTI) (Boutrot and Zipfel, 2017). Specific recognition, on the other hand, primarily involves receptors encoded by resistance genes (R genes) that detect the presence of pathogen effector proteins and trigger effector-triggered immunity (ETI) (Jones and Dangl, 2006). Both types of recognition occur dynamically and continuously, converging into signaling pathways that activate essential mechanisms for downstream responses to pathogen recognition (Lu and Tsuda, 2021; Yuan et al., 2021).

The R genes have been extensively studied in several crops to facilitate their greater use in plant breeding. (Jupe et al., 2013; Wan et al., 2013; Lozano et al., 2015; Inturrisi et al., 2020; Steuernagel et al., 2020). The protein products of these genes recognize directly or indirectly the effectors secreted by pathogens (Kourelis and Van Der Hoorn, 2018) and trigger a series of signaling steps that lead to the hypersensitive response (HR) that activates cell death and potentially leads to systemic acquired resistance (SAR) (Kachroo and Robin, 2013; Jones et al., 2016). The largest and most diverse group of R genes found in plants belong to the nucleotide-binding site leucine-rich repeat family (NLR or NBS-LRR) (Jones et al., 2016). The proteins encoded by these genes are typically modular, and many have a variable N-terminal domain-containing Toll/interleucina-1 receptors (TIR) or coiled-coil (CC). As well, the nucleotide-binding domain (NB-ARC or NBS) is a canonical feature of NRLs, that is

shared with human apoptotic protease-activating factor-1 (*APAF-1*) and *Caenorhabditis elegans* death-4 (*CED-4*) proteins. A C-terminal region comprising a variable number of leucine-rich repeats (LRRs) is another common feature in NLR genes (Jones et al., 2016; Shao et al., 2019). The NB-ARC domain is highly conserved and is involved in the active and inactive state of the NLS protein (Bonardi et al., 2012; Jones et al., 2016). This domain presents motifs characteristic of the ATPase family, such as p-loop and kinase 2, besides the RNBS (Resistance Nucleotide Binding Site) A, RNBS-C, and RNBS-D motifs (Van Ghelder et al., 2019). Mutations in specific residues within these motifs may cause the loss of protein function or self-activation and interfere with the regulation or activation of defense mechanisms. (Monteiro and Nishimura, 2018; Bezerra-Neto et al., 2020).

The domains described above provide a more refined classification of NLRs proteins, which may be characterized into two main groups: TNLs (TIR-NLRs) or non-TNL (which include CNLs - CC-NLRs). The truncation of a single domain, such as LRR (CN or TN), TIR or CC (NL), in both C and N terminal domains (N), may also be used to classify NLR genes (Monteiro and Nishimura, 2018). Additionally, atypical or non-canonical integrated domains (IDs) that act as decoys and play roles in oligomerization or downstream signaling may be present, demonstrating the structural diversity of this NLR family (Kroj et al., 2016; Wang et al., 2021). The number of NLRs in plant genomes varies greatly and is often organized in tandem, which facilitates duplication, contraction, and transposition and provides a reservoir of genetic variation that allows plant evolutionary dynamics to respond to changes phytopathogen populations (Barragan and Weigel, 2021). These genes are often under selection pressure, resulting in a large number of pseudogenes and variable loci content within the same species, among species, and across plant populations (Schatz et al., 2014; Steuernagel et al., 2015; Sun et al., 2020; Hufford et al., 2021).

The knowledge of how NLRs are distributed throughout the genome and their diversity is of great interest. It may reveal new sources of resistance that may be used to develop new cultivars (Monteiro and Nishimura, 2018). The growing number of sequenced plant genomes facilitates the search for novel NLR and has led to the genome-wide analysis of NLR genes (Denoeud et al., 2014; Song et al., 2015; Scott et al., 2020; Wang et al., 2021). However, its large number, frequently clustered genomic distribution, and low expression in uninfected tissues make the cataloging NLR genes a great challenge and often underestimate the number of NLRs in genomes (Jupe et al., 2013; Steuernagel et al., 2015, 2020). To mitigate this problem, some specific gene/loci NLR annotation pipelines have been developed to augment standard gene annotation software and improve our ability to identify and locate genes belonging to this family. Some examples of these pipelines are NBSPred (Kushwaha et al., 2016), NLGenomeSweeper (Toda et al., 2020), and NLR-annotator, a new version of NLR-parser (Steuernagel et al., 2015, 2020). The NLR-Annotator is a tool used to annotate NLR loci that use 20 highly curated motifs present in NLR proteins and does not depend on the support of transcript data (Jupe et al., 2012; Steuernagel et al., 2020). Since it was published, this tool has been applied in several studies to prospect and annotate R genes in recently sequenced genomes (Muliyar et al., 2020; Read et al., 2020; Scott et al., 2020), to checking the completeness of previous annotations (Muliyar et al., 2020), and for studies of the resistance-related locus (Jost et al., 2020).

The genome of *C. canephora* was published in 2014, which allowed the first genome-wide NLR study in coffee (Denoeud et al., 2014). In 2018, the *C. arabica* and *C. eugenioides* genomes were deposited at the NCBI, providing an essential resource for studying the structure and evolution of NLRs in arabica coffee and the contribution of the genomes that gave rise to this species. For coffee production to continue advancing in producing regions

worldwide, adequate disease management is of great importance. A range of strategies must be used to control the main phytosanitary problems. Using these genomic resources is essential for informing breeding strategies aimed at developing resistance to disease in coffee. Given the above, this study aimed to: (i) identify NLR loci in *C. arabica*, *C. canephora*, and *C. eugenioides* genomes using the NLR-annotator tool and discuss the improvements in annotation derived from the use of a specific pipeline for NLR genes in coffee, (ii) catalog, classify and characterize the distribution of NLRs loci in the *coffee spp.* genomes, and (iii) understand the contribution of *C. canephora* and *C. eugenioides* to the NLRs repertoire of *C. arabica*.

Materials and Methods

Coffee Genomic Resources

Three genomes were used in this study. The *C. arabica* (Caturra red - Cara_1.0, GenBank assembly accession: GCA_003713225.1) and *C. eugenioides* (Ceug_1.0, GenBank assembly accession: GCA_003713205.1) genomes are available from the NCBI (National Center for Biotechnology Information) database (https://www.ncbi.nlm.nih.gov/) and the *C. canephora* genome is available at Coffee genome hub (https://coffee-genome.org/) (Denoeud et al., 2014). For the three species, the genome files, set of predicted proteins, predicted genes, and GFF (General Feature Format) were used.

Identification of loci NLR in Coffea spp. genomes

The identification of NLR loci in *Coffea spp.* was accomplished using the NLR-Annotator, (Steuernagel et al., 2020) using the default parameters. The tool uses combinations of short motifs of 15 to 50 amino acids to classify a genomic locus as an NLR. These motifs were defined using domains of known NLR proteins used as a training set in a study carried out by Jupe et al. 2012 (Supplementary Table 1).

In summary, the pipeline is divided into three steps: 1) dissection of genomic input sequence into 20-kb fragments overlapping by 5 kb; 2) translating each fragment into all six reading frames and searching for the motifs associated with NLR by NLR-Parser that to create an xml-based reporting file. The NLR-Parser searches for combinations of doublets or triplets of motifs, disregarding motifs that occur randomly. In this step, the positions of each motif are transferred to the nucleotide positions, and 3) the NLR-Annotator uses the xml file as input, integrates data from all fragments, evaluates positions and combinations of motifs. In this step, the NB-ARC motifs are used as a seed to annotate NLR loci (Supplementary Table 1), generate output files (.gff, .bed - Browser Extensible Data, .txt and file of the NB-ARC motifs as multiple alignments to complete loci) based on coordinates and orientation the initial input genomic sequence (Steuernagel et al., 2020).

Each section of the genomic sequence associated with a single NLR is called an 'NLR locus' and refers to an NB-ARC domain (or associated motif) followed or not by one or more leucine-rich repeats (LRRs). From the sets of motifs that are identified, these loci are classified as complete (containing the P-loop, at least three consecutive NB-ARC motifs, and at least one LRR), complete (pseudogenes), partial and partial (pseudogenes) (Supplementary Table 1). Therefore, the NLR-Annotator identifies the NLR loci that are either active genes or pseudogenes. The number of NLR loci and their classification is described in the output file.txt. For more details, see at Steuernagel et al., 2020 and https://github.com/steuernb/NLR-Annotator.

Validation of sensitivity and specificity of the NLR-annotator in coffee genomes

To validate the sensitivity and specificity of the NLR-annotator in the coffee genomes, we initially classified the protein sequences of the three genomes using PfamScan (https://pfam.xfam.org/) version 1.5 with an e-value of less than 1E- 5 and models from Pfam-A. Subsequently, proteins that had the NB-ARC domain (PF00931) were filtered, and from this process, we obtained the ID of the genes corresponding to each protein. With the list of gene model IDs of the NLR family, it was then possible to filter the GFF files and obtain the positions of the genes that had already been annotated in each genome.

We look for overlapping intervals to compare the NLR loci detected by NLR-Annotator and the NLR genes annotated in the genomes. We used the information from .gff files from both annotations for an overlay analysis using bedtools intersect (version 2.29.2). An overlap was only considered if both, the locus, and gene, were on the same strand. This analysis made it possible to distinguish the loci identified by NLR-annotator that were or were not overlapping on the gene models from reference genomes.

For NLR genes already annotated in the genomes and not identified by NLR-annotator, a search for motifs by NLR-Parser was performed to obtain the xml and txt output (options -c and -o) as well the detection of conserved domains using NCBI Conserved Domain Database (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) for nucleotides sequences. Standard parameters were used for the conserved domains analysis, except for the threshold (E-value), which was set to 1E-5. The Graphical summary was set to provide a concise view of the results. To characterize the NLR loci only found by NLR-annotator and to make sure they were homologous with NLRs already annotated in plants, we aligned these loci sequences with NCBI's non-redundant protein database (nr) (https://blast.ncbi.nlm.nih.gov) using BLASTx (BLAST - version 2.10.1 with max_target_seqs 5). For loci that did not have homology with NLRs proteins, a conserved domain analysis was also performed as previously described.

The sensitivity of the pipeline was calculated as the ratio of the number of loci identified by NLR-annotator (including motifs detected by NLR-Parser in the second step of the pipeline) to the number of NLRs genes already annotated in the genomes. The specificity was calculated as the ratio of the number of loci identified by NLR-annotator that are related to NLRs genes or presents characteristic domains of that family (overlap with the annotations already described in the studied genomes, homology with NLR proteins by BLASTx or NB-ARC domains by conserved domains analysis) to the total number of loci identified.

Distribution of NLR loci in Coffee's chromosomes

For visualize the distribution of NLR loci on chromosomes of the three analyzed coffee species, the annotation file from NLR annotator (.txt) were used to extract the genomic position and classifications of the loci. The chromosomes size information in Mb was obtained from the NCBI (for *C. arabica* and *C. eugenioides*) and Coffee genome hub (for *C. canephora*). The visualization was performed using the R software with the chromoMap package. (Anand and Lopez, 2020). The chromoMap, divides the chromosomes as a continuous composition of loci. Each locus, consist of a specific genomic range determined algorithmically based on chromosome length and then the annotations are inserted. The detailed annotation information on each locus NLRs (complete, complete (pseudogene), partial and partial (pseudogene)) is displayed in an HTML file.

Prediction of genes in the complete loci found only by the NLR-annotator

Gene prediction was performed using the AUGUSTUS program version 3.3.3 (http://bioinf.uni-greifswald.de/augustus/) (Stanke et al., 2006) using gene models from *Solanum lycopersicum* and allowing for the prediction of only complete genes.

Orthologous groups and Phylogenetic Analyses

The complete loci identified in the coffee genomes, being those that overlap with gene models of the reference genomes or that were annotated by AUGUSTUS as putative genes, were selected for ortholog and phylogenetic analysis. In order to make a comparison with the set of coffee NLRs, 326 NLR loci identified in tomato (*Solanum licopersicum* - Heinz 1706) by Andolfo et al., 2014, 755 loci identified in potato (*Solanum tuberosum*) by Jupe et al., 2013, 67 NLR reference genes (functionally characterized protein) obtained from The Plant Resistance Genes database - PRGDB (http://prgdb.org/prgdb/, S2 table) (Osuna-Cruz et al., 2018) and the *CED-4* gene from *Caenorhabditis elegans* (outgroup) were also added. All these sequences were classified according to the rules of motifs established by the NLR-annotator and only those considered as complete NLR were used in the analysis. This criterion was used to standardize the methodology for classifying loci as complete or not.

The amino acid sequences of the NB-ARC domain were extracted from the set of complete NLR loci for the 5 species compared (*C. arabica*, *C. canephora*, *C. eugenioides*, *S. licopersicum* and *S. tuberosum*) and the reference genes by the NLR -annotator (parameter -a). All NB-ARC domain of these complete loci (hereafter called as NLRs) were compared with each other using BLASTP all-by-all (E value <1e-10). The markov clustering algorithm was performed with inflation value of 1.5 and then NLRs in the same cluster were classified as orthologous subgroups by OrthoMCL version 1.4 (standard parameters) (Li et al., 2003). In order to analyze and visualize the number of orthogroups shared between the species or the ones that are unique to a single species we used the UpSetR package in R (Conway et al., 2017).

The NLRs clustered into single-copy orthogroups (orthogroups that have one copy of each NLR present once in each of the 5 genomes or reference NLRs) by OrthoMCL were used to construct a phylogenetic tree. The sequences were aligned by MAFFT software version 6.903 (Katoh et al., 2002), using --auto parameter to select the best alignment strategy. The tree was inferred using RAxML version 8.2.10 (Stamatakis, 2014) with the PROTGAMMAAUTO model (the JTT model was selected as having the highest likelihood) with 100 bootstrap replicates. A second phylogenetic tree to classify the coffee NLRs was also constructed with the above-mentioned parameters using the entire set of complete NLRs. Coffee NLRs were classified in the tree, from the previous classification described to 67 reference NLRs (Supplementary Table 2) and the tomato and potato NLRs (Jupe et al., 2012, 2013; Andolfo et al., 2014). The trees were visualized and edited using the Interactive Tree of Life (iTOL) tool (Letunic and Bork, 2021).

Results

NLRs Identification, validation of the sensitivity and specificity of NLR-annotator in coffee genomes

A total of 1318 loci were identified for *C. arabica*, 932 for *C. canephora*, and 1081 for *C. eugenioides* (Supplementary Table 3). In each species, we identified some loci that are in the same position but were separated by the NLR-annotator as two distinct NLRs. We found 7, 5, and 2 repeated loci in the species mentioned above, respectively (Supplementary Table 3

highlighted in blue). Considering these repeated loci, for counting the number and distribution of the NLRs loci on the chromosomes, the most complete classification was considered. After the identification of these regions, it was found that there were 1311 non-redundant loci for *C. arabica* (627 from the *C. canephora* subgenome - CaC, 650 from the *C. eugenioides* subgenome - CaE and 34 Unassigned - Un), 927 for *C. canephora* (559 mapped on chromosomes and 367 Un) and 1079 for *C. eugenioides* (944 mapped on chromosomes and 135 Un). The number of complete loci found in each species was 809 (*C. arabica*), 562 (*C. canehora*), and 695 (*C.eugenioides*).

To examine whether there was a consensus between the gene models for NLRs that have previously been annotated in the genomes and loci identified by NLR-annotator, an overlap analysis was performed. PfamScan analysis on the set of predicted proteins and subsequent selection of NLR proteins annotated in each genome showed that 1015, 709, and 869 genes encoded proteins (including isoforms) with the NB-ARC domain in the *C. arabica*, *C. canephora*, and *C. eugenioides* genomes, respectively (Figure 1, Supplementary Table 4). The overlap between the genomic positions of these genes and the positions of loci from NLR-annotator showed that of 1311, 927, and 1079 loci identified by NLR-annotator for *C. arabica*, *C. canephora*, and *C.eugenioides* respectively, 1013 (99,80%), 687 (96,90%), and 857 (98,62%) overlaps the genes already annotated in the reference genomes. It was also detected that 298, 240 and 222 NLRs do not overlap (Figure 1, Table 1). We also noticed that there are cases in which more than one NLR loci overlapped with a single NLR gene and that the opposite was also found in all three genomes. A Venn diagram representing these data is shown in figure 1 (intersection Figure 1 and highlighted in Supplementary Table 5).

The overlap analysis also made it possible to identify genes annotated in the reference genomes that did not overlap with any locus from NLR-annotator. To examine these genes, an NLR-parser analysis with the options -c (file.xml) and -o (file.txt) was performed on this set. Among the genes not identified by NLR-annotator for *C. arabica* (18), *C. canephora* (25), and *C. eugenioides* (24), 7, 3, and 4 were below the standard threshold (1E-5) for the search of MAST motifs by NLR-Parser, respectively. Additionally, 9, 17, and 16 genes present motifs detectable by the standard threshold but did not contain at least three consecutive motifs belonging to the NB-ARC domain and motifs in random order and therefore were not annotated in the third step of the NLR-annotator (Supplementary Table 6). After this analysis, we also confirmed genes not found by NLR-annotator. Two genes were not found for *C. arabica* (LOC113737176 and LOC113735982), five genes for *C. canephora*, (Cc02_g12220, Cc03_g10360, Cc07_g18800, Cc00_g21910 and Cc00_g35420) and four genes for *C. eugenioides* (LOC113766771, LOC113766774, LOC113766615 and LOC113777141). For more details about this analysis, see supplementary text 1, supplementary table 6 and supplementary figure 1. After these analyses, it was possible to verify that the NLR annotator showed a sensitivity of 99.8%, 99.4%, and 99.7% for *C. arabica*, *C. canephora* and *C. eugenioides*, respectively.

As stated above, the overlap analysis also made it possible to identify that the NLR-annotator annotated complete, complete (pseudogene), partial and partial (pseudogene) loci that did not overlap with genes already annotated in reference genomes (Figure 1, Table 1). To further investigate these loci and make sure that they were indeed related to genes encoding NLR proteins, a BLASTx analysis was performed. This analysis showed that of the 298, 240, and 222 loci in *C. arabica*, *C canephora*, and *C. eugenioides*, only 7, 4, and 6 did not show homology, with resistance proteins being found among the five best hits, respectively. (Supplementary Table 7, 8, and 9, highlighted in orange).

To describe the sequences that did not show homology to NLRs proteins by BLASTx, a conserved domains analysis was performed (Supplementary Figure 2). Many of these loci do not show homology with NLRs proteins because most of the sequence contains domains related to the family of proteins involved in the activity of transposable elements such as ribonuclease H (RNase H) and reverse transcriptases (RTs). However, it was also possible to identify characteristic domains of NLR proteins such as NB-ARC, Toll / interleukin-1 receptor (TIR), RX-CC_like, and Rx_N, suggesting that these loci cannot be considered false positives. Only three loci did not present characteristic domains, Chr11c_nlr_73_Ca, chr0_nlr_300_Cc and Chr8_nlr_67_Ce, all partial (pseudogene). These loci were removed from further analysis. From these results, it was possible to verify that the specificity of the NLR-annotator was 99.9% in all three genomes.

Distribution of NLR loci in the Coffea spp. Genome

Considering all detected loci, in *C. canephora*, chromosomes 3 and 11 have the most significant number of identified loci, including complete, complete (pseudogene), partial and partial (pseudogene). For *C. eugenioides*, chromosomes 3 and 11 also contains many NLR, this large number was also found in the chromosomes 5 and 8. The total amount of loci found on chromosome 8 is almost the same as that of 11, however, the number of complete NLR on chromosome 8 is higher, with 16 more loci identified. For *C. arabica*, chromosomes 3 and 11 from *C. canephora* and *C. eugenioides* subgenomes, respectively, also have the largest number of NLR, what was also detected to the chromosome 8 from *C. eugenioides* subgenoma. For *C. arabica*, in general the *C. eugenoides* subgenome has a slightly higher number of NLRs loci, as already reported above. The number of loci of this subgenome on chromosomes 8 and 11 stand out in comparison to *C. canephora* subgenome, with 34 and 30 more loci, respectively. The chromosomes with the fewest loci for the three species are 9 and 10, and the chromosome 4 for *C. eugenioides*. The number of loci in unmapped sequences (Unassigned) for the *C. canephora* reference genome represent 39.7%, which was much higher than those found in *C. eugenioides* (12.5%) and *C. arabica* (2,6%) (Figure 2A).

The chromosomal location of these loci in the three species demonstrated that most loci are organized in clusters and are unevenly distributed across the entire chromosome. In addition, there are clusters that have the four different types of loci or at least two types, presenting a stretch of complete, complete (pseudogene), partial and/or partial (pseudogene). Not all loci were clustered, we also found loci of the four types that were physically isolated in chromosomes (Figure 2B).

Gene prediction for complete loci found only by NLR-annotator

It was considering that the NLR-annotator is not a gene predictor but a tool to annotate loci associated with NLRs, the gene-finding program AUGUSTUS was used to characterize the loci found only by NLR-annotator that were classified as complete (S3 Table highlighted in orange). This analysis aimed to verify whether these complete loci could be considered potential gene models. This analysis showed that of the 70 and 67 complete loci for *C. arabica* and *C. canephora*, 64 and 66, are potential gene models, respectively. For *C. eugenioides*, all 71 loci were predicted as potential genes. The loci not identified as potential genes are in red in supplementary table 3.

Ortholog Groups and Phylogenetic Analysis

To ortholog groups analysis by OrthoMCL, 803 complete loci of *C. arabica*, 561 of *C. canephora* and 695 of *C. eugenioides* were used. Six and 1 loci of *C. arabica* and *C. canephora*, respectively, were removed from analysis because they are complete loci that are not overlapping gene models or were not identified as putative genes by AUGUSTUS analysis. Additionally, 151 tomato loci (out of 326) and 403 potato loci (out of 755) that were classified as complete loci by NLR-annotator, moreover 67 reference NLRs and *CED-4* were also used. Out of a total of 2681 NLRs, 2038 (76%) were grouped into 593 ortholog groups, hereinafter called orthogroups (Figure 3, Supplementary Table 10). The number of coffee NLRs present within these orthogroups were 591, 427, 584, which represents 73.6%, 76.1% and 84% of the total NLRs found for *C. arabica, C. canephora* and *C. eugenioides*, respectively. Two hundred and seventy-two orthogroups were in single-copy, grouping 647 NLRs, of which only 7 are reference NLRs.

There were 10 orthogroups shared by all species and reference NLRs and 11 were shared only among species. As expected, the greatest number of orthogroups were shared among coffee NLRs, 200 orthogroups with 783 NLRs were shared only among *C. arabica* (296: 163 CaE, 130 CaC e 3 un), *C. canephora* (215) e *C. eugenioides* (272), respectively. The comparison between *C. arabica* NLRs with only one of the diploid species showed that *C. eugenioides* shares a slightly higher number of orthogroups (78) than *C. canephora* (71) and also of NLRs within these orthogroups (orthogroup Ca/Ce = 87/96 NLRs, orthogroup Ca/Cc = 86/74 NLRs). When the comparison was only between the two diploid species, it was observed that 34 orthogroups are shared only between them. The number of orthogroups shared between NLRs of the same coffee species was 31 in *C. eugenioides*, 24 in *C. arabica* and 9 in *C. canephora*.

The orthogroups NLR shared only between the three coffee species and one solanum specie, was higher among potato (9) than tomato (3) NLRs, however it should be considered that the number of potato NLRs in the analysis was almost 3 times bigger than tomato. Forty-six orthogroups have gathered NLRs from at least one coffee species and one solanum species. Fifteen orthogroups were shared between reference NLRs, and at least one coffee specie and solanum, grouping 21 reference NLRs. (Figure 3, Supplementary Table 10, highlighted in light blue), of these, *C. canephora* and/or *C. eugenioides* are present in three orthogroups with reference genes in which *C. arabica* is absent (ORTHOMCL16: Cc, Soly, Soltu e *Hero*; ORTHOMCL17: Ce, Soly, Soltu e *Rpiblb1*; ORTHOMCL24: Cc, Ce, Soly, Soltu e *VAT*). Four orthogroups clustered only the three coffee species and reference NLRs (ORTHOMCL1, ORTHOMCL19, ORTHOMCL119 and ORTHOMCL199, Figure 3 and Supplementary Table 10, highlighted in dark blue), gathering *Lr10*, *MLA1*, *MLA10*, *MLA13*, *Mla12*, *Mla6*, *Pi36*, *Pikm2TS*, *FOM-2*, *Rdg2a* e *Pm3*. The percentage of orphans (i.e., NLRs not assigned to any ortholog group by OrthoMCL) among coffee NLRs was highest in *C. arabica* (26.4% -212) followed by *C. canephora* (23.9% - 134) and *C. eugenioides* (16% -111) (Supplementary Table 11).

The phylogenetic tree of single-copy orthologous NLRs showed that most clades are shared only among coffee species (Figure 4), but it was also possible to observe clades that clustered NLRs from solanum, coffee and reference. Among the clades that clustered coffee NLRs, in 71 of them it was possible to observe groupings of orthologs between *C. arabica*, *C. canephora* and *C. eugenioides*, and most of which are located within the same chromosome. One of these clades, in addition to grouping NLRs of the three coffee species, presents the reference NLR *RPS2* (RESISTANCE to P. SYRINGAE 2) (Figure 4). This clade was supported by high bootstrap value (100%) and were grouped in the ORTHOMCL45 (Supplementary Table 10). All loci in this cluster were found on chromosome 6 for the three

species. Clades that clustered NLRs of *C. arabica* and *C. canephora*, *C. arabica* and *C. eugenioides* and a few *C. canephora* and *C. eugenioides* can also be observed. These are within the same chromosome or not.

Phylogenetic analysis for coffee NLRs classification revealed that members of the NLR superfamily are grouped into 2 main groups: TIR-NLR (including TNL and NLs) and non-TNLs (including CNLs and NLs) (Figure 5). NLRs belonging to the non-TNLs group outnumbered those in the TNL group in coffee genomes. We also found that the non-TNLs group is divided into 13 subgroups and that all subgroups had NLRs from all studied coffee genomes. It also occurred in the TNL group. Within non-TNLs subgroups it was possible to observe clades with a greater number of NLRs of *C. arabica* that are shared with *C. eugenioides* (bands on the outer ring of the tree with a predominance of green and blue colors). There were exclusive coffee clades and those that were shared with potato, tomato, and reference NLRs.

Discussion

NLRs Identification and use of NLR-annotator in coffee genomes

In this study, we investigated loci related to genes of the NLR family in three coffee genomes. The annotation of genes in this family is a high priority in plant genome sequencing and annotation projects (Steuernagel et al., 2015) because losses from pathogens are among the main problems for sustainable agriculture. A better understanding of disease resistance in crops will provide plant breeders with tools that may be used to produce long-term solutions for dealing with future environmental change. A catalog of NLRs loci, whether complete genes or pseudogenes, within and between species, provides a toolbox for exploring NLRs hitherto not described (Jones et al., 2016). Given the importance of coffee and the availability of the recent *C. arabica* and *C. eugenioides* genomes, the study of NLRs loci in these species represents an essential source of information for the development of new disease-resistant cultivars.

As NLR-annotator has already been validated in *C. canephora* (Steuernagel et al., 2020), we initially used this genome to assure the reproducibility of the tool in our study and then use it in the *C. arabica* and *C. eugenioides* genomes. The software identified 932 loci NLR for *C. canephora*, which is in accordance with Steuernagel et al., 2020, and the prediction of two distinct NLRs loci within the same genomic position was also reported. This repeatable annotation is the result of a complete NB-ARC domain preceded by a truncated NB-ARC domain, which makes the tool use the two NB-ARC domains as distinct seeds to identify two NLRs for the same locus, a fact that has also been reported with the use of this tool on wheat (Steuernagel et al., 2020). The sensitivities and specificities of this tool in coffee genomes were extremely high (above 99%). In the *Arabidopsis thaliana* genome, which represents a well-annotated model plant genome, this tool achieved 95% sensitivity, and all loci found were validated to be associated with NLRs (specificity of 100%) (Steuernagel et al., 2020). In rice (Nipponbare reference genome), the detection success rate was 99.2% (Read et al., 2020).

As NLR genes have repeated and clustered genomic distribution in plants, their representation in genomes using standard gene callers can be underestimated (Jupe et al., 2013; Steuernagel et al., 2015, 2020). In addition to the high rate of specificity and sensitivity, the NLR-annotator allowed for the identification of complete loci for coffee in regions distinct from the gene models already annotated in the reference genomes. This study, therefore, increased the detection capability the number of possible NLR genes in coffee species for the reference

genomes used. The complete loci identified by NLR-annotator that did not overlap the reference genome annotation have also been reported in rice (Read et al., 2020). It is also relevant to highlight that as this tool is not limited to searching for functional genes, the complete (pseudogene) loci that did not overlap the annotations of the reference genome were also identified for *C. arabica* (90), *C. canephora* (56), and *C eugenioides* (65). The location of these loci also represents an important resource since non-functional alleles identified in sequenced accessions may represent functional NLRs in other individuals of the same species (Steuernagel et al., 2020). The caturra cultivar (*C. arabica*) sequenced and used in this study, for example, is used as a susceptible control to differentiate *Hemileia vastatrix* races among differential clones (Zambolim and Caixeta, 2021). Pseudogene regions in this genome may indicate functional genes in other coffee cultivars.

Our data showed that 18 of the 20 loci found only by NLR-annotator that did not present homology to NLRs proteins by BLASTx analysis, have protein domains involved in the activity of transposable elements (TE). It is known that TE are abundant in plant genomes and that they play an important role in adaptive evolution, contributing to the evolutionary dynamics of plant-pathogen interactions (Malacarne et al., 2012; Zhang et al., 2014; Kim et al., 2017). Many R genes are flanked by TE, which in addition to being sources of genetic variability, are involved in suppressing or increasing the expression of these genes (Seidl and Thomma, 2017). The Ty3-gypsy-like TE performance has been reported in a region around the S_H3 locus for CLR resistance. This TE has been described in *C. arabica* subgenomes, replacing the orthologous counterpart of *C. canephora* to *C. eugenioides* (homoeologous non-reciprocal transposition - HNRT) (Cenci et al., 2012). Moreover, there is evidence of functional R genes that have evolved through TE-mediated duplications (Seidl and Thomma, 2017), which demonstrates their importance in the evolutionary changes and expansion of NLR receptors of the plant defense system and justifies the presence of domains related to TE in the studied loci (Zhang et al., 2014; Kim et al., 2017).

Distribution of NLR loci in the *Coffea spp.* Genome

Although *C. arabica* results from the natural interspecific hybridization between *C. canephora* and *C. eugenioides*, the number of loci found was not proportional to the sum of the two subgenomes, showing that this species has a smaller number of NLRs loci. The *C. canephora* genome size is about 690 Mbp, and the *C. eugenioides* is 665 Mbp (Noirot et al., 2003; Clarindo and Carvalho, 2011; Hamon et al., 2015). The *C. arabica* genome, on the other hand, is slightly smaller than the sum of its two combined parental genomes (about 1276 Mbp) (Hamon et al., 2015). This may explain the smaller number of NLRs in this species. Genome contraction is common in amphidiploids, which may be related to chromosomal rearrangements, including duplication, insertions, and deletions after initial hybridization (Hamon et al., 2015). An example of the number of NLRs reduced compared to the sum of the corresponding parents was reported in *Brassica juncea* (Indian mustard), a species formed by hybridization between the diploid Brassica species, *B. rapa,* and *B. nigra* (Inturrisi et al., 2020). Moreover, differences in the genome assembly quality may also have interfered with the identification of NLR loci.

Among the three species analyzed, the only one with a genome-wide NLR study already reported is *C. canephora* (Denoeud et al., 2014). The NLR gene data from that study agrees with much of our findings. A large number of NLR loci in unanchored scaffolds for this species has also been described. Here 210 complete NLR loci were identified in unanchored scaffolds for *C. canephora*, while in the first description of the manually curated genes, 213 were not mapped. The number of mapped NLR genes was 348, while in our study, there were

352 complete loci. In *C. canephora*, it has also been reported that NLRs genes are located on all chromosomes, but with an increased number found on chromosomes 1, 3, 5, 8, 11, which together represented 70.1% of the mapped NLR genes. Although we have highlighted chromosomes 3 and 11 as having a greater number of NLR, chromosomes 1,3, 5, 8 and 11 also present large numbers of NLR loci in the three species studied here, together representing 68.2, 71.4 and 70.0% of the total of NLR loci mapped for *C. arabica, C. canephora*, and *C. eugenioides*, respectively. Moreover, the few NLR genes on chromosomes 9 and 10 have also been previously reported (Denoeud et al., 2014). These comparisons show that the three species display a conserved pattern to the chromosomal distribution of NLR loci.

The NLR loci found for the three studied coffee species are arranged in clusters that group complete loci, pseudogenes and partial. These genes tend to group providing birth-and-death events for functional NLRs (Ling et al., 2021). In these clusters it is possible to find tandem gene duplications, recombination hotspots or active transposon elements functioning as a reservoir of genetic variation to generate specificity for new pathogen variants (Michelmore and Meyers, 1998; Zhang et al., 2014). Within plant genomes many R genes have been found to reside in clusters (Jupe et al., 2012; Andolfo et al., 2014, 2021; Seo et al., 2016; Zheng et al., 2016; Read et al., 2020). The S_H3 locus in coffee, for example, corresponds to a complex cluster of multiple genes, including CNL-like NLR genes (Ribas et al., 2011; Cenci et al., 2012). The number of complete or functional loci in plants represents a fraction of the total number of loci found (Jupe et al., 2012; Seo et al., 2016). This happens precisely because the evolutionary dynamics within these clusters favor the coexistence of functional genes, pseudogenes, and partial genes, which differ between plants in consequence of evolutionary routes for certain pathosystems.

Recent discoveries show that NLRs can be multi domain receptors, that is, present domains integrated to the canonical form NLR or TNL/CNL (Bailey et al., 2018; Wang et al., 2021). Knowing regions of the genome that have this canonical form can facilitate the description of non-canonical integrated domains that are upstream or downstream from the more conserved region (Monteiro and Nishimura, 2018). Another relevant information is that activation of NLRs often happens in complexes and there is evidence that truncated NLRs can form heterocomplexes with complete NLRs, or act as main receptors in defense activation (Monteiro and Nishimura, 2018). NLRs truncated as *CbCN* (*Capsicum baccatum* – CC-NB-ARC), and *TN2* (TIR-NB-ARC) act in resistance to pathogens (Zhao et al., 2015; Son et al., 2021). This evidence reinforces the importance of knowing loci related to R genes, whether complete, pseudogenes or partial.

Ortholog groups and Phylogenetic Analyses

The genus Coffea belongs to the Rubiaceae family, which is in the asterid clade, as well as to the Solanaceae family. Many studies use species of the genus Solanum to obtain insights about the genomic and evolutionary architecture of coffea (Lin et al., 2005; Lefebvre-Pautigny et al., 2010; Denoeud et al., 2014). Species of the genus Solanum are also used as a model for understanding the molecular processes related to plant-pathogen interaction, which makes us understand that comparative approaches with these species can lead to discoveries of NLR loci or functionally important gene families in coffee (Andolfo et al., 2021). Our results showed that of the 17 reference NLRs cloned and characterized in species of the genus Solanum and used in this study, 8 of them are present in shared orthogroups with coffee (Supplementary Table 2 and Supplementary Table 10), being 1 TNL (*Gro1-4)* and 7 CNL (*Hero, Prf, Rpi-blb1*, *Rx2*, *Sw-5*, *Tm-2a*, *Tm-2*). These genes are involved in the mechanism of resistance to a diverse group of pathogens including viruses, oomycetes, bacteria and

nematodes (Bendahmane et al., 2000; Van Der Vossen et al., 2003; Paal et al., 2004; Andolfo et al., 2021). In total 46 orthogroups were shared between coffee and solanum indicating that these orthologs were probably present before the speciation of these two groups. Reference NLRs characterized in species such as *Hordeum vulgare*, *Oryza sativa*, *Triticum aestivum*, *Glycine max*, *Arabidopsis thaliana* e *Cucumis melo* also share othogroups with coffee NLRs, all of which belong to the CNL class (Supplementary Table 2 and Supplementary Table 10). These orthogroups are important as it can indicate possible roles to be investigated. An interesting orthogroup and that obtained a high support in the phylogenetic tree was the one that clustered the *RPS2* reference NLR and NLRs present in the three coffee species. The *RPS2 i*s a resistance gene of *Arabidopsis thaliana* that confers resistance against *Pseudomonas syringae* bacteria that express avirulence gene *avrRpt2* (Bent et al., 1994; Mindrinos et al., 1994).

In general, plant species exhibit differences in the number of NLR genes. Amplification of certain groups is also detected (Seo et al., 2016). This diversity has not been associated with genome size or phylogenetic relationships, but rather as a consequence of the specialization of each particular host (Wan et al., 2013; Lozano et al., 2015; Seo et al., 2016; Zheng et al., 2016; Steuernagel et al., 2020). In all three coffee species, a set of orphan genes and orthogroups that share NLRs within the same species were detected. In tomato, 45 of ~320 NLRs sequences are more similar to each other than to any other sequences compared (Andolfo et al., 2021) and orthogroups that share NLRs within the same species in solanum are attributed to duplication events that generate different gene repertoires resulting in species-specific subfamilies (Seo et al., 2016).

The single-copy ortrogroups provide more reliable results of the evolutionary processes between groups of evaluated genes, making it possible to identify true orthologs between different groups of plants (Zimmer et al., 2007; Duarte et al., 2010). The results from single-copy ortrogroups tree showed that some orthologous NLR seem to have a common ancestor only among coffee species. The *S_H3* locus, for example, was described as being shared only among coffee species suggesting that the ancestral copy *S_H3*-CNL was inserted into the *S_H3* locus after the divergence between the Solanum and Coffea lineages (Ribas et al., 2011). The clades that clustered orthologous NLRs from *C. arabica*, *C. canephora* and *C. eugenioides* probably represent NLR present in both ancestral diploids genomes, and that were passed to *C. arabica* genome. On the other hand, clades that clustered only *C. canephora* and *C. eugenioides* may represent ancestral NLRs that were lost in *C. arabica* or that underwent so many modifications in this species that it makes it difficult to find homology between these NLRs. Nucleotide level changes, such as deletions, insertions and rearrangements were observed in coffee RGA (Resistance gene analogues) (Noir et al., 2001; Hendre et al., 2011). It is also known that it is very likely that the sequenced genotypes of *C. canephora* and *C. eugenioides* present significant differences from ancestral donors of *C. arabica* subgenomes, which may explain the lack of homology in certain NLR groups (Cenci et al., 2012).

The NB-ARC is the most conserved domain in the NLR gene family. Despite this conservation, from this domain it is possible to distinguish the TIR (TNL) and non-TIR classes based on different residues inside the motifs present in this region (Jones et al., 2016; Shao et al., 2019; Van Ghelder et al., 2019). Therefore, this domain is used to describe the phylogenetic relationships between the sequences of this group and classify them (Andolfo et al., 2014). The classification of NLRs in coffee revealed that the non-TNL class were in greater numbers than those of the TNL group in each of the three analyzed coffee genomes. It is known that non-TNL genes that include many CNL are widely distributed in monocots and dicots, while TNL are mainly found in dicots (McHale et al., 2006; Zheng et al., 2016). The

low frequency of TNLs in coffee agrees with results found for species of the solanum group, such as pepper, tomato and potato (Andolfo et al., 2014; Seo et al., 2016). Our results are also consistent with the low frequency of TNLs found in previous coffee studies (Hendre et al., 2011; Denoeud et al., 2014). Thus, it is possible to suggest that, as in solanum, non-TNLs represent an important repertoire of resistance genes in coffee. Additionally, the TNL group and the non-TNLs subgroups had NLRs from *C.arabica*, *C. canephora* and *C. eugenioides*, indicating conservation of the NLR classes across coffee genomes and suggesting that all subgroups were present in a common ancestor, thus as described for comparisons of species of the solanum group (Seo et al., 2016).

In the two phylogenetic trees analyzed, the clades group coffee NLRs that are mostly present in the same chromosomes but grouping NLRs in different chromosomes were also detected. Genes located on the same chromosome tend to group into subclades in the phylogenetic tree. However, rearrangement events of the chromosomes can affect NLR loci modify their genomic order or location (Cenci et al., 2012; Zheng et al., 2016; Ling et al., 2021).

Considering the relevance of coffee, few studies have been conducted addressing the identification of NLR in genomes of this crop. RGA studies using degenerate primers for NB-ARC region have already been performed (Noir et al., 2001; Hendre et al., 2011; Kumar, 2012), in addition to studies in S_H3 locus (Cenci et al., 2010, 2012; Lashermes et al., 2010), but very little is known about the NLR family in cultivated (*C.arabica* e *C. canephora*) and uncultivated coffee species (such as *C. eugenioides*). This is the first study focused on the wide identification of NLRs in *C. arabica* genome, besides to adding information to the existing report for *C. canephora* (Denoeud et al., 2014). The Genome-wide identification of coffe NLRs allow for more in-depth future molecular studies and represents a potential approach for candidates genes cloning and subsequent use of functional NLR genes, expanding the range of NLR in the breeding of this crop (Seo et al., 2016).

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

Anand, L., and Lopez, C. M. R. (2020). chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes.

Andolfo, G., D'agostino, N., Frusciante, L., and Ercolano, M. R. (2021). The tomato interspecific NB-LRR gene arsenal and its impact on breeding strategies. *Genes.* 12, 1–12. doi:10.3390/genes12020184.

Andolfo, G., Jupe, F., and Witek, K. (2014). Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. *BMC plant* 14, 120, 2014. doi:10.1186/1471-2229-14-120.

Bailey, P. C., Schudoma, C., Jackson, W., Baggs, E., Dagdas, G., Haerty, W., et al. (2018). Dominant integration locus drives continuous diversification of plant immune receptors

with exogenous domain fusions. *Genome Biol.* 19, 1–18. doi:10.1186/s13059-018-1392-6.

Barragan, A. C., and Weigel, D. (2021). Plant NLR Diversity: The Known Unknowns of Pan-NLRomes. *Plant Cell* 33, 814–831. doi:/10.1093/plcell/koaa002.

Bawin, Y., Ruttink, T., Staelens, A., Haegeman, A., Stoffelen, P., Mwanga Mwanga, J. C. I., et al. (2020). Phylogenomic analysis clarifies the evolutionary origin of *Coffea arabica*. *J. Syst. Evol.* doi:10.1111/jse.12694.

Bendahmane, A., Querci, M., Kanyuka, K., and Baulcombe, D. C. (2000). Agrobacterium transient expression system as a tool for the isolation of disease resistance genes: Application to the *Rx2* locus in potato. *Plant J.* 21, 73–81. doi:/10.1046/j.1365-313x.2000.00654.x.

Bent, A. F., Kunkel, B. N., Dahlbeck, D., Brown, K. L., Schmidt, R., Giraudat, J., et al. (1994). RPS2 of Arabidopsis thaliana: A leucine-rich repeat class of plant disease resistance genes. *Science* 265, 1856–1860. doi:10.1126/science.8091210.

Bezerra-Neto, J. P., Araújo, F. C., Ferreira-Neto, J. R. C., Silva, R. L. O., Borges, A. N. C., Matos, M. K. S., et al. (2020). "NBS-LRR genes-Plant health sentinels: Structure, roles, evolution and biotechnological applications," in *Applied Plant Biotechnology for Improving Resistance to Biotic Stress* (Elsevier), 63–120. doi:10.1016/B978-0-12-816030-5.00004-5.

Bonardi, V., Cherkis, K., Nishimura, M. T., and Dangl, J. L. (2012). A new eye on NLR proteins: Focused on clarity or diffused by complexity? *Curr. Opin. Immunol.* 24, 41–50. doi:10.1016/j.coi.2011.12.006.

Boutrot, F., and Zipfel, C. (2017). Function, discovery, and exploitation of plant pattern recognition receptors for broad-spectrum disease resistance. *Annu. Rev. Phytopathol.* 55, 257–286. doi:10.1146/annurev-phyto-080614-120106.

Cabral, P. G. C., Maciel-Zambolim, E., Oliveira, S. A. S., Caixeta, E. T., and Zambolim, L. (2016). Genetic diversity and structure of *Hemileia vastatrix* populations on *Coffea spp.* *Plant Pathol.* 65, 196–204. doi:10.1111/ppa.12411.

Cenci, A., Combes, M. C., and Lashermes, P. (2010). Comparative sequence analyses indicate that Coffea (Asterids) and Vitis (Rosids) derive from the same paleo-hexaploid ancestral genome. *Mol. Genet. Genomics* 283, 493–501. doi:10.1007/s00438-010-0534-7.

Cenci, A., Combes, M. C., and Lashermes, P. (2012). Genome evolution in diploid and tetraploid Coffea species as revealed by comparative analysis of orthologous genome segments. *Plant Mol. Biol.* 78, 135–145. doi:10.1007/s11103-011-9852-3.

Cerda, R., Avelino, J., Gary, C., Tixier, P., Lechevallier, E., and Allinne, C. (2017). Primary and secondary yield losses caused by pests and diseases: Assessment and modeling in coffee. *PLoS One* 12, 1–17. doi:10.1371/journal.pone.0169133.

Clarindo, W. R., and Carvalho, C. R. (2011). Flow cytometric analysis using SYBR Green I for genome size estimation in coffee. *Acta Histochem.* 113, 221–225. doi:10.1016/j.acthis.2009.10.005.

675 Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: An R package for the
676 visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940.
677 doi:10.1093/bioinformatics/btx364.

678 Davis, A. P., Govaerts, R., Bridson, D. M., and Stoffelen, P. (2006). An annotatated
679 taxonomic of the genus coffea (Rubiaceae). *Bot. J. Linn. Soc.* 152, 465–512.
680 doi:10.1111/j.1095-8339.2006.00584.x.

681 Denoeud, F., Carretero-Paulet, L., Dereeper, A., Droc, G., Guyot, R., Pietrella, M., et al.
682 (2014). The coffee genome provides insight into the convergent evolution of caffeine
683 biosynthesis. *Science* 345, 1181–1184. doi:10.1126/science.1255274.

684 Duarte, J. M., Wall, P. K., Edger, P. P., Landherr, L. L., Ma, H., Pires, J. C., et al. (2010).
685 Identification of shared single copy nuclear genes in Arabidopsis, Populus, Vitis and
686 Oryza and their phylogenetic utility across various taxonomic levels. *BMC Evol. Biol.*
687 10, 1–18. doi:10.1186/1471-2148-10-61.

688 Hamon, P., Hamon, S., Razafinarivo, N. J., Guyot, R., Siljak-Yakovlev, S., Couturon, E., et
689 al. (2015). "Coffea Genome Organization and Evolution," in *Coffee in Health and*
690 *Disease Prevention*, 29–37. doi:10.1016/B978-0-12-409517-5.00004-8.

691 Hendre, P. S., Bhat, P. R., Krishnakumar, V., Aggarwal, R. K., and Donini, P. (2011).
692 Isolation and characterization of resistance gene analogues from Psilanthus species that
693 represent wild relatives of cultivated coffee endemic to India. *Genome* 54, 377–390.
694 doi:10.1139/g11-004.

695 Hufford, M. B., Seetharam, A. S., Woodhouse, M. R., Chougule, K. M., Coletta, D., Tittes,
696 S., et al. (2021). De novo assembly, annotation, and comparative analysis of 26 diverse
697 maize genomes. *bioRxiv Prepr.* doi:10.1101/2021.01.14.426684.

698 Inturrisi, F., Bayer, P. E., Yang, H., Tirnaz, S., Edwards, D., and Batley, J. (2020). Genome-
699 wide identification and comparative analysis of resistance genes in *Brassica juncea*. *Mol.*
700 *Breed.* 40, 1–14. doi:10.1101/2020.12.15.422814.

701 Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329.
702 doi:10.1038/nature05286.

703 Jones, J. D. G., Vance, R. E., and Dangl, J. L. (2016). Intracellular innate immune
704 surveillance devices in plants and animals. *Science* 354. doi:10.1126/science.aaf6395.

705 Jost, M., Singh, D., Lagudah, E., Park, R. F., and Dracatos, P. (2020). Fine mapping of leaf
706 rust resistance gene Rph13 from wild barley. *Theor. Appl. Genet.* 133, 1887–1895.
707 doi:10.1007/s00122-020-03564-6.

708 Jupe, F., Pritchard, L., Etherington, G. J., MacKenzie, K., Cock, P. J. A., Wright, F., et al.
709 (2012). Identification and localisation of the NB-LRR gene family within the potato
710 genome. *BMC Genomics* 13, 1–14. doi:10.1186/1471-2164-13-75.

711 Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., et al. (2013).
712 Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR
713 gene family from sequenced plant genomes and rapid mapping of resistance loci in
714 segregating populations. *Plant J.* 76, 530–544. doi:10.1111/tpj.12307.

715 Kachroo, A., and Robin, G. P. (2013). Systemic signaling during plant defense. *Curr. Opin.*
716 *Plant Biol.* 16, 527–533. doi:10.1016/j.pbi.2013.06.019.

717 Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid
718 multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30,
719 3059–66. doi:/10.1093/nar/gkf436.

720 Kim, S., Park, J., Yeom, S. I., Kim, Y. M., Seo, E., Kim, K. T., et al. (2017). New reference
721 genome sequences of hot pepper reveal the massive evolution of plant disease-resistance
722 genes by retroduplication. *Genome Biol.* 18, 1–11. doi:10.1186/s13059-017-1341-9.

723 Kourelis, J., and Van Der Hoorn, R. A. (2018). Defended to the Nines: 25 years of Resistance
724 Gene Cloning Identifies Nine Mechanisms for R Protein Function. *Plant Cell* 30, 285–
725 299. doi:10.1105/tpc.17.00579.

726 Krishnan, S. (2017). Sustainable Coffee Production. *Oxford Res. Encycl. Environ. Sci.*
727 doi:10.1093/acrefore/9780199389414.013.224.

728 Kroj, T., Chanclud, E., Michel-Romiti, C., Grand, X., and Morel, J. B. (2016). Integration of
729 decoy domains derived from protein targets of pathogen effectors into plant immune
730 receptors is widespread. *New Phytol.* 210, 618–626. doi:10.1111/nph.13869.

731 Kumar, D. (2012). Isolation of Nucleotide Binding Site (NBS)-Leucine Rich Repeat (LRR)
732 Resistant Gene Analogs (Rgas) In Arabica Coffee (*Coffea arabica* L. Cv S.288). *J.*
733 *Biotechnol. Biomater.* 02. doi:10.4172/2155-952x.1000146.

734 Kushwaha, S. K., Chauhan, P., Hedlund, K., and Ahren, D. (2016). NBSPred: A support
735 vector machine-based high-throughput pipeline for plant resistance protein NBSLRR
736 prediction. *Bioinformatics* 32, 1223–1225. doi:10.1093/bioinformatics/btv714.

737 Lashermes, P., Combes, M. C., Ribas, A., Cenci, A., Mahé, L., and Etienne, H. (2010).
738 Genetic and physical mapping of the SH3 region that confers resistance to leaf rust in
739 coffee tree (*Coffea arabica* L.). *Tree Genet. Genomes* 6, 973–980. doi:10.1007/s11295-
740 010-0306-x.

741 Lashermes, P., Combes, M. C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F., et al. (1999).
742 Molecular characterization and origin of the Coffea arabica L. genome. *Mol. Gen. Genet.*
743 261, 259–266. doi:10.1007/s004380050965.

744 Lefebvre-Pautigny, F., Wu, F., Philippot, M., Rigoreau, M., Priyono, Zouine, M., et al.
745 (2010). High resolution synteny maps allowing direct comparisons between the coffee
746 and tomato genomes. *Tree Genet. Genomes* 6, 565–577. doi:10.1007/s11295-010-0272-
747 3.

748 Letunic, I., and Bork, P. (2021). Interactive tree of life (iTOL) v3: an online tool for the
749 display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 49, W293–
750 W296. doi:10.1093/nar/gkw290.

751 Li, L., Stoeckert, C. J. J., and Roos, D. S. (2003). OrthoMCL: Identification of Ortholog
752 Groups for Eukaryotic Genomes. *Genome Res.* 13, 2178–2189.
753 doi:10.1101/gr.1224503.candidates.

754 Lin, C., Mueller, L. A., Carthy, J. M., Crouzillat, D., Pétiard, V., and Tanksley, S. D. (2005).
755 Coffee and tomato share common gene repertoires as revealed by deep sequencing of
756 seed and cherry transcripts. *Theor. Appl. Genet.* 112, 114–130. doi:10.1007/s00122-005-
757 0112-2.

758 Ling, J., Xie, X., Gu, X., Zhao, J., Ping, X., Li, Y., et al. (2021). High-quality chromosome-
759 level genomes of *Cucumis metuliferus* and *Cucumis melo* provide insight into Cucumis
760 genome evolution. *Plant J.* 107, 136–148. doi:10.1111/tpj.15279.

761 Lozano, R., Hamblin, M. T., Prochnik, S., and Jannink, J. L. (2015). Identification and
762 distribution of the NBS-LRR gene family in the Cassava genome. *BMC Genomics* 16, 1–
763 14. doi:10.1186/s12864-015-1554-9.

764 Lu, Y., and Tsuda, K. (2021). Intimate Association of PRR- and NLR-Mediated Signaling in
765 Plant Immunity. *Mol. Plant-Microbe Interact.* 34, 3–14. doi:10.1094/mpmi-08-20-0239-
766 ia.

767 Malacarne, G., Perazzolli, M., Cestaro, A., Sterck, L., Fontana, P., van de Peer, Y., et al.
768 (2012). Deconstruction of the (paleo)polyploid grapevine genome based on the analysis
769 of transposition events involving NBS resistance genes. *PLoS One* 7.
770 doi:10.1371/journal.pone.0029762.

771 McCook, S., and Vandermeer, J. (2015). The Big Rust and the Red Queen: Long-Term
772 Perspectives on Coffee Rust Research. *Phytopathology* 105, 1164–1173.
773 doi:10.1094/PHYTO-04-15-0085-RVW.

774 McHale, L., Tan, X., Koehl, P., and Michelmore, R. W. (2006). Plant NBS-LRR proteins:
775 Adaptable guards. *Genome Biol.* 7. doi:10.1186/gb-2006-7-4-212.

776 Michelmore, R. W., and Meyers, B. C. (1998). Clusters of resistance genes in plants evolve
777 by divergent selection and a birth-and-death process. *Genome Res.* 8, 1113–1130.
778 doi:10.1101/gr.8.11.1113.

779 Mindrinos, M., Katagiri, F., Yu, G. L., and Ausubel, F. M. (1994). The A. thaliana disease
780 resistance gene RPS2 encodes a protein containing a nucleotide-binding site and leucine-
781 rich repeats. *Cell* 78, 1089–1099. doi:10.1016/0092-8674(94)90282-8.

782 Monteiro, F., and Nishimura, M. T. (2018). Structural, functional, and genomic diversity of
783 plant NLR proteins: An evolved resource for rational engineering of plant immunity.
784 *Annu. Rev. Phytopathol.* 56, 243–267. doi:10.1146/annurev-phyto-080417-045817.

785 Muliyar, R. K., Chowdappa, P., Behera, S. K., Kasaragod, S., Gangaraj, K. P., Kotimoole, C.
786 N., et al. (2020). Assembly and Annotation of the Nuclear and Organellar Genomes of a
787 Dwarf Coconut (Chowghat Green Dwarf) Possessing Enhanced Disease Resistance.
788 *Omi. A J. Integr. Biol.* 24, 726–742. doi:10.1089/omi.2020.0147.

789 Noir, S., Combes, M.-C., Anthony, F., and Lashermes, P. (2001). Origin, diversity and
790 evolution of NBS-type disease-resistance gene homologues in coffee trees (Coffea L.).
791 *Mol. Genet. Genomics* 265, 654–662. doi:10.1007/s004380100459.

792 Noirot, M., Poncet, V., Barre, P., Hamon, P., Hamon, S., and De Kochko, A. (2003). Genome
793 size variations in diploid African Coffea species. *Ann. Bot.* 92, 709–714.

794 doi:10.1093/aob/mcg183.

795 Osuna-Cruz, C. M., Paytuvi-Gallart, A., Di Donato, A., Sundesha, V., Andolfo, G., Cigliano,
796 R. A., et al. (2018). PRGdb 3.0: A comprehensive platform for prediction and analysis of
797 plant disease resistance genes. *Nucleic Acids Res.* 46, D1197–D1201.
798 doi:10.1093/nar/gkx1119.

799 Paal, J., Henselewski, H., Muth, J., Meksem, K., Menéndez, C. M., Salamini, F., et al. (2004).
800 Molecular cloning of the potato Gro1-4 gene conferring resistance to pathotype Ro1 of
801 the root cyst nematode Globodera rostochiensis, based on a candidate gene approach.
802 *Plant J.* 38, 285–297. doi:10.1111/j.1365-313X.2004.02047.x.

803 Read, A. C., Moscou, M. J., Zimin, A. V., Pertea, G., Meyer, R. S., Purugganan, M. D., et al.
804 (2020). Genome assembly and characterization of a complex zfBED-NLR gene-
805 containing disease resistance locus in Carolina Gold Select rice with Nanopore
806 sequencing. *PLoS Genet.* 16, 1–24. doi:10.1371/journal.pgen.1008571.

807 Ribas, A. F., Cenci, A., Combes, M. C., Etienne, H., and Lashermes, P. (2011). Organization
808 and molecular evolution of a disease-resistance gene cluster in coffee trees. *BMC*
809 *Genomics* 12. doi:10.1186/1471-2164-12-240.

810 Schatz, M. C., Maron, L. G., Stein, J. C., Hernandez Wences, A., Gurtowski, J., Biggers, E.,
811 et al. (2014). Whole genome de novo assemblies of three divergent strains of rice, Oryza
812 sativa, document novel gene space of aus and indica. *Genome Biol.* 15, 506.
813 doi:10.1186/s13059-014-0506-z.

814 Scott, A. D., Zimin, A. V, Puiu, D., Workman, R., Britton, M., Zaman, S., et al. (2020). The
815 giant sequoia genome and proliferation of disease resistance genes. *bioRxiv* 28, 1–43.
816 doi:10.1101/2020.03.17.995944.

817 Seidl, M. F., and Thomma, B. P. H. J. (2017). Transposable Elements Direct The Coevolution
818 between Plants and Microbes. *Trends Genet.* 33, 842–851.
819 doi:10.1016/j.tig.2017.07.003.

820 Seo, E., Kim, S., Yeom, S. I., and Choi, D. (2016). Genome-wide comparative analyses reveal
821 the dynamic evolution of nucleotide-binding leucine-rich repeat gene family among
822 solanaceae plants. *Front. Plant Sci.* 7, 1–13. doi:10.3389/fpls.2016.01205.

823 Shao, Z. Q., Xue, J. Y., Wang, Q., Wang, B., and Chen, J. Q. (2019). Revisiting the Origin of
824 Plant NBS-LRR Genes. *Trends Plant Sci.* 24, 9–12. doi:10.1016/j.tplants.2018.10.015.

825 Son, S., Kim, S., Lee, K., Oh, J., Choi, I., Do, J., et al. (2021). The *Capsicum baccatum*-
826 Specific Truncated NLR Protein CbCN Enhances the Innate Immunity against
827 Colletotrichum acutatum. *Int. J. Mol. Sci.* 22. doi:10.3390/ijms22147672.

828 Song, W., Wang, B., Li, X., Wei, J., Chen, L., Zhang, D., et al. (2015). Identification of
829 Immune Related LRR-Containing Genes in Maize (*Zea mays L.*) by Genome-Wide
830 Sequence Analysis. *Int. J. Genomics* 2015. doi:10.1155/2015/231358.

831 Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis
832 of large phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033.

833 Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006).
834 AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* 34,
835 W435–W439. doi:10.1093/nar/gkl200.

836 Steuernagel, B., Jupe, F., Witek, K., Jones, J. D. G., and Wulff, B. B. H. (2015). NLR-parser:
837 Rapid annotation of plant NLR complements. *Bioinformatics* 31, 1665–1667.
838 doi:10.1093/bioinformatics/btv005.

839 Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H. J.,
840 Yu, G., et al. (2020). The NLR-annotator tool enables annotation of the intracellular
841 immune receptor repertoire. *Plant Physiol.* 183, 468–482. doi:10.1104/pp.19.01273.

842 Sun, X., Jiao, C., Schwaninger, H., Chao, C. T., Ma, Y., Duan, N., et al. (2020). Phased
843 diploid genome assemblies and pan-genomes provide insights into the genetic history of
844 apple domestication. *Nat. Genet.* 52, 1423–1432. doi:10.1038/s41588-020-00723-9.

845 Toda, N., Rustenholz, C., Baud, A., Le Paslier, M. C., Amselem, J., Merdinoglu, D., et al.
846 (2020). NLGenomeSweeper: A tool for genome-wide NBS-LRR resistance gene
847 identification. *Genes.* 11, 333. doi:10.3390/genes11030333.

848 Tran, H. T. M., Ramaraj, T., Furtado, A., Lee, L. S., and Henry, R. J. (2018). Use of a draft
849 genome of coffee (*Coffea arabica*) to identify SNPs associated with caffeine content.
850 *Plant Biotechnol. J.* 9, 1756–1766. doi:10.1111/pbi.12912.

851 Van Der Vossen, E., Sikkema, A., Te Lintel Hekkert, B., Gros, J., Stevens, P., Muskens, M.,
852 et al. (2003). An ancient R gene from the wild potato species *Solanum bulbocastanum*
853 confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and
854 tomato. *Plant J.* 36, 867–882. doi:10.1046/j.1365-313X.2003.01934.x.

855 Van Ghelder, C., Parent, G. J., Rigault, P., Prunier, J., Giguère, I., Caron, S., et al. (2019). The
856 large repertoire of conifer NLR resistance genes includes drought responsive and highly
857 diversified RNLs. *Sci. Rep.* 9, 1–13. doi:10.1038/s41598-019-47950-7.

858 Wan, H., Yuan, W., Bo, K., Shen, J., Pang, X., and Chen, J. (2013). Genome-wide analysis of
859 NBS-encoding disease resistance genes in *Cucumis sativus* and phylogenetic study of
860 NBS-encoding genes in Cucurbitaceae crops. *BMC Genomics* 14, 109.
861 doi:10.1186/1471-2164-14-109.

862 Wang, Z., Ren, H., Xu, F., Lu, G., Cheng, W., Que, Y., et al. (2021). Genome-Wide
863 Characterization of NLRs in *Saccharum spontaneum* L. and Their Responses to Leaf
864 Blight in Saccharum. *Agronomy* 11, 153. doi:10.3390/plants10020322.

865 Yuan, M., Jiang, Z., Bi, G., Nomura, K., Liu, M., He, S. Y., et al. (2021). Pattern-recognition
866 receptors are required for NLR-mediated plant immunity. *Nature*, 1–5.
867 doi:10.1101/2020.04.10.031294.

868 Zambolim, L., and Caixeta, E. T. (2021). An overview of physiological specialization
869 ofcoffee leaf rust – new designation of Pathotypes. *Int. J. Curr. Res.* 13, 15564–15575.

870 Zhang, R., Murat, F., Pont, C., Langin, T., and Salse, J. (2014). Paleo-evolutionary plasticity
871 of plant disease resistance genes. *BMC Genomics* 15. doi:10.1186/1471-2164-15-187.

Zhao, T., Rui, L., Li, J., Nishimura, M. T., Vogel, J. P., Liu, N., et al. (2015). A Truncated NLR Protein, TIR-NBS2, Is Required for Activated Defense Responses in the exo70B1 Mutant. *PLoS Genet.* 11, 1–28. doi:10.1371/journal.pgen.1004945.

Zheng, F., Wu, H., Zhang, R., Li, S., He, W., Wong, F. L., et al. (2016). Molecular phylogeny and dynamic evolution of disease resistance genes in the legume family. *BMC Genomics* 17, 1–13. doi:10.1186/s12864-016-2736-9.

Zimmer, A., Lang, D., Richardt, S., Frank, W., Reski, R., and Rensing, S. A. (2007). Dating the early evolution of plants: Detection and molecular clock analyses of orthologs. *Mol. Genet. Genomics* 278, 393–402. doi:10.1007/s00438-007-0257-6.

Figure 1. Venn diagrams representing the overlap between the loci from NLR-annotator and NLR genes annotated in the *C. arabica, C. canephora,* and *C. eugenioides* refence genomes. The colors represent the origin of the annotation, with blue indicating those annotated by NLR-annotator and green indicating those found in the reference genome. The intersection refers to the overlap that occurred once or more than once.

Figure 2. Number and chromosomal distribution of NLR loci in *C. arabica, C. canephora* e *C. eugenioides*. **A**. The chromosomes with the highest number of NLR loci are highlighted in dark blue, and the smallest number of NRL are highlighted in light blue. CPL identifies the completeness of NLR as: C= complete, Cps= complete (pseudogene), P= partial, Pps= partial pseudogene and Un= unassigned. **B.** The chromosomes represented in *C. arabica* refer to the two subgenomes, the first originating from *C. canephopr*a and the second originating from *C. eugenioides*. The chromosomal distribution represented in this figure does not show all loci for each region but identifies all regions that present NLRs loci. A browse to view these chromosomes and observe all regions in detail may be found at Supplementary Figure 3, 4 and 5.

Figure 3. **Upset plot of orthologous NLR groups (orthogroups) among five species (*C. arabica* (Ca) *C. canephora* (Cc), *C. eugenioides* (Ce), *S. tuberosum* (Soltu) and *S. lycopersicum* (Soly) and NLRs de referência (Ref).** The orthogroups that gather combination for species/Ref NLRs is shown by the interconnected blue dots on the bottom panel. Unconnected blue dots show orthogroups within the same species. The ‘Set size’ represents the total number of orthogroups per species/Ref. The ‘intersection size’ shows the number of orthogroups inside and between species/Ref. The orthogroups were clustered by OrthoMCL.

Figure 4. **Phylogenetic tree of single-copy NLR orthogroups.** The phylogenetic tree was constructed based on 647 NLRs (domain NB-ARC) single-copy orthologs using RAxML. The colored ring indicates coffee NLR clades, the green color represents *C. arabica* (Ca), red, *C. canephora* (Cc) and blue *C. eugenioides* (Ce). Labels in black are coffee NLRs, in green, *S. lycopersicum* (Soly), in blue, *S. tuberosum* (Soltu) and pink are reference proteins (Ref). Bootstrap values above 70% are indicated on each branch with a brown circle. Pink backgrounds indicate clades that group orthologs of Ca, Cc and Ce. The clade highlighted in purple shows the coffee NLR and RPS2 grouping.

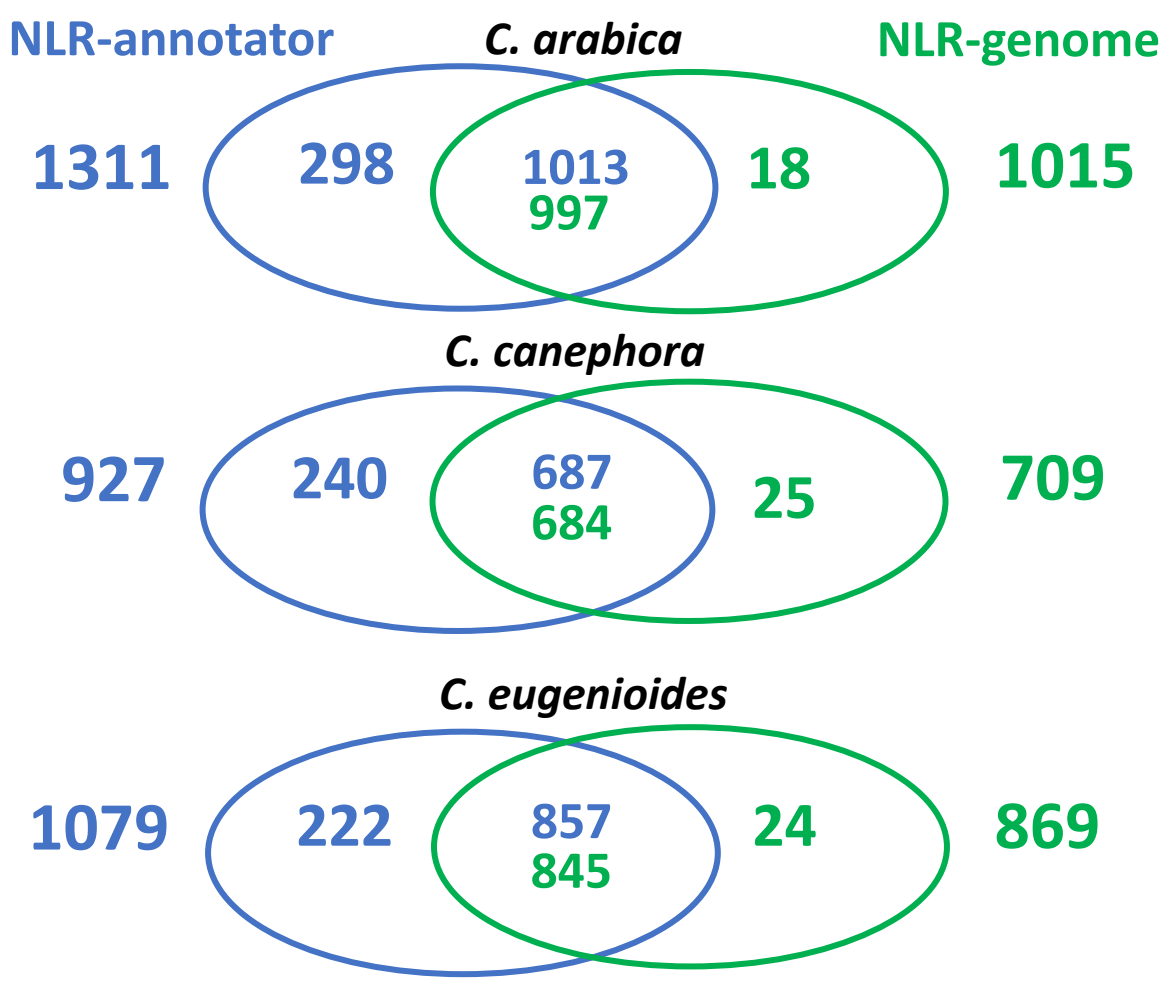
Figure 5. **Phylogenetic tree for coffee NLRs classification**. NB-ARC domains from 2681 NLRs gathering *C. arabica* (*Ca*), *C. canephora (Cc), C. eugenioides* (Ce), *S. lycopersicum, S. tuberosum*, reference NLRs (pink balls) and CED-4 (root) were used in the construction. The tree was constructed using RAxML. The classifications of the reference NLRs and some *S. lycopersicum, S. tuberosum* NLRs were used to classify the coffee NLRs into TNLs and Non TNLs groups (inner ring - TNL = yellow, CNL = gray and NL = Purple). Subgroups in Non TNLs are indicated from I to XIII and alternating colors (green and purple). Gray and yellow background highlight coffee NLRs and the outer ring separates the NLRs for Ca, Cc and Ce in green, red and blue respectively.

Table and Figures

Table 1. The loci from NLR-annotator that did not overlap with annotations of NLR genes from coffee reference genomes.

Species	Complete	Complete (pseudogene)	Partial	Partial (pseudogene)	Total
C. arabica	70	90	65	73	298
C. canephora	67	56	73	44	240
C. eugenioides	71	65	37	49	222

Figure 1



945 **Figure 2**

A

Coffea arabica

CPL	Chr1c	Chr1e	Chr2c	Chr2e	Chr3c	Chr3e	Chr4c	Chr4e	Chr5c	Chr5e	Chr6c	Chr6e	Chr7c	Chr7e	Chr8c	Chr8e	Chr9c	Chr9e	Chr10c	Chr10e	Chr11c	Chr11e	Un	Total
C	42	48	17	22	71	65	25	21	42	49	23	15	44	43	35	54	11	7	16	9	54	73	23	809
Cps	21	13	7	8	38	24	5	6	28	19	3	6	14	15	12	23	1	4	1	1	12	23	5	289
P	5	4	4	6	13	12	1	4	4	5	10	2	5	6	4	3	1	0	5	8	10	7	6	125
Pps	5	2	2	3	8	4	2	1	1	2	2	4	2	4	4	9	2	0	4	3	10	13	0	88
Total	**73**	**67**	**30**	**39**	**130**	**105**	**33**	**32**	**75**	**75**	**38**	**27**	**65**	**68**	**55**	**89**	**15**	**11**	**26**	**21**	**86**	**116**	**34**	**1311**

Coffea canephora

CPL	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Un	Total
C	49	19	70	23	42	16	29	37	1	11	55	210	562
Cps	14	9	25	7	14	4	7	12	1	1	14	66	174
P	3	7	12	3	5	2	5	2	3	3	8	68	121
Pps	6	2	13	1	2	2	2	3	1	1	13	23	70
Total	**72**	**37**	**120**	**34**	**63**	**24**	**43**	**54**	**6**	**16**	**90**	**367**	**927**

Coffea eugenioides

CPL	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9	Chr10	Chr11	Un	Total
C	56	30	98	17	86	40	56	106	18	10	90	88	695
Cps	14	17	33	7	34	16	18	27	5	3	29	36	239
P	3	10	13	2	9	4	8	9	1	3	16	5	83
Pps	2	3	11	1	5	3	3	6	3	4	14	6	62
Total	**75**	**60**	**155**	**27**	**134**	**63**	**85**	**148**	**27**	**20**	**149**	**135**	**1079**

B

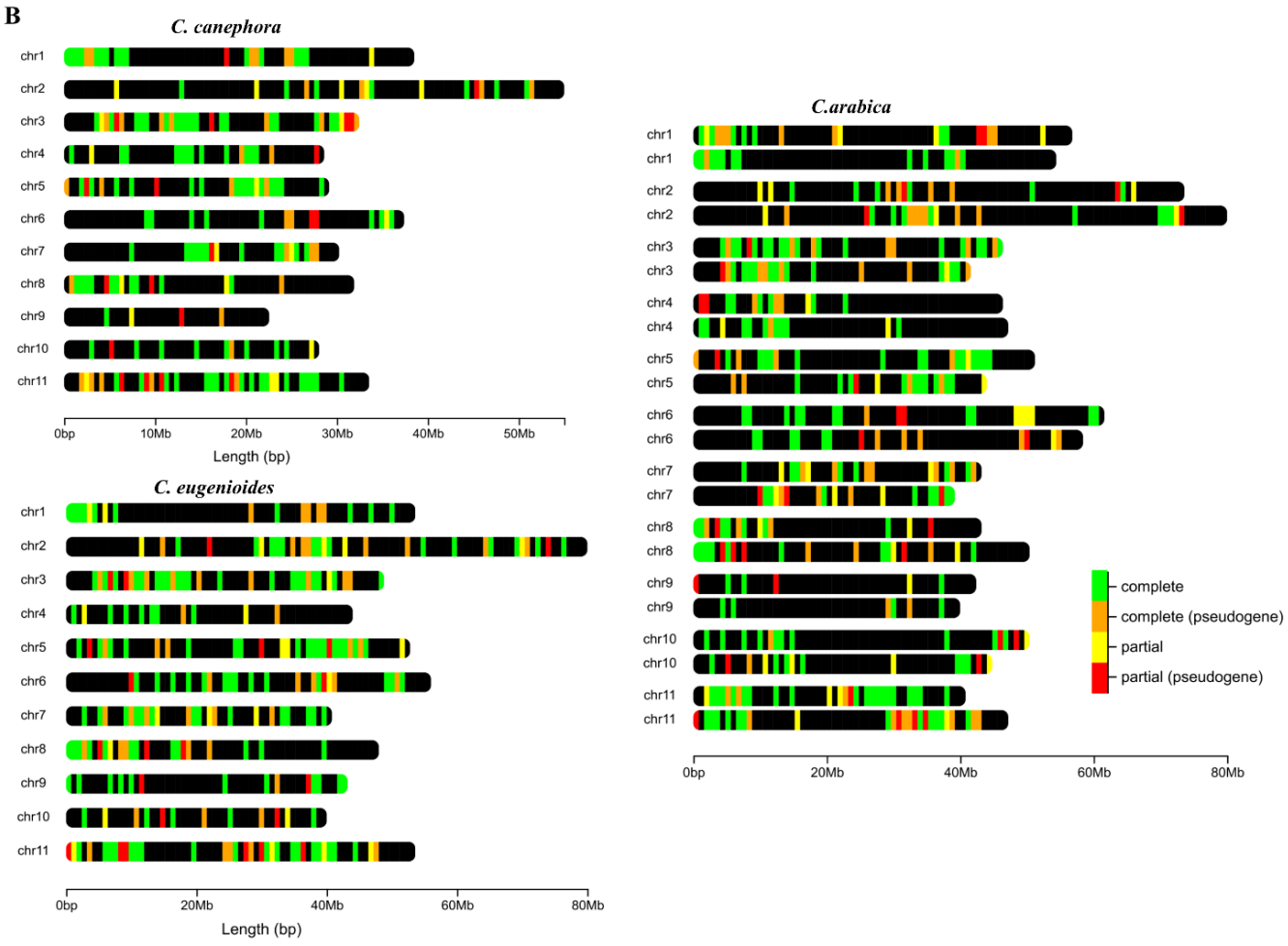


Figure 3

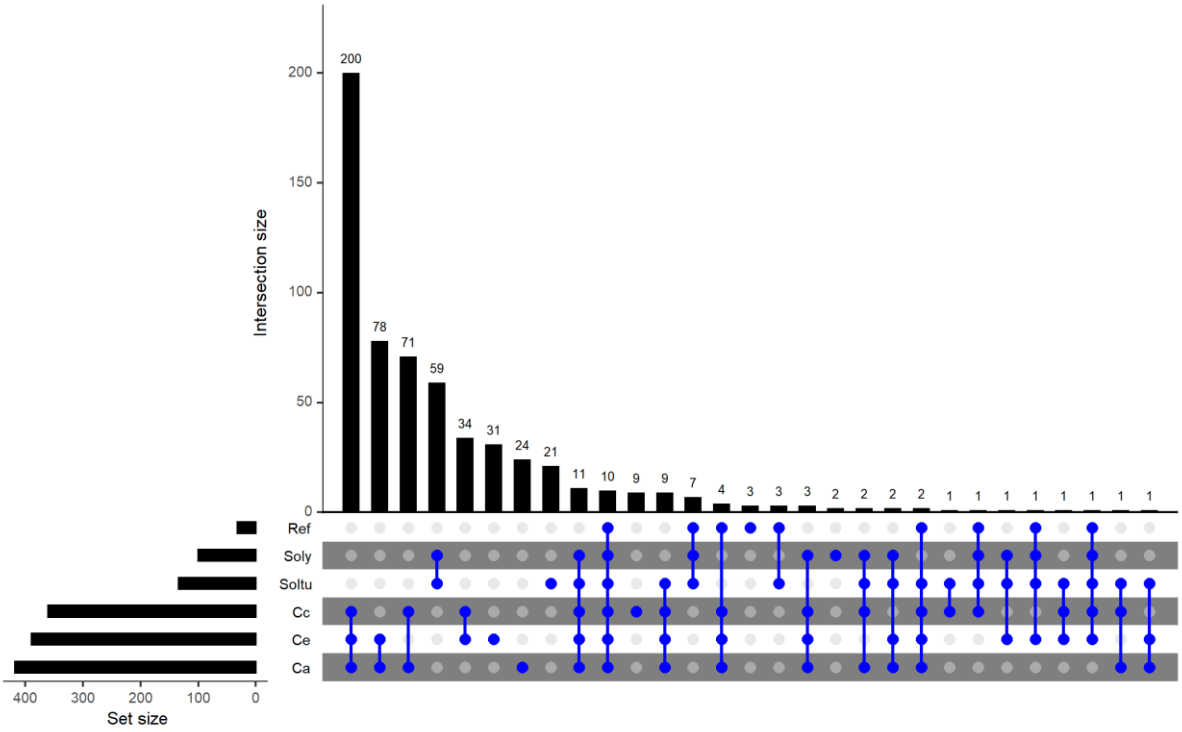

Intersection size
200
150
100
50
0
200
78
71
59
34
31
24
21
11
10
9
9
7
4
3
3
3
2
2
2
2
1
1
1
1
1
1
1
1
Ref
Soly
Soltu
Cc
Ce
Ca
400
300
200
100
0
Set size

Figure 4

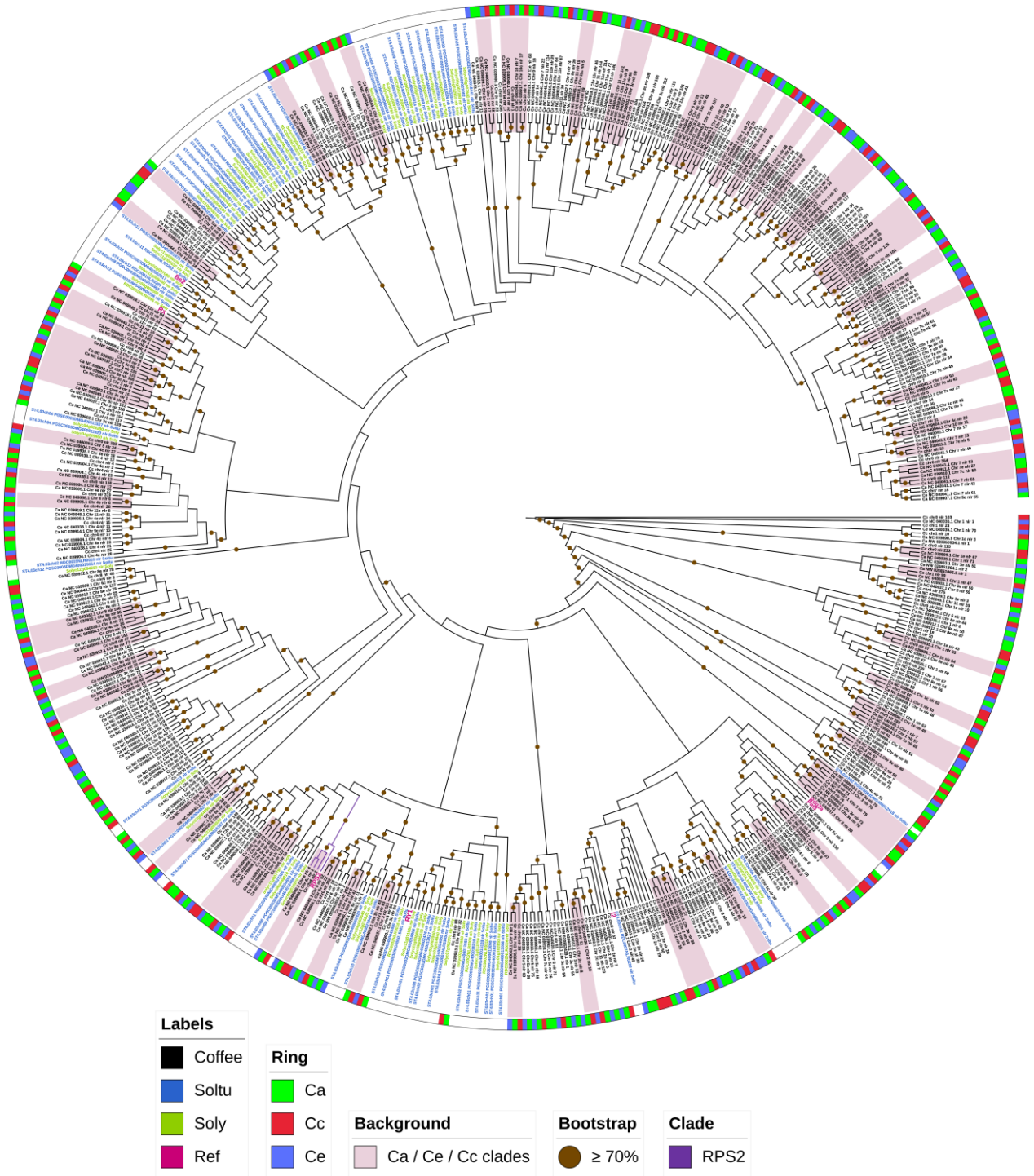
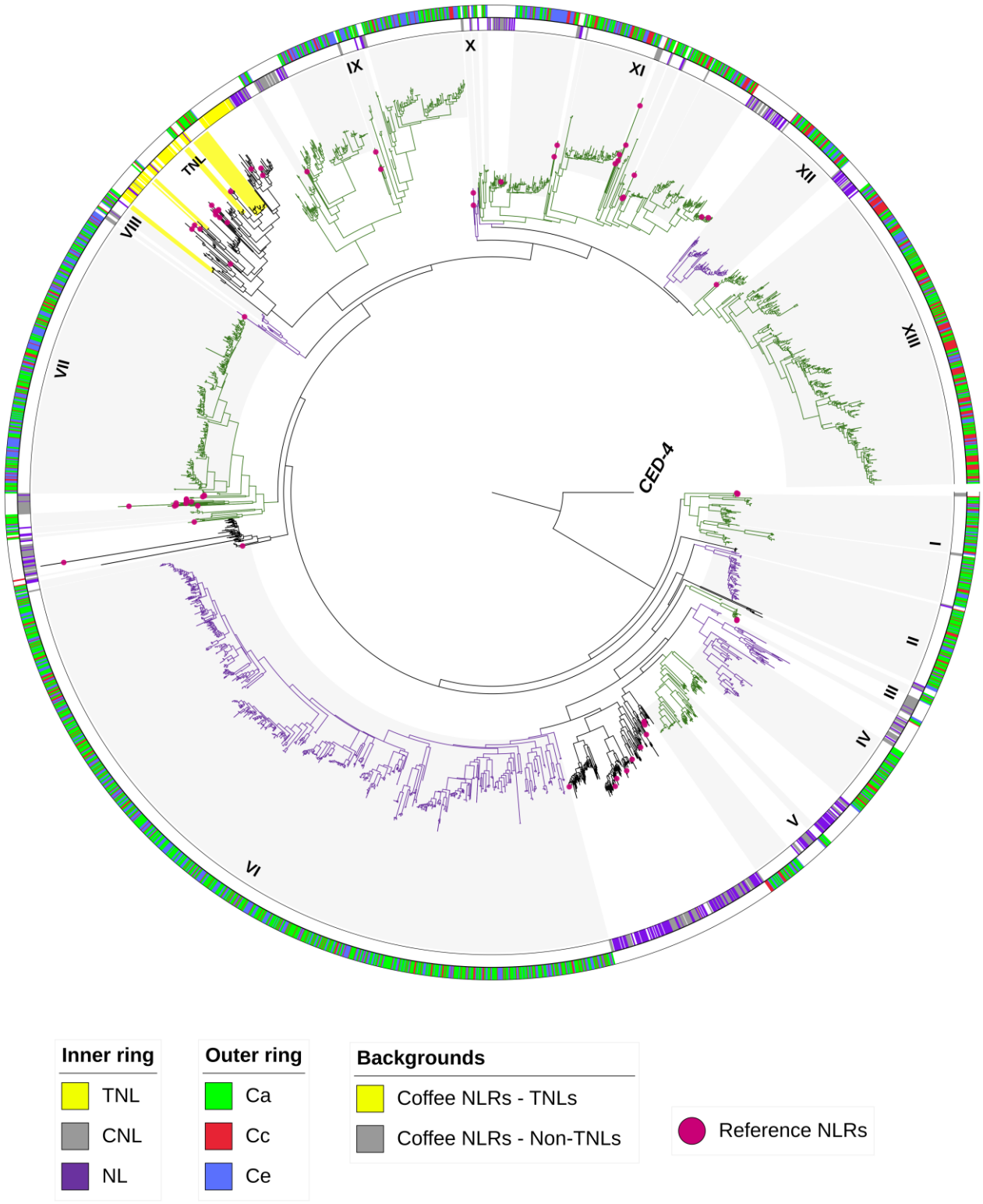

Labels
Coffee
Soltu
Soly
Ref
Ring
Ca
Cc
Ce
Background
Ca / Ce / Cc clades
Bootstrap
≥ 70%
Clade
RPS2

Figure 5


X
IX
XI
XII
XIII
TNL
VIII
VII
CED-4
I
II
III
IV
V
VI
Inner ring
TNL
CNL
NL
Outer ring
Ca
Cc
Ce
Backgrounds
Coffee NLRs - TNLs
Coffee NLRs - Non-TNLs
Reference NLRs

Supplementary Material

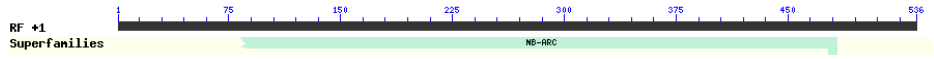
Supplementary Text 1. NLRs genes not found by the NLR-annotator in the coffee genomes.

The overlap analysis made it possible to identify genes annotated in the reference genomes that did not overlap with any locus from NLR-annotator. To examine these genes, an NLR-parser analysis was performed on this set. After this analysis, we observed genes that did not obtain motifs detectable by NLR-parser and therefore did not present annotations in NLR-annotator. These loci were considered as "not found" by the tool. For *C. arabica*, two genes were not found, the LOC113737176 (XP_027120243.1 disease resistance protein RGA2-like) and the LOC113735982 (XP_027118739.1 probable disease resistance protein At4g19060) and for *C. canephora*, five genes were not found: Cc02_g12220 (Putative disease resistance protein At4g19050), Cc03_g10360 (Hypothetical protein with PFAM:PF00931), Cc07_g18800 (Hypothetical protein with PFAM:PF00931), Cc00_g21910 (Putative NBS-coding resistance gene protein -Fragment), and Cc00_g35420 (Putative Disease resistance protein RGA2).

For *C. eugenioides*, two situations occurred: the LOC113766771 (XP_027166726.1- probable disease resistance protein At5g43730) and the LOC113766774 (XP_027166730.1 - disease resistance protein At4g27190-like) did not present motifs detectable by NLR-Parser and consequently, were not annotated by NLR-annotator. As well, two genes (LOC113766615 - putative disease resistance protein RGA4 and LOC113777141 - putative late blight resistance protein homolog R1B-17) had at least three consecutive motifs belonging to the NB-ARC domain and detectable by NLR-Paser (standard threshold) but were not annotated. For the latter situation, a manual search of the positions in the txt file (Supplementary Table 3) was undertaken to make sure that there were no errors in the overlap analysis. The locus in this interval (Chr3 start at 40629207 and ends at 40630439, Ch7 start at 40630439 and ends at 22112852) was not detected by the NLR-annotator, and these genes were also classified as "not found" (Supplementary Table 6).

To ensure that the genes that NLR-annotator did not found encode proteins with NB-ARC domains, in addition to the PfamScan analysis, a conserved domains analysis using the nucleotide sequence was performed (Supplementary Figure 1). The goal here was to use a strategy similar to that used by the NLR-annotator. For *C. arabica*, we observed that the two genes belong in the NB-ARC superfamily, but with incomplete domain and variation in domain length. This also occurred in *C. canephora*, except for one gene (Cc07_g18800) that did not present detectable domains, despite showing a small fragment of the NB-ARC with PfamScan analysis (highlighted in orange in Supplementary Table 4). Additionally, in this genome, the Cc00_g35420 gene presented an NB-ARC incomplete domain and contained an N-terminal with Rx_N (pfam18052), a domain found in many plant resistance proteins. For *C. eugenioides*, two genes (LOC113766615 and LOC113766771) presented NB-ARC incomplete domains of different lengths, and the LOC113766774 gene, in addition to presenting NB-ARC complete domain, contains leucine-rich repeat. The gene LOC113777141 has two NB-ARC fragments separated by a domain belonging to the retrotran gag two superfamilies, containing a gag-polypeptide of LTR (long terminal repeat), which is a known type of retrotransposon in plants.

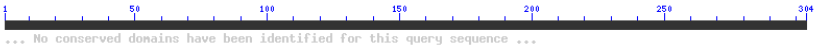
ID	Name	Accession	Description	Interval	E-value
LOC113737176_Ca	NB-ARC super family	cl26397	NB-ARC domain	208-438	8.92e-11



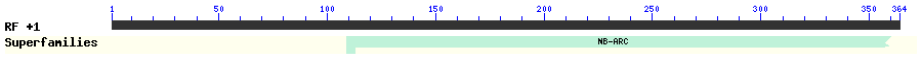
ID	Name	Accession	Description	Interval	E-value
LOC113735982_Ca	NB-ARC super family	cl26397	NB-ARC domain	82-483	1.90e-11

ID	Name	Accession	Description	Interval	E-value
Cc02_g12220_Cc	NB-ARC super family	cl26397	NB-ARC domain	1193-1771	3.08e-08

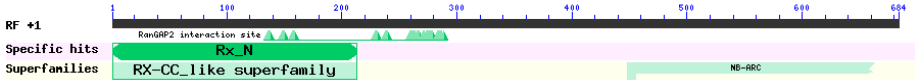
ID	Name	Accession	Description	Interval	E-value
Cc03_g10360_Cc	NB-ARC super family	cl26397	NB-ARC domain	1561-1836	7.70e-13



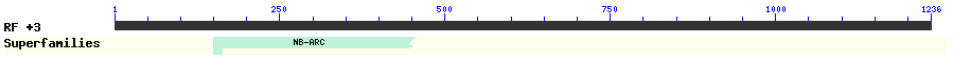
ID	Name	Accession	Description	Interval	E-value
Cc07_g18800_Cc	-	-	-	-	-



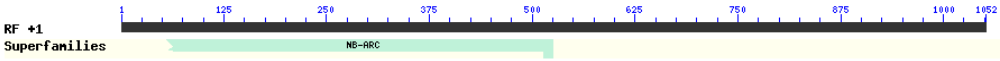
ID	Name	Accession	Description	Interval	E-value
Cc00_g21910_Cc	NB-ARC super family	cl26397	NB-ARC domain	109-360	7.62e-08



ID	Name	Accession	Description	Interval	E-value
Cc00_g35420_Cc	Rx_N	pfam18052	Rx N-terminal domain; This entry represents the N-terminal domain found in many plants	1-213	1.29e-17
	NB-ARC super family	cl26397	NB-ARC domain	448-663	1.78e-10

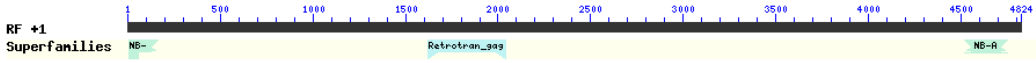


ID	Name	Accession	Description	Interval	E-value
LOC113766615_Ce	NB-ARC super family	cl26397	NB-ARC domain	150-455	1.53e-11



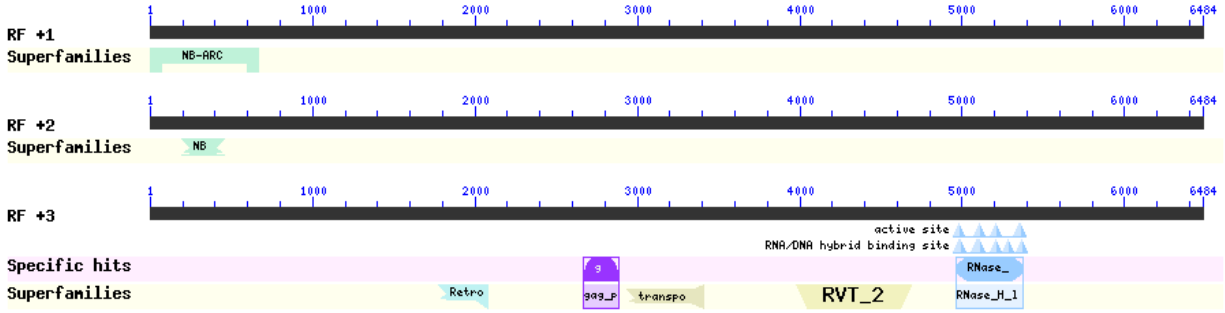
ID	Name	Accession	Description	Interval	E-value
LOC113766771_Ce	NB-ARC super family	cl26397	NB-ARC domain	55-525	9.82e-07

ID	Name	Accession	Description	Interval	E-value
LOC113766774_Ce	NB-ARC super family	cl26397	NB-ARC domain	637-1326	9.69e-11

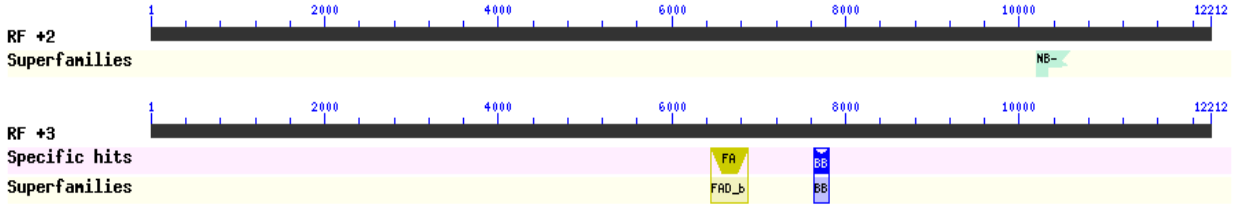


ID	Name	Accession	Description	Interval	E-value
LOC113777141_Ce	Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	1624-2040	2.22e-14
	NB-ARC super family	cl26397	NB-ARC domain	7-165	1.58e-10
	NB-ARC super family	cl26397	NB-ARC domain	4519-4749	7.73e-10

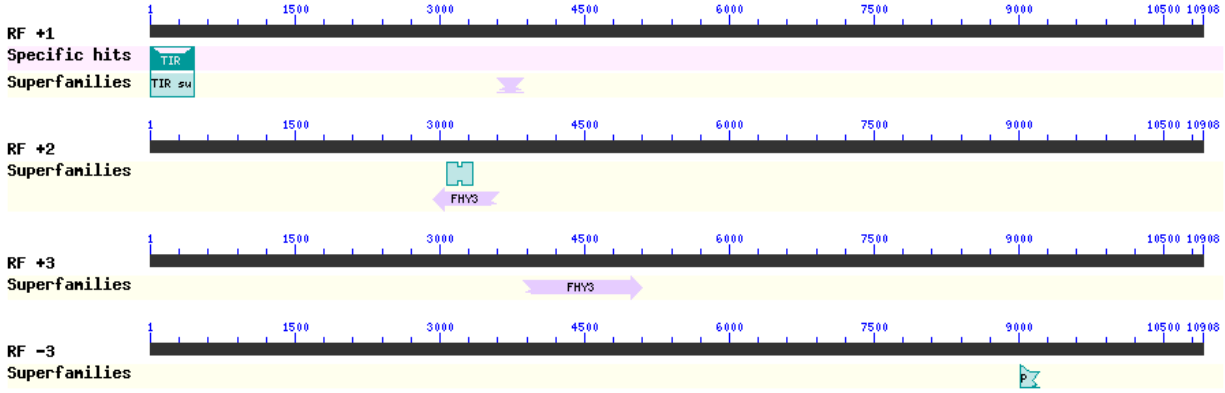
Supplementary Figure 1. **Conserved domains analysis for genes not found by the NLR-annotator in the coffee genomes**. For the analysis, nucleotide sequences were used. The table shows the list of domain hits found for each of the sequences. The conserved domains were detected using NCBI Conserved Domain Database and the graphical summary was set to concise view. If the alignment omitted more than 20% of the either the N- or C-terminus or both, the partial nature of the hit is indicated in the graphical display as domain with jagged edges. At the end of each ID, the source genome is identified, Ca: *C. arabica*, Cc: *C. eugenioides* and Ce: *C. eugenioides*. RF: reading frame.



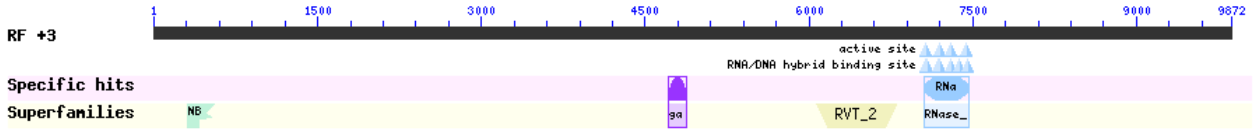
Chr_4c_nlr_22_Ca partial				
Name	**Accession**	**Description**	**Interval**	**E-value**
NB-ARC super family	cl26397	NB-ARC domain	1-663	1.67e-06
NB-ARC super family	cl26397	NB-ARC domain	194-457	4.21e-12
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements	4965-5378	2.30e-60
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	3978-4682	8.87e-51
Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	1779-2075	5.08e-16
Transpos_IS481 super family	cl41329	IS481 family transposase	2934-3404	3.45e-12
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	2670-2885	3.14e-11



Chr_3c_nlr_44_Ca complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
NB-ARC super family	cl26397	NB-ARC domain	10199-10585	2.48e-10
FAD_binding_4	pfam01565	FAD binding domain; This family consists of various enzymes that use FAD as a co-factor	6453-6872	5.37e-19
BBE	pfam08031	Berberine and berberine like; This domain is found in the berberine bridge and berberine	7632-7805	5.77e-14

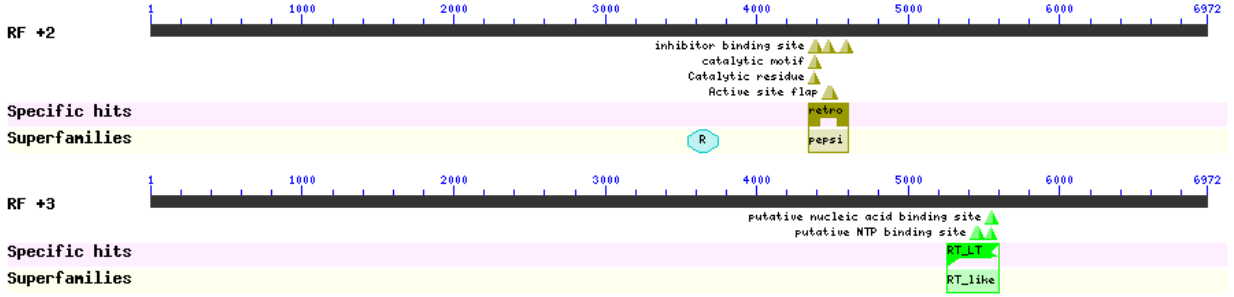


Chr_10c_nlr_18_Ca partial (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
TIR	pfam01582	TIR domain; The Toll/interleukin-1 receptor (TIR)	4-450	1.05e-54
FHY3 super family	cl31971	Protein FAR-RED ELONGATED HYPOCOTYL 3	3598-3861	5.24e-22
FAR1 super family	cl40636	FAR1 DNA-binding domain	3074-3334	1.24e-19
FHY3 super family	cl31971	Protein FAR-RED ELONGATED HYPOCOTYL 3	2936-3610	7.11e-17
FHY3 super family	cl31971	Protein FAR-RED ELONGATED HYPOCOTYL 3	3861-5096	1.62e-62
PH-like super family	cl17171	Pleckstrin homology-like domain; The PH-like family includes the PH domain	9019-9207	4.83e-08

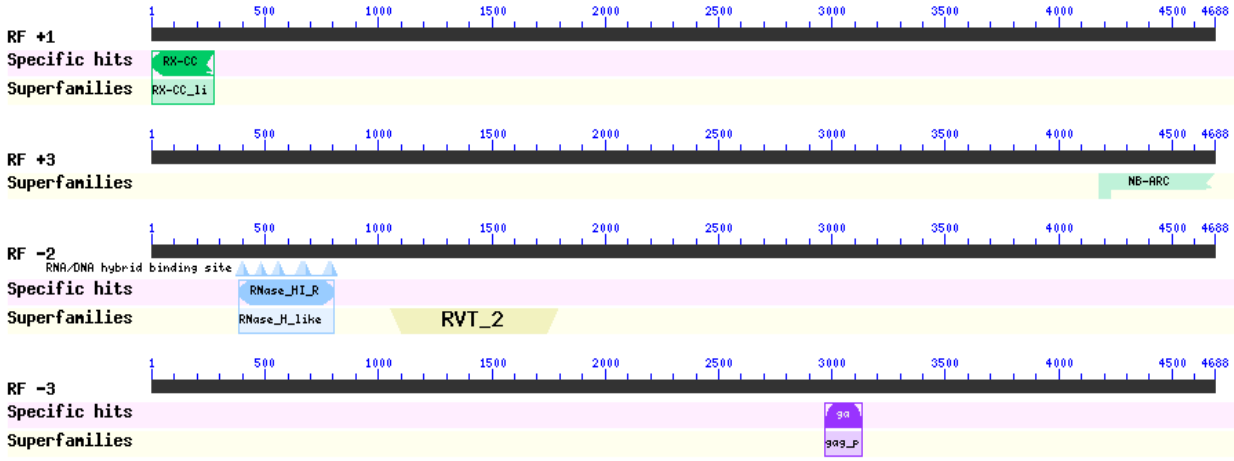


Chr_8c_nlr_13_Ca complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	6069-6800	2.16e-76

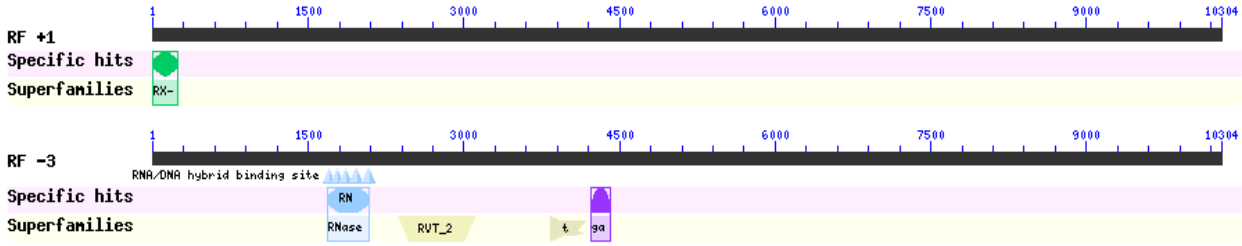
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	7053-7469	3.08e-74
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	4710-4874	1.19e-08
NB-ARC super family	cl26397	NB-ARC domain;	309-554	2.54e-06



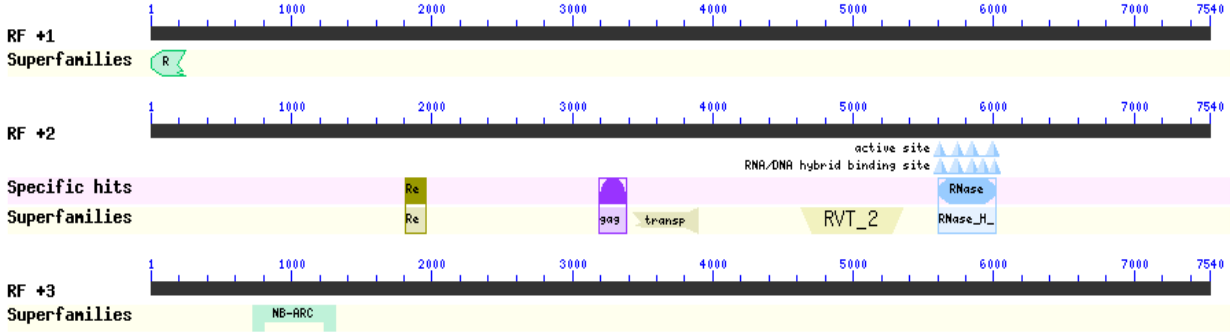
Chr_11c_nlr_73_Ca partial (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
Retropepsin_like	cd00303	Retropepsins; pepsin-like aspartate proteases	4346-4606	5.67e-11
Retrotrans_gag super family	cl29674	Retrotransposon gag protein; Gag or Capsid-like proteins from LTR retrotransposons	3542-3745	1.56e-07
RT_LTR	cd01647	RT_LTR: Reverse transcriptases (RTs)	5256-5597	2.09e-40



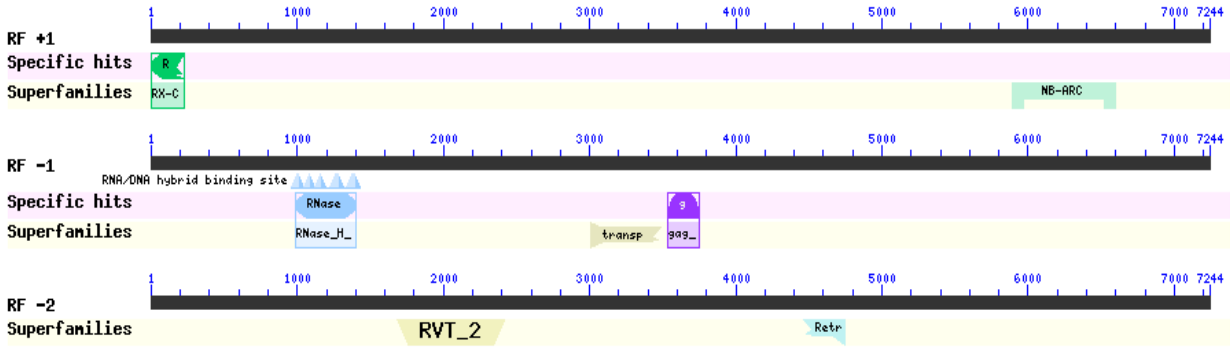
Chr_6e_nlr_10_Ca partial (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RX-CC_like	cd14798	Coiled-coil domain of the potato virus X resistance protein and similar proteins	1-273	3.80e-15
NB-ARC super family	cl26397	NB-ARC domain	4176-4685	2.49e-12
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	386-802	4.93e-78
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	1055-1786	1.73e-68
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	2965-3132	1.08e-06



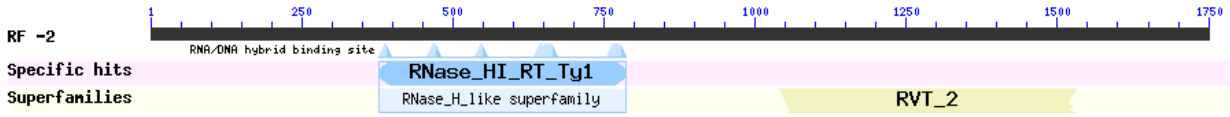
Chr_1e_nlr_61_Ca complete (pseudogene)				
Name	Accession	Description	Interval	E-value
Rx_N	pfam18052	This entry represents the N-terminal domain found in many plant resistance proteins	1-243	9.30e-08
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	2365-3102	1.26e-70
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	1678-2091	3.28e-64
Transpos_IS481 super family	cl41329	IS481 family transposase	3832-4170	7.73e-17
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	4222-4410	1.42e-12



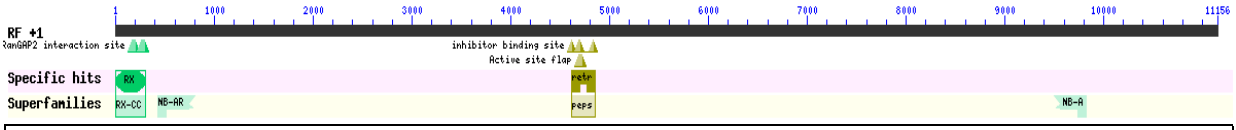
Chr_11e_nlr_43_Ca complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RX-CC_like super family	cl36576	Coiled-coil domain of the potato virus X resistance protein and similar proteins	1-249	9.15e-09
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	4622-5350	3.81e-78
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	5600-6019	4.36e-77
Retrotran_gag_3	pfam14244	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	1814-1954	4.42e-15
Transpos_IS481 super family	cl41329	IS481 family transposase	3431-3889	2.46e-10
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	3191-3382	1.51e-08
NB-ARC super family	cl26397	NB-ARC domain	723-1313	1.69e-39



Chr_10e_nlr_2_Ca complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
NB-ARC super family	cl26397	NB-ARC domain	5896-6600	5.29e-44
RX-CC_like	cd14798	Coiled-coil domain of the potato virus X resistance protein and similar proteins	1-225	1.03e-16
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	993-1400	8.74e-66
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	3534-3746	6.29e-12
Transpos_IS481 super family	cl41329	IS481 family transposase	3012-3482	7.08e-11
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	1679-2419	1.18e-48
Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	4460-4741	2.61e-09

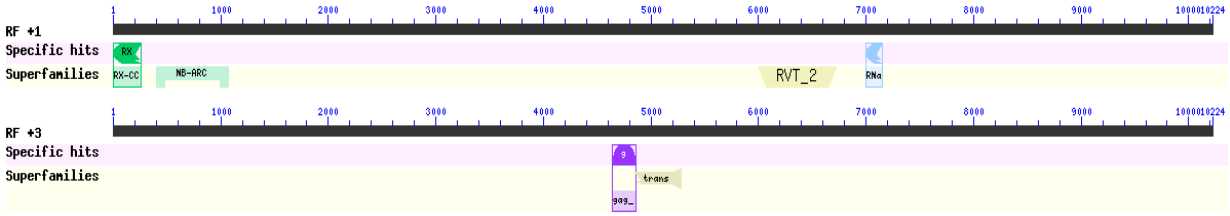


chr0_nlr_300_Cc partial (pseudogene)				
Name	Accession	Description	Interval	E-value
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	378-785	3.94e-67
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	1041-1529	4.39e-47

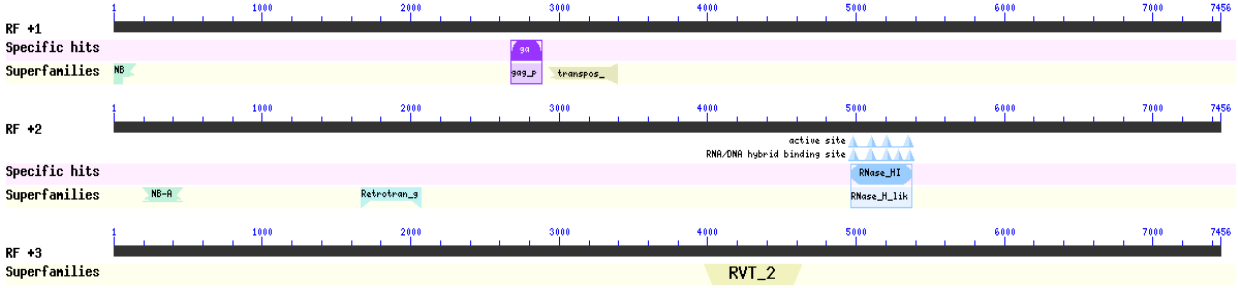


chr8_nlr_26_Cc complete (pseudogene)				
Name	Accession	Description	Interval	E-value
NB-ARC super family	cl26397	NB-ARC domain	427-801	1.88e-22
NB-ARC super family	cl26397	NB-ARC domain	9502-9828	2.21e-15

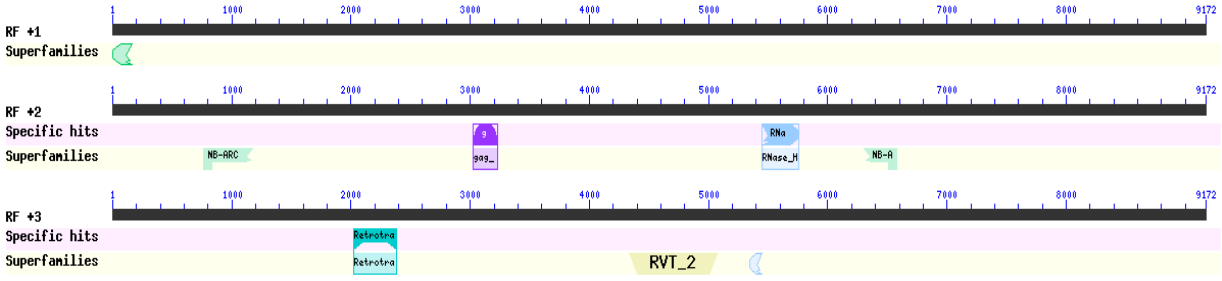
RX-CC_like	cd14798	Coiled-coil domain of the potato virus X resistance protein and similar proteins	4-300	2.90e-11
Retropepsin_like	cd00303	Retropepsins; pepsin-like aspartate proteases; The family includes pepsin-like aspartate proteases from retroviruses, retrotransposons and retroelements	4612-4851	3.61e-06



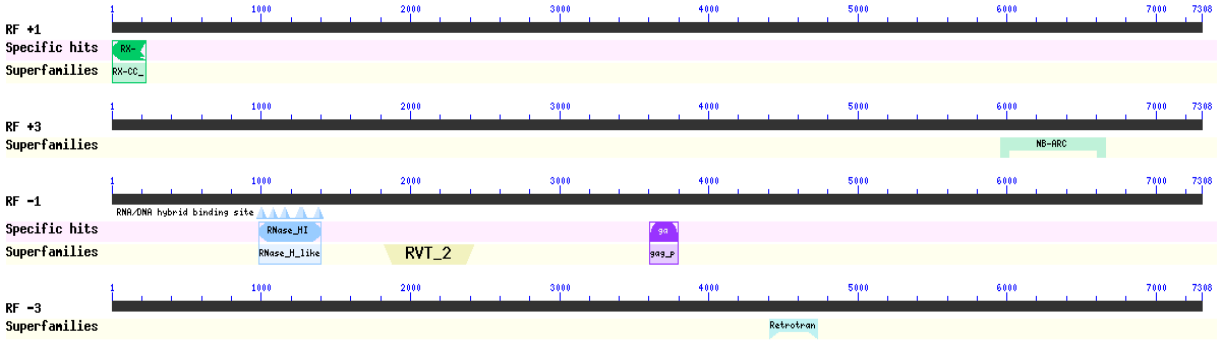
chr7_nlr_33_Cc complete (pseudogene)				
Name	Accession	Description	Interval	E-value
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	5992-6726	3.87e-67
NB-ARC super family	cl26397	NB-ARC domain	403-1068	1.85e-50
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	7000-7149	1.94e-19
RX-CC_like	cd14798	Coiled-coil domain of the potato virus X resistance protein and similar proteins	1-261	1.47e-11
Transpos_IS481 super family	cl41329	IS481 family transposase	4848-5276	1.29e-14
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	4638-4862	3.10e-12



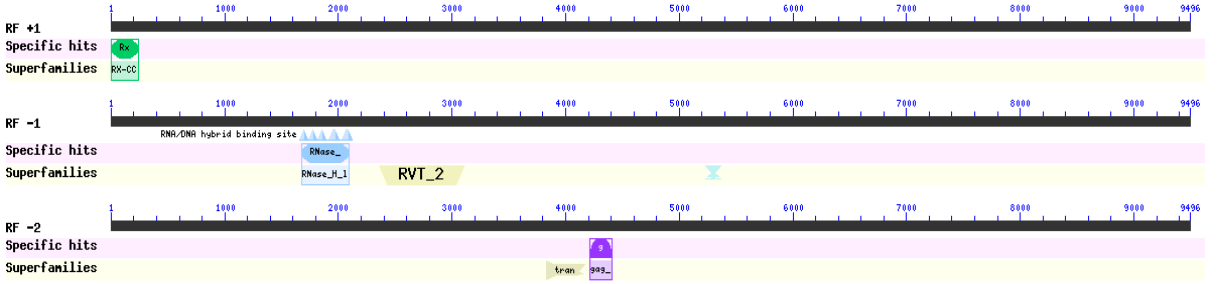
chr4_nlr_22 Cc complete				
Name	Accession	Description	Interval	E-value
Transpos_IS481 super family	cl41329	IS481 family transposase	2935-3393	1.65e-13
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	2671-2886	4.35e-11
NB-ARC super family	cl26397	NB-ARC domain	1-147	8.67e-07
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	4964-5377	1.25e-52
Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	1664-2068	6.38e-22
NB-ARC super family	cl26397	NB-ARC domain	194-457	4.89e-12
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	3978-4631	7.19e-46



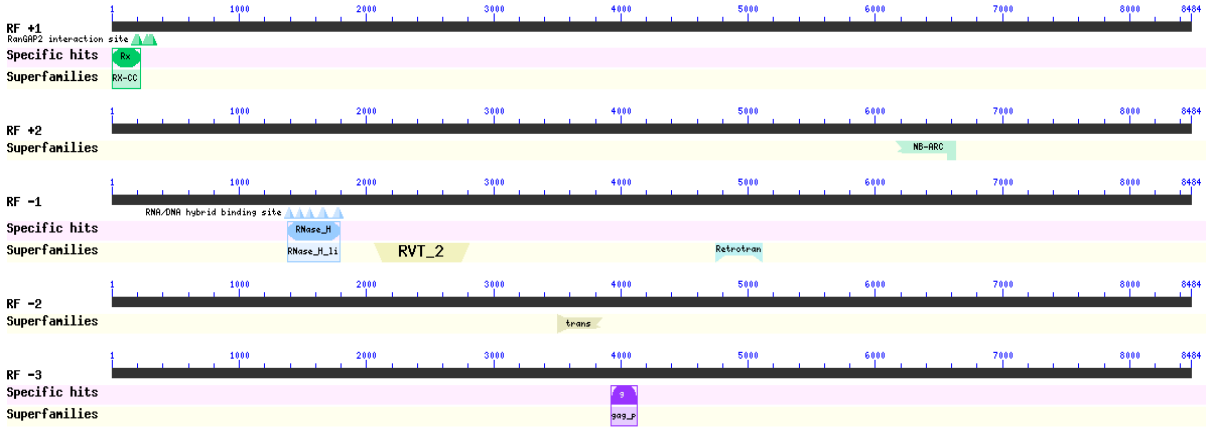
Chr_5_nlr_92_Ce complete				
Name	**Accession**	**Description**	**Interval**	**E-value**
RX-CC_like super family	cl36576	Coiled-coil domain of the potato virus X resistance protein and similar proteins	1-165	1.40e-07
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	5450-5752	1.79e-44
NB-ARC super family	cl26397	NB-ARC domain	767-1171	1.40e-26
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	3026-3229	8.44e-13
NB-ARC super family	cl26397	NB-ARC domain	6305-6580	5.76e-09
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	4332-5072	3.19e-69
Retrotran_gag_2	pfam14223	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	2022-2381	3.58e-31
RNase_H_like super family	cl14782	Ribonuclease H-like superfamily, including RNase H, HI, HII, HIII, and RNase-like domain IV	5343-5444	2.92e-08



Chr_10_nlr_19_Ce complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RX-CC_like	cd14798	Coiled-coil domain of the potato virus X resistance protein and similar proteins; The potato	1-225	8.90e-19
NB-ARC super family	cl26397	NB-ARC domain	5955-6662	2.74e-43
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	986-1399	1.04e-62
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	1823-2422	1.44e-40
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	3602-3796	2.47e-09
Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	4407-4730	2.75e-10

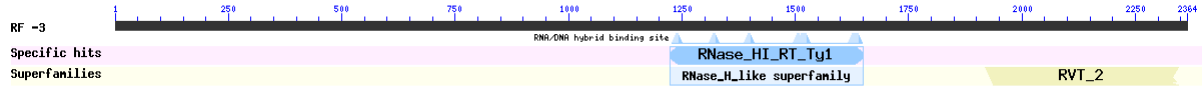


Chr_1_nlr_53_Ce complete (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
Rx_N	pfam18052	This entry represents the N-terminal domain found in many plant resistance proteins	1-243	8.01e-10
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	2363-3100	2.71e-68
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	1676-2089	1.39e-64
Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	5228-5365	4.61e-07
Transpos_IS481 super family	cl41329	IS481 family transposase	3826-4164	1.06e-15
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	4216-4407	1.18e-12

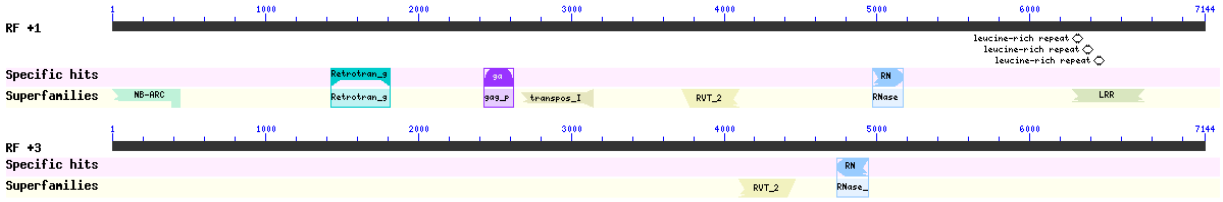


Chr_8_nlr_53 complete (pseudogene)				
Name	Accession	Description	Interval	E-value
Rx_N	pfam18052	This entry represents the N-terminal domain found in many plant resistance proteins	4-225	1.00e-16
NB-ARC super family	cl26397	NB-ARC domain	6167-6631	6.51e-24
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	2062-2802	1.81e-65
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	1381-1791	2.54e-57

Retrotran_gag_2 super family	cl26047	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	4744-5106	2.87e-22
Transpos_IS481 super family	cl41329	IS481 family transposase	3507-3848	3.91e-10
gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	3923-4126	4.06e-13



Chr_8_nlr_67_Ce partial (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H (RNase H)	1224-1649	7.59e-41
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	1920-2345	8.89e-26



Chr_11_nlr_129_Ce partial (pseudogene)				
Name	**Accession**	**Description**	**Interval**	**E-value**
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	3721-4095	1.01e-43
Retrotran_gag_2	pfam14223	gag-polypeptide of LTR copia-type; This family is found in Plants and fungi	1429-1815	2.51e-29
NB-ARC super family	cl26397	NB-ARC domain	1-438	5.97e-26
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	4972-5169	5.72e-23
Transpos_IS481 super family	cl41329	IS481 family transposase	2674-3144	4.64e-18
Gag_pre-integrs	pfam13976	GAG-pre-integrase domain; This domain is found associated with retroviral insertion elements	2431-2622	5.34e-10
LRR super family	cl34836	Leucine-rich repeat (LRR) protein [Transcription]	6277-6747	2.28e-07
RNase_HI_RT_Ty1	cd09272	Ty1/Copia family of RNase HI in long-term repeat retroelements; Ribonuclease H	4737-4943	4.01e-30
RVT_2 super family	cl06662	Reverse transcriptase (RNA-dependent DNA polymerase)	4095-4466	5.86e-23

Supplementary Figure 2. Conserved domains analysis for loci from NLR-annotator that did not present homology to NLRs proteins by BLASTx analysis. For the analysis, nucleotide sequences were used. The table shows the list of domain hits found for each of the sequences and the domains related to NLR proteins are highlighted in green. The conserved domains were detected using NCBI Conserved Domain Database and the graphical summary was set to concise view. If the alignment omitted more than 20% of the either the n- or c-terminus or both, the partial nature of the hit is indicated in the graphical display as domain with jagged edges. At the end of each ID, the source genome is identified, Ca: *C. arabica*, Cc: *C. eugenioides* and Ce: *C. eugenioides*. RF: reading frame.

Supplementary Figure 3. chromosomal distribution of NLR loci in *C. arabica*. For more details, mouse-over the picture.
To view the figure, access the link and download the HTML file:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Supplementary Figure 4. chromosomal distribution of NLR loci in *C. canephora*. For more details, mouse-over the picture.
To view the figure, access the link and download the HTML file:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Supplementary Figure 5. chromosomal distribution of NLR loci in *C. eugenioides*. For more details, mouse-over the picture.
To view the figure, access the link and download the HTML file:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Supplementary Table 1. Motif from Jupe et al., 2012 and NB-ARC motif combinations from Steuernagel et al., 2020.

Motif Table from Jupe et al., 2012				
Motif	**Sequence**	**Domain**	**Group**	**Similar to**
1	PIWGMGGVGKTTLARAVYNDP	NB-ARC	CNL/TNL	P-loop
2	LKPCFLYCAIFPEDYMIDKNKLIWLWMAE	NB-ARC	CNL	RNBS-D
3	CGGLPLAIKVWGGMLAGKQKT	NB-ARC	CNL/TNL	GLPL
4	YLVVLDDVWDTDQWD	NB-ARC	CNL/TNL	Kin-2
5	NGSRIIITTRNKHVANYMCT	NB-ARC	CNL/TNL	RNBS-B
6	HFDCRAWVCVSQQYDMKKVLRDIIQQVGG	NB-ARC	CNL	RNBS-A
7	CRMHDMMHDMCWYKAREQNFV	linker	CNL/TNL	MHDV
8	MEDVGEYYFNELINRSMFQPI	linker	CNL/TNL	-
9	LIHLRYLNLSGTNIKQLPASI	LRR1	CNL/TNL	Motif1 LDL
10	LSHEESWQLFHQHAF	NB-ARC	CNL/TNL	RNBS-C
11	MPNLETLDIHNCPNLEEIP	LRR	CNL/TNL	-
12	IMPVLRLSYHHLPYH	NB-ARC	CNL/TNL	-
13	QIVIPIFYDVDPSDVRHQTGSFGEAFWKHCSR	TIR	TNL	TIR-3
14	AIKDIQEQLQKVADRRDRNKVFVPHPTRPIAIDPCLRALYAEATELVGIY	monocot	-	-
15	KNYATSRWCLNELVKIMECKE	TIR	TNL	TIR-2
16	DAAYDAEDVIDSFKYHA	pre-NB	CNL	EDVID
17	FAIPKLGDFLTQEYYLHKGIKKEIEWLKRELEFMQA	pre-NB	CNL	-
18	KYDVFLSFRGADTRRTFTSHLYEALKNRGINTF	TIR	TNL	TIR-1
19	IKMVEITGYRGTRFPNWMGHPVYCNMVSISIRNCKNCSCLP	LRR	CNL/TNL	-
20	ETSSFELMDLLGERWVPPVHLREFKSFMPSQLSALRGWIQRDPSHLSNLS	monocot	-	-
NB-ARC Motif Combinations from Steuernagel, et al. 2020				
1,6,4				
1,4,5				
6,4,5				
5,10,3				
10,3,12				
3,12,2				
Consensus structure of an NLR From Steuernagel et al., 2020				
				

Supplementary Table 2. Reference NLRs. All reference NRLs were obtained from the PRGDB. The reference NLRs highlighted in light blue and dark blue refer to NLRs that share orthogroups with coffee, according to table S10

Reference NLRs				
Plants				
PRGdb ID	**Gene name**	**Class**	**Gene type**	**Species**
PRGDB130	Bs2	CNL	reference	Capsicum chacoense
PRGDB138	Gpa2	CNL	reference	Solanum tuberosum
PRGDB140	Hero	CNL	reference	Solanum lycopersicum
PRGDB142	Mi1.2	CNL	reference	Solanum lycopersicum
PRGDB146	R1	CNL	reference	Solanum demissum
PRGDB148	Rpi-blb1	CNL	reference	Solanum bulbocastanum
PRGDB1484	HRT	CNL	reference	Arabidopsis thaliana
PRGDB1488	Pi-ta	CNL	reference	Oryza sativa
PRGDB1489	RCY1	CNL	reference	Arabidopsis thaliana
PRGDB149	Rpi-blb2	CNL	reference	Solanum bulbocastanum
PRGDB1491	RPM1	CNL	reference	Arabidopsis thaliana
PRGDB1492	RPP13	CNL	reference	Arabidopsis thaliana
PRGDB1494	RPP8	CNL	reference	Arabidopsis thaliana
PRGDB1495	Rps2	CNL	reference	Arabidopsis thaliana
PRGDB1497	RPS5	CNL	reference	Arabidopsis thaliana
PRGDB150	Rx	CNL	reference	Solanum tuberosum
PRGDB150957	PIB	CNL	reference	Oryza sativa
PRGDB150958	XA1	CNL	reference	Oryza sativa
PRGDB151	Rx2	CNL	reference	Solanum acaule
PRGDB153	Sw-5	CNL	reference	Solanum lycopersicum
PRGDB154	Tm-2a	CNL	reference	Solanum lycopersicum
PRGDB157	Tm-2	CNL	reference	Solanum lycopersicum
PRGDB161432	MLA1	CNL	reference	Hordeum vulgare
PRGDB161433	Mla6	CNL	reference	Hordeum vulgare subsp. vulgare
PRGDB135725	MLA10	CNL	reference	Hordeum vulgare
PRGDB161434	Mla12	CNL	reference	Hordeum vulgare subsp. vulgare
PRGDB161435	MLA13	CNL	reference	Hordeum vulgare
PRGDB161436	Pi36	CNL	reference	Oryza sativa Indica Group
PRGDB161437	Rp1-D	CNL	reference	Zea mays
PRGDB161438	Pm3	CNL	reference	Triticum aestivum
PRGDB161439	Lr10	CNL	reference	Triticum aestivum
PRGDB161440	Pl8	CNL	reference	Helianthus annuus
PRGDB161442	Pi9	CNL	reference	Oryza sativa Indica Group
PRGDB161443	Piz-t	CNL	reference	Oryza sativa Japonica Group
PRGDB161444	Pi2	CNL	reference	Oryza sativa Indica Group
PRGDB161446	Cre1	CNL	reference	Aegilops tauschii
PRGDB161453	Pikm1-TS	CNL	reference	Oryza sativa Japonica Group
RGDB161454	Pikm2-TS	CNL	reference	Oryza sativa Japonica Group
PRGDB161456	Rdg2a	CNL	reference	Hordeum vulgare subsp. vulgare
PRGDB161457	Pid3	CNL	reference	Oryza sativa Japonica Group

PRGDB161458	Pi5-1	CNL	reference	Oryza sativa Japonica Group
PRGDB161459	Pi5-2	CNL	reference	Oryza sativa Japonica Group
PRGDB161460	Pit	CNL	reference	Oryza sativa Japonica Group
PRGDB161465	FOM-2	CNL	reference	Cucumis melo
PRGDB161468	Lr21	CNL	reference	Triticum aestivum
PRGDB161469	Lr1	CNL	reference	Triticum aestivum
PRGDB161471	VAT	CNL	reference	Cucumis melo
PRGDB145	Prf	CNL	reference	Solanum pimpinellifolium
PRGDB131	Bs4	TNL	reference	Solanum lycopersicum
PRGDB135724	P2	TNL	reference	Linum usitatissimum
PRGDB139	Gro1.4	TNL	reference	Solanum tuberosum
PRGDB1486	L6	TNL	reference	Linum usitatissimum
PRGDB1487	M	TNL	reference	Linum usitatissimum
PRGDB1493	RPP5	TNL	reference	Arabidopsis thaliana
PRGDB1496	Rps4	TNL	reference	Arabidopsis thaliana
PRGDB150959	RPP1	TNL	reference	Arabidopsis thaliana
PRGDB150960	RPP4	TNL	reference	Arabidopsis thaliana
PRGDB150963	RRS1	TNL	reference	Arabidopsis thaliana
PRGDB152	RY-1	TNL	reference	Solanum tuberosum subsp andigena
PRGDB161441	SSI4	TNL	reference	Arabidopsis thaliana
PRGDB161464	KR1	TNL	reference	Glycine max
PRGDB161467	RAC1	TNL	reference	Arabidopsis thaliana
PRGDB144	N	TNL	reference	Nicotiana glutinosa
PRGDB141	I-2	NL	reference	Solanum lycopersicum
PRGDB1500	Rps1-k-2	NL	reference	Glycine max
PRGDB1501	Rps1-k-1	NL	reference	Glycine max
PRGDB147	R3a	NL	reference	Solanum tuberosum
NCBI ID				**Species**
NP_001021202.1	CED-4	-	-	Caenorhabditis elegans

Supplementary Table 3. Txt output file from NLR-annotator analysis in the genomes of *C. arabica, C. canephora* and *C.eugenioides.* The NLR loci in the same genomic position that were identified as distinct NLRs by NLR-annotator are highlighted in blue. The complete loci that do not overlap with gene models from the reference genomes are highlighted in orange. Complete loci not identified by AUGUSTUS as putative genes are in red
To view the table, access the link:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Supplementary Table 4. PfamScan analysis for the set of predicted proteins belonging to the NB-ARC family. The PfamSancan analysis highlighted in blue shows all the predicted NB-ARC domains for each of the proteins, which occurred once or more than once. The "description gene/protein" highlighted in purple shows each of the NLR proteins (including isoforms), their description, and the gene encoding it.
To view the table, access the link:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Supplementary Table 5. Overlap analysis between NLR loci annotated by NLR-annotator and NLR genes model annotated in each genome. Regions that have more than one overlap are highlighted in orange, and the ID of the genes or NLR-annotator locus that overlaps in more than one region is in bold.
To view the table, access the link:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Table S6. NLR-Parser analysis of gene models that did not overlap with loci from NLR-annotator. Genes classified as not found are highlighted in orange.

C. arabica						
Chr	**start**	**end**	**strand**	**Gene ID**	**Situation**	**Motif List - NLR-Parser**
1c	6512878	6523148	-	LOC113736959	without 3 consecutive NB-ARC associated motifs	16,1,6,10,3,2,8,9,9,11,9,11,11,11,11,11,11,11,9,11
1c	7444225	7444668	+	LOC113737176	not found	N/A
1e	3276674	3277597	-	LOC113688369	below the threshold	N/A
1e	3750078	3754745	-	LOC113698956	without 3 consecutive NB-ARC associated motifs	16,1,6,10,3,2,2,8,9,9,11,9,11,11,11,11,11,11,11,9,11
3c	35930310	35942468	+	LOC113735021	below the threshold	N/A
3c	36025944	36044624	+	LOC113735022	without 3 consecutive NB-ARC associated motifs	4,11,18,1,10,3,2,9,11,9
3c	36310659	36357338	+	LOC113735027	without 3 consecutive NB-ARC associated motifs	8,2,1,6,3,2,9,11,9,4
3c	38706643	38707176	-	LOC113735982	not found	N/A
3e	33116426	33117508	-	LOC113737673	below the threshold	N/A
3e	33157516	33191039	+	LOC113736865	without 3 consecutive NB-ARC associated motifs	8,6,5,8,6,12,2,1,3,2,9,11,9,11
3e	33416306	33417337	-	LOC113737677	below the threshold	N/A
8e	5996662	5998340	-	LOC113704444	below the threshold	N/A
8e	6113951	6115675	-	LOC113704447	without 3 consecutive NB-ARC associated motifs	11,9,11
7e	1319762	1327541	+	LOC113701944	below the threshold	N/A
8c	5007613	5009028	-	LOC113705702	below the threshold	N/A
8c	5031746	5035876	-	LOC113705703	without 3 consecutive NB-ARC associated motifs	6,8,11,11,9,11,9,11,9,9,11,11,9,9,9,11,9,11,11,11,11,11,11,11,11
8c	5292356	5297998	-	LOC113705523	without 3 consecutive NB-ARC associated motifs	1,11,9,11,9,11,9,9,9,11,11,9,9,9,11,9,11,11,11,11,11,11,11,11,11
10c	12139259	12145211	-	LOC113714883	without 3 consecutive NB-ARC associated motifs	6,8,11,9,11,9,11,9,9,11,11,9,9,11,9,11,11,11,11,11,11,11,11
C.canephora						
Chr	**start**	**end**	**strand**	**Gene ID**	**Situation**	**Motif List - NLR-Parser**
1	3052489	3076874	+	Cc01_g01770	without 3 consecutive NB-ARC associated motifs	17,15,1,4,9,9,11,11,11,11,11,11,11,11,11,11,11,18
1	7030448	7034678	-	Cc01_g03240	without 3 consecutive NB-ARC associated motifs	16,1,6,10,3,2,8,9,9,11,9,11,11,11,11,11,11,11,11,9,11
2	10176400	10178449	+	Cc02_g12220	not found	
3	13384680	13386551	+	Cc03_g10360	not found	
3	26987218	26999539	+	Cc03_g13770	below the threshold	N/A
3	27099395	27108875	+	Cc03_g13800	without 3 consecutive NB-ARC associated motifs	1,10,3,2,9,11,9

5	21717343	21718097	+	Cc05_g06950	without 3 consecutive NB-ARC associated motifs	17,16,1
5	23644535	23650504	+	Cc05_g09100	without 3 consecutive NB-ARC associated motifs	1,10,11,9,11,9,9,9,11,9,11,11,9,9,11,9,11,11,11,11,11,12
7	2123155	2123892	+	Cc07_g03110	without 3 consecutive NB-ARC associated motifs	17,16,1
7	4004632	4008454	+	Cc07_g05700	below the threshold	N/A
7	16681864	16682166	+	Cc07_g18800	not found	
7	24709506	24714731	-	Cc07_g20750	below the threshold	N/A
8	2676910	2677614	-	Cc08_g02190	without 3 consecutive NB-ARC associated motifs	17,16,1
0	58890586	58894463	+	Cc00_g07290	without 3 consecutive NB-ARC associated motifs	17,16,1,6,19,8,7,9,9,11,10,11
0	84715359	84718639	-	Cc00_g10090	without 3 consecutive NB-ARC associated motifs	1,6,8,11,9,11,9,9,9,9,11,11,9,9,11,9,11
0	116040312	116043383	+	Cc00_g17360	without 3 consecutive NB-ARC associated motifs	17,1,10,3,2,9,11,9,11,11,11,11
0	138585790	138588278	-	Cc00_g21830	without 3 consecutive NB-ARC associated motifs	1,6,8,11,9,11,9,11,9,9,11,11,9
0	139277121	139277483	-	Cc00_g21910	not found	N/A
0	156558503	156559243	+	Cc00_g24980	without 3 consecutive NB-ARC associated motifs	17,16,1,6
0	165812081	165812791	+	Cc00_g26400	without 3 consecutive NB-ARC associated motifs	17,16,1,6
0	181282461	181283961	+	Cc00_g29640	without 3 consecutive NB-ARC associated motifs	17,16,1,6,11,11
0	183665123	183667650	-	Cc00_g30260	without 3 consecutive NB-ARC associated motifs	1,8,11,9,11,9,11,9
0	184677779	184678558	-	Cc00_g30580	without 3 consecutive NB-ARC associated motifs	17,16,1,6
0	203517297	203517947	-	Cc00_g35410	without 3 consecutive NB-ARC associated motifs	17,16,1
0	203550867	203551550	-	Cc00_g35420	not found	N/A
C. eugenioides						
Chr	**start**	**end**	**strand**	**Gene ID**	**Situation**	**Motif List - NLR-Parser**
11	22314608	22315932	-	LOC113752059	without 3 consecutive NB-ARC associated motifs	17,16,1,6
11	50613214	50617035	+	LOC113753345	below the threshold	N/A
3	16302436	16317186	-	LOC113766181	without 3 consecutive NB-ARC associated motifs	4,5,18,1,10,3,2,9,11,9
3	40629207	40630439	+	LOC113766615	not found	1,4,5,3
8	7316120	7325726	-	LOC113780105	without 3 consecutive NB-ARC associated motifs	1,11,9,11,9,11,9,9,9,11,11,20,9,11,9,2
8	7362566	7364434	-	LOC113780107	without 3 consecutive NB-ARC associated motifs	1,11,9,11,9,11
8	7423952	7429039	+	LOC113780108	without 3 consecutive NB-ARC associated motifs	6,11,11,9,11,9,9,9,9,11,11,9,11,9,11,11,11,11,11,11,11
8	7476020	7481925	+	LOC113779458	without 3 consecutive NB-ARC associated motifs	6,8,11,9,11,9,11,9,9,11,11,9,9,9,11,9,11,11,11,11,11,11,11,11,11
3	46512774	46514805	+	LOC113766765	without 3 consecutive NB-ARC associated motifs	1,10,3,2,18,9,11
3	46604388	46606692	+	LOC113766766	without 3 consecutive NB-ARC associated motifs	1,10,3,2,18,9,11,9

3	46638024	46657976	-	LOC113766768	without 3 consecutive NB-ARC associated motifs	20,1,10,3,2,9,11,9,6
3	46782057	46783106	-	LOC113766771	not found	N/A
3	46938907	46947033	-	LOC113766774	not found	N/A
1	5822459	5826205	-	LOC113766334	without 3 consecutive NB-ARC associated motifs	16,1,6,10,3,2,2,8,9,9,11,9,11,11,11,11,11,11,11,9,11
7	22108032	22112852	-	LOC113777141	not found	1,6,4
3	47011507	47013825	-	LOC113766777	without 3 consecutive NB-ARC associated motifs	1,10,3,2,9,11,9
3	47802769	47824783	-	LOC113766786	without 3 consecutive NB-ARC associated motifs	2,4,1,10,3,2,9,11,9,18,20,10,13
3	48042814	48058270	-	LOC113766791	below the threshold	N/A
5	40196396	40203450	-	LOC113771639	below the threshold	N/A
5	45613260	45625908	+	LOC113771845	without 3 consecutive NB-ARC associated motifs	1,10,11,9,11,9,9,11,9,11,11,9,9,11,9,11,11,11,11,11,11,12,2,8
2	31850051	31851919	+	LOC113759493	below the threshold	N/A
2	54134203	54136209	-	LOC113759896	without 3 consecutive NB-ARC associated motifs	1,10,3,2,9,11
2	59401301	59403620	-	LOC113759976	without 3 consecutive NB-ARC associated motifs	1,10,3,2,18,9,11,9
Un	3866	6528	+	LOC113758937	without 3 consecutive NB-ARC associated motifs	1,10,11,9,11,9,9,11,9,11,11,9,9

Table S7. BLASTx analysis for *C.arabica*. Complete loci are highlighted in green, and loci that did not present homology with NLR proteins are highlighted in orange.
To view the table, access the link: https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Table S8. BLASTx analysis for *C.canephora*. Complete loci are highlighted in green, and loci that did not present homology with NLR proteins are highlighted in orange.
To view the table, access the link: https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Table S9. BLASTx analysis for *C. eugenioides*. Complete loci are highlighted in green, and loci that did not present homology with NLR proteins are highlighted in orange.
To view the table, access the link: https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Table S10. NLRs orthogroups defined by OrthoMCL analysis.
To view the table, access the link:
https://1drv.ms/u/s!As084N7WlXAIhqJVvP2MB4rLi6jQoQ?e=5aSbCP

Table S11. NLRs orphans. These NLRs are the ones not assigned to any ortholog group by OrthoMCL (S10 Table).

C.arabica	C.canephora	*C. eugenioides*
Ca_NC_03989.1_Chr_1c_nlr_10	Cc_chr0_nlr_10	Ce_NC_040035.1_Chr_1_nlr_16
Ca_NC_039898.1_Chr_1c_nlr_11	Cc_chr0_nlr_106	Ce_NC_040035.1_Chr_1_nlr_18
Ca_NC_039898.1_Chr_1c_nlr_14	Cc_chr0_nlr_109	Ce_NC_040035.1_Chr_1_nlr_2
Ca_NC_039898.1_Chr_1c_nlr_16	Cc_chr0_nlr_113	Ce_NC_040035.1_Chr_1_nlr_3
Ca_NC_039898.1_Chr_1c_nlr_17	Cc_chr0_nlr_114	Ce_NC_040035.1_Chr_1_nlr_37
Ca_NC_039898.1_Chr_1c_nlr_19	Cc_chr0_nlr_117	Ce_NC_040035.1_Chr_1_nlr_4
Ca_NC_039898.1_Chr_1c_nlr_2	Cc_chr0_nlr_126	Ce_NC_040035.1_Chr_1_nlr_60
Ca_NC_039898.1_Chr_1c_nlr_27	Cc_chr0_nlr_129	Ce_NC_040035.1_Chr_1_nlr_61
Ca_NC_039898.1_Chr_1c_nlr_30	Cc_chr0_nlr_131	Ce_NC_040035.1_Chr_1_nlr_64
Ca_NC_039898.1_Chr_1c_nlr_35	Cc_chr0_nlr_141	Ce_NC_040035.1_Chr_1_nlr_65
Ca_NC_039898.1_Chr_1c_nlr_36	Cc_chr0_nlr_159	Ce_NC_040035.1_Chr_1_nlr_74
Ca_NC_039898.1_Chr_1c_nlr_57	Cc_chr0_nlr_165	Ce_NC_040036.1_Chr_2_nlr_4
Ca_NC_039898.1_Chr_1c_nlr_59	Cc_chr0_nlr_169	Ce_NC_040036.1_Chr_2_nlr_42
Ca_NC_039898.1_Chr_1c_nlr_60	Cc_chr0_nlr_175	Ce_NC_040036.1_Chr_2_nlr_5
Ca_NC_039898.1_Chr_1c_nlr_66	Cc_chr0_nlr_18	Ce_NC_040037.1_Chr_3_nlr_1
Ca_NC_039898.1_Chr_1c_nlr_71	Cc_chr0_nlr_208	Ce_NC_040037.1_Chr_3_nlr_107
Ca_NC_039899.1_Chr_1e_nlr_1	Cc_chr0_nlr_212	Ce_NC_040037.1_Chr_3_nlr_116
Ca_NC_039899.1_Chr_1e_nlr_11	Cc_chr0_nlr_217	Ce_NC_040037.1_Chr_3_nlr_121
a_NC_039899.1_Chr_1e_nlr_18	Cc_chr0_nlr_221	Ce_NC_040037.1_Chr_3_nlr_124
Ca_NC_039899.1_Chr_1e_nlr_21	Cc_chr0_nlr_239	Ce_NC_040037.1_Chr_3_nlr_127
Ca_NC_039899.1_Chr_1e_nlr_26	Cc_chr0_nlr_242	Ce_NC_040037.1_Chr_3_nlr_132
Ca_NC_039899.1_Chr_1e_nlr_30	Cc_chr0_nlr_247	Ce_NC_040037.1_Chr_3_nlr_134
Ca_NC_039899.1_Chr_1e_nlr_33	Cc_chr0_nlr_248	Ce_NC_040037.1_Chr_3_nlr_139
Ca_NC_039899.1_Chr_1e_nlr_44	Cc_chr0_nlr_25	Ce_NC_040037.1_Chr_3_nlr_20
Ca_NC_039899.1_Chr_1e_nlr_47	Cc_chr0_nlr_256	Ce_NC_040037.1_Chr_3_nlr_21
Ca_NC_039899.1_Chr_1e_nlr_49	Cc_chr0_nlr_262	Ce_NC_040037.1_Chr_3_nlr_24
Ca_NC_039899.1_Chr_1e_nlr_52	Cc_chr0_nlr_272	Ce_NC_040037.1_Chr_3_nlr_25
Ca_NC_039899.1_Chr_1e_nlr_53	Cc_chr0_nlr_277	Ce_NC_040037.1_Chr_3_nlr_26
Ca_NC_039899.1_Chr_1e_nlr_54	Cc_chr0_nlr_278	Ce_NC_040037.1_Chr_3_nlr_40
Ca_NC_039899.1_Chr_1e_nlr_55	Cc_chr0_nlr_28	Ce_NC_040037.1_Chr_3_nlr_65
Ca_NC_039899.1_Chr_1e_nlr_60	Cc_chr0_nlr_284	Ce_NC_040037.1_Chr_3_nlr_66
Ca_NC_039899.1_Chr_1e_nlr_8	Cc_chr0_nlr_289	Ce_NC_040037.1_Chr_3_nlr_67
Ca_NC_039900.1_Chr_2c_nlr_10	Cc_chr0_nlr_299	Ce_NC_040037.1_Chr_3_nlr_82
Ca_NC_039900.1_Chr_2c_nlr_26	Cc_chr0_nlr_304	Ce_NC_040037.1_Chr_3_nlr_91
Ca_NC_039901.1_Chr_2e_nlr_10	Cc_chr0_nlr_31	Ce_NC_040037.1_Chr_3_nlr_97
Ca_NC_039901.1_Chr_2e_nlr_23	Cc_chr0_nlr_314	Ce_NC_040038.1_Chr_4_nlr_25
Ca_NC_039901.1_Chr_2e_nlr_27	Cc_chr0_nlr_315	Ce_NC_040039.1_Chr_5_nlr_101
Ca_NC_039901.1_Chr_2e_nlr_31	Cc_chr0_nlr_316	Ce_NC_040039.1_Chr_5_nlr_104
Ca_NC_039901.1_Chr_2e_nlr_32	Cc_chr0_nlr_317	Ce_NC_040039.1_Chr_5_nlr_106

Ca_NC_039901.1_Chr_2e_nlr_33	Cc_chr0_nlr_327	Ce_NC_040039.1_Chr_5_nlr_109
Ca_NC_039902.1_Chr_3c_nlr_111	Cc_chr0_nlr_33	Ce_NC_040039.1_Chr_5_nlr_12
Ca_NC_039902.1_Chr_3c_nlr_113	Cc_chr0_nlr_335	Ce_NC_040039.1_Chr_5_nlr_2
Ca_NC_039902.1_Chr_3c_nlr_117	Cc_chr0_nlr_336	Ce_NC_040039.1_Chr_5_nlr_35
Ca_NC_039902.1_Chr_3c_nlr_127	Cc_chr0_nlr_339	Ce_NC_040039.1_Chr_5_nlr_36
Ca_NC_039902.1_Chr_3c_nlr_15	Cc_chr0_nlr_34	Ce_NC_040039.1_Chr_5_nlr_41
Ca_NC_039902.1_Chr_3c_nlr_19	Cc_chr0_nlr_342	Ce_NC_040039.1_Chr_5_nlr_43
Ca_NC_039902.1_Chr_3c_nlr_24	Cc_chr0_nlr_343	Ce_NC_040039.1_Chr_5_nlr_6
Ca_NC_039902.1_Chr_3c_nlr_26	Cc_chr0_nlr_344	Ce_NC_040039.1_Chr_5_nlr_70
Ca_NC_039902.1_Chr_3c_nlr_28	Cc_chr0_nlr_35	Ce_NC_040039.1_Chr_5_nlr_74
Ca_NC_039902.1_Chr_3c_nlr_33	Cc_chr0_nlr_362	Ce_NC_040039.1_Chr_5_nlr_76
Ca_NC_039902.1_Chr_3c_nlr_34	Cc_chr0_nlr_363	Ce_NC_040039.1_Chr_5_nlr_92
Ca_NC_039902.1_Chr_3c_nlr_37	Cc_chr0_nlr_365	Ce_NC_040039.1_Chr_5_nlr_97
Ca_NC_039902.1_Chr_3c_nlr_41	Cc_chr0_nlr_367	Ce_NC_040039.1_Chr_5_nlr_98
Ca_NC_039902.1_Chr_3c_nlr_51	Cc_chr0_nlr_37	Ce_NC_040040.1_Chr_6_nlr_12
Ca_NC_039902.1_Chr_3c_nlr_55	Cc_chr0_nlr_40	Ce_NC_040040.1_Chr_6_nlr_14
Ca_NC_039902.1_Chr_3c_nlr_56	Cc_chr0_nlr_41	Ce_NC_040040.1_Chr_6_nlr_15
Ca_NC_039902.1_Chr_3c_nlr_58	Cc_chr0_nlr_45	Ce_NC_040040.1_Chr_6_nlr_19
Ca_NC_039902.1_Chr_3c_nlr_61	Cc_chr0_nlr_52	Ce_NC_040041.1_Chr_7_nlr_42
Ca_NC_039902.1_Chr_3c_nlr_7	Cc_chr0_nlr_58	Ce_NC_040041.1_Chr_7_nlr_46
Ca_NC_039902.1_Chr_3c_nlr_93	Cc_chr0_nlr_70	Ce_NC_040041.1_Chr_7_nlr_59
Ca_NC_039902.1_Chr_3c_nlr_97	Cc_chr0_nlr_73	Ce_NC_040041.1_Chr_7_nlr_75
Ca_NC_039903.1_Chr_3e_nlr_12	Cc_chr0_nlr_76	Ce_NC_040042.1_Chr_8_nlr_103
Ca_NC_039903.1_Chr_3e_nlr_13	Cc_chr0_nlr_82	e_NC_040042.1_Chr_8_nlr_107
Ca_NC_039903.1_Chr_3e_nlr_14	Cc_chr0_nlr_86	Ce_NC_040042.1_Chr_8_nlr_116
Ca_NC_039903.1_Chr_3e_nlr_2	Cc_chr0_nlr_90	Ce_NC_040042.1_Chr_8_nlr_136
Ca_NC_039903.1_Chr_3e_nlr_20	Cc_chr0_nlr_92	Ce_NC_040042.1_Chr_8_nlr_139
Ca_NC_039903.1_Chr_3e_nlr_24	Cc_chr0_nlr_94	Ce_NC_040042.1_Chr_8_nlr_140
Ca_NC_039903.1_Chr_3e_nlr_26	Cc_chr0_nlr_97	Ce_NC_040042.1_Chr_8_nlr_16
Ca_NC_039903.1_Chr_3e_nlr_29	Cc_chr0_nlr_99	Ce_NC_040042.1_Chr_8_nlr_18
Ca_NC_039903.1_Chr_3e_nlr_45	Cc_chr1_nlr_17	Ce_NC_040042.1_Chr_8_nlr_2
Ca_NC_039903.1_Chr_3e_nlr_6	Cc_chr1_nlr_19	Ce_NC_040042.1_Chr_8_nlr_22
Ca_NC_039903.1_Chr_3e_nlr_69	Cc_chr1_nlr_2	Ce_NC_040042.1_Chr_8_nlr_28
Ca_NC_039903.1_Chr_3e_nlr_72	Cc_chr1_nlr_21	Ce_NC_040042.1_Chr_8_nlr_3
Ca_NC_039903.1_Chr_3e_nlr_8	Cc_chr1_nlr_25	Ce_NC_040042.1_Chr_8_nlr_46
Ca_NC_039903.1_Chr_3e_nlr_83	Cc_chr1_nlr_4	Ce_NC_040042.1_Chr_8_nlr_49
Ca_NC_039903.1_Chr_3e_nlr_89	Cc_chr1_nlr_43	Ce_NC_040042.1_Chr_8_nlr_58
Ca_NC_039903.1_Chr_3e_nlr_9	Cc_chr1_nlr_50	Ce_NC_040042.1_Chr_8_nlr_62
Ca_NC_039903.1_Chr_3e_nlr_92	Cc_chr1_nlr_51	Ce_NC_040042.1_Chr_8_nlr_79
Ca_NC_039903.1_Chr_3e_nlr_96	Cc_chr1_nlr_53	Ce_NC_040042.1_Chr_8_nlr_87
Ca_NC_039904.1_Chr_4c_nlr_13	Cc_chr1_nlr_54	Ce_NC_040042.1_Chr_8_nlr_89
Ca_NC_039904.1_Chr_4c_nlr_15	Cc_chr1_nlr_58	Ce_NC_040042.1_Chr_8_nlr_91
Ca_NC_039904.1_Chr_4c_nlr_26	Cc_chr1_nlr_6	Ce_NC_040042.1_Chr_8_nlr_95
Ca_NC_039904.1_Chr_4c_nlr_29	Cc_chr1_nlr_65	Ce_NC_040043.1_Chr_9_nlr_17
Ca_NC_039904.1_Chr_4c_nlr_5	Cc_chr1_nlr_68	Ce_NC_040043.1_Chr_9_nlr_25
Ca_NC_039905.1_Chr_4e_nlr_21	Cc_chr1_nlr_7	Ce_NC_040045.1_Chr_11_nlr_101
Ca_NC_039905.1_Chr_4e_nlr_24	Cc_chr11_nlr_13	Ce_NC_040045.1_Chr_11_nlr_121
Ca_NC_039905.1_Chr_4e_nlr_26	Cc_chr11_nlr_36	Ce_NC_040045.1_Chr_11_nlr_132

Ca_NC_039905.1_Chr_4e_nlr_28	Cc_chr11_nlr_41	Ce_NC_040045.1_Chr_11_nlr_2
Ca_NC_039905.1_Chr_4e_nlr_29	Cc_chr11_nlr_72	Ce_NC_040045.1_Chr_11_nlr_27
Ca_NC_039905.1_Chr_4e_nlr_30	Cc_chr11_nlr_78	Ce_NC_040045.1_Chr_11_nlr_31
Ca_NC_039905.1_Chr_4e_nlr_31	Cc_chr2_nlr_32	Ce_NC_040045.1_Chr_11_nlr_33
Ca_NC_039906.1_Chr_5e_nlr_12	Cc_chr2_nlr_33	Ce_NC_040045.1_Chr_11_nlr_36
Ca_NC_039906.1_Chr_5e_nlr_23	Cc_chr2_nlr_9	Ce_NC_040045.1_Chr_11_nlr_53
Ca_NC_039906.1_Chr_5e_nlr_35	Cc_chr3_nlr_100	Ce_NC_040045.1_Chr_11_nlr_55
Ca_NC_039906.1_Chr_5e_nlr_36	Cc_chr3_nlr_18	Ce_NC_040045.1_Chr_11_nlr_67
Ca_NC_039906.1_Chr_5e_nlr_44	Cc_chr3_nlr_24	Ce_NC_040045.1_Chr_11_nlr_96
Ca_NC_039906.1_Chr_5e_nlr_55	Cc_chr3_nlr_41	Ce_NW_020861822.1_nlr_2
Ca_NC_039906.1_Chr_5e_nlr_60	Cc_chr3_nlr_44	Ce_NW_020861823.1_nlr_2
Ca_NC_039906.1_Chr_5e_nlr_68	Cc_chr3_nlr_46	Ce_NW_020862264.1_nlr_1
Ca_NC_039906.1_Chr_5e_nlr_70	Cc_chr3_nlr_59	Ce_NW_020862380.1_nlr_1
Ca_NC_039906.1_Chr_5e_nlr_71	Cc_chr3_nlr_63	Ce_NW_020862810.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_11	Cc_chr3_nlr_65	Ce_NW_020862919.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_18	Cc_chr3_nlr_70	Ce_NW_020862967.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_31	Cc_chr3_nlr_85	Ce_NW_020863939.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_32	Cc_chr3_nlr_87	Ce_NW_020864008.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_38	Cc_chr4_nlr_16	Ce_NW_020864288.1_nlr_2
Ca_NC_039907.1_Chr_5c_nlr_39	Cc_chr4_nlr_17	Ce_NW_020864351.1_nlr_1
Ca_NC_039907.1_Chr_5c_nlr_5	Cc_chr4_nlr_20	Ce_NW_020864351.1_nlr_3
Ca_NC_039907.1_Chr_5c_nlr_57	Cc_chr4_nlr_22	Ce_NW_020864659.1_nlr_11
Ca_NC_039907.1_Chr_5c_nlr_7	Cc_chr4_nlr_26	Ce_NW_020864659.1_nlr_4
Ca_NC_039907.1_Chr_5c_nlr_71	Cc_chr4_nlr_5	Ce_NW_020864860.1_nlr_2
Ca_NC_039908.1_Chr_6c_nlr_14	Cc_chr5_nlr_10	-
Ca_NC_039908.1_Chr_6c_nlr_24	Cc_chr5_nlr_2	-
Ca_NC_039908.1_Chr_6c_nlr_29	Cc_chr5_nlr_22	-
Ca_NC_039908.1_Chr_6c_nlr_32	Cc_chr5_nlr_3	-
Ca_NC_039908.1_Chr_6c_nlr_33	Cc_chr5_nlr_43	-
Ca_NC_039908.1_Chr_6c_nlr_35	Cc_chr6_nlr_21	-
Ca_NC_039908.1_Chr_6c_nlr_38	Cc_chr6_nlr_24	-
Ca_NC_039909.1_Chr_6e_nlr_2	Cc_chr6_nlr_7	-
Ca_NC_039909.1_Chr_6e_nlr_3	Cc_chr7_nlr_11	-
Ca_NC_039910.1_Chr_7c_nlr_22	Cc_chr7_nlr_13	-
Ca_NC_039910.1_Chr_7c_nlr_25	Cc_chr7_nlr_20	-
Ca_NC_039910.1_Chr_7c_nlr_32	Cc_chr7_nlr_23	-
Ca_NC_039910.1_Chr_7c_nlr_33	Cc_chr7_nlr_38	-
Ca_NC_039910.1_Chr_7c_nlr_37	Cc_chr7_nlr_39	-
Ca_NC_039910.1_Chr_7c_nlr_38	Cc_chr8_nlr_1	-
Ca_NC_039910.1_Chr_7c_nlr_47	Cc_chr8_nlr_15	-
Ca_NC_039910.1_Chr_7c_nlr_48	Cc_chr8_nlr_16	-
Ca_NC_039910.1_Chr_7c_nlr_49	Cc_chr8_nlr_18	-
Ca_NC_039910.1_Chr_7c_nlr_55	Cc_chr8_nlr_19	-
Ca_NC_039910.1_Chr_7c_nlr_56	Cc_chr8_nlr_21	-
Ca_NC_039911.1_Chr_7e_nlr_10	Cc_chr8_nlr_31	-
Ca_NC_039911.1_Chr_7e_nlr_19	Cc_chr8_nlr_39	-
Ca_NC_039911.1_Chr_7e_nlr_23	Cc_chr8_nlr_40	-
Ca_NC_039911.1_Chr_7e_nlr_25	-	-

Ca_NC_039911.1_Chr_7e_nlr_28	-	-
Ca_NC_039911.1_Chr_7e_nlr_30	-	-
Ca_NC_039911.1_Chr_7e_nlr_33	-	-
Ca_NC_039911.1_Chr_7e_nlr_35	-	-
Ca_NC_039911.1_Chr_7e_nlr_38	-	-
a_NC_039911.1_Chr_7e_nlr_41	-	-
Ca_NC_039911.1_Chr_7e_nlr_56	-	-
Ca_NC_039911.1_Chr_7e_nlr_57	-	-
Ca_NC_039911.1_Chr_7e_nlr_7	-	-
Ca_NC_039912.1_Chr_8e_nlr_15	-	-
Ca_NC_039912.1_Chr_8e_nlr_18	-	-
Ca_NC_039912.1_Chr_8e_nlr_22	-	-
Ca_NC_039912.1_Chr_8e_nlr_33	-	-
Ca_NC_039912.1_Chr_8e_nlr_34	-	-
Ca_NC_039912.1_Chr_8e_nlr_42	-	-
Ca_NC_039912.1_Chr_8e_nlr_45	-	-
Ca_NC_039912.1_Chr_8e_nlr_48	-	-
Ca_NC_039912.1_Chr_8e_nlr_51	-	-
Ca_NC_039912.1_Chr_8e_nlr_60	-	-
Ca_NC_039912.1_Chr_8e_nlr_69	-	-
Ca_NC_039912.1_Chr_8e_nlr_80	-	-
Ca_NC_039912.1_Chr_8e_nlr_82	-	-
Ca_NC_039913.1_Chr_8c_nlr_1	-	-
Ca_NC_039913.1_Chr_8c_nlr_12	-	-
Ca_NC_039913.1_Chr_8c_nlr_14	-	-
Ca_NC_039913.1_Chr_8c_nlr_37	-	-
Ca_NC_039913.1_Chr_8c_nlr_38	-	-
Ca_NC_039913.1_Chr_8c_nlr_39	-	-
Ca_NC_039913.1_Chr_8c_nlr_4	-	-
Ca_NC_039913.1_Chr_8c_nlr_48	-	-
Ca_NC_039913.1_Chr_8c_nlr_53	-	-
Ca_NC_039913.1_Chr_8c_nlr_55	-	-
Ca_NC_039913.1_Chr_8c_nlr_6	-	-
Ca_NC_039913.1_Chr_8c_nlr_7	-	-
Ca_NC_039914.1_Chr_9c_nlr_15	-	-
Ca_NC_039914.1_Chr_9c_nlr_3	-	-
Ca_NC_039915.1_Chr_9e_nlr_1	-	-
Ca_NC_039917.1_Chr_10c_nlr_12	-	-
Ca_NC_039917.1_Chr_10c_nlr_20	-	-
Ca_NC_039917.1_Chr_10c_nlr_22	-	-
Ca_NC_039917.1_Chr_10c_nlr_24	-	-
Ca_NC_039917.1_Chr_10c_nlr_6	-	-
Ca_NC_039918.1_Chr_11c_nlr_11	-	-
Ca_NC_039918.1_Chr_11c_nlr_25	-	-
Ca_NC_039918.1_Chr_11c_nlr_27	-	-
Ca_NC_039918.1_Chr_11c_nlr_28	-	-
Ca_NC_039918.1_Chr_11c_nlr_54	-	-
Ca_NC_039918.1_Chr_11c_nlr_62	-	-

Ca_NC_039919.1_Chr_11e_nlr_102	-	-
Ca_NC_039919.1_Chr_11e_nlr_104	-	-
Ca_NC_039919.1_Chr_11e_nlr_109	-	-
Ca_NC_039919.1_Chr_11e_nlr_111	-	-
Ca_NC_039919.1_Chr_11e_nlr_112	-	-
Ca_NC_039919.1_Chr_11e_nlr_20	-	-
Ca_NC_039919.1_Chr_11e_nlr_23	-	-
Ca_NC_039919.1_Chr_11e_nlr_38	-	-
Ca_NC_039919.1_Chr_11e_nlr_42	-	-
Ca_NC_039919.1_Chr_11e_nlr_49	-	-
Ca_NC_039919.1_Chr_11e_nlr_56	-	-
Ca_NC_039919.1_Chr_11e_nlr_66	-	-
Ca_NC_039919.1_Chr_11e_nlr_7	-	-
Ca_NC_039919.1_Chr_11e_nlr_75	-	-
Ca_NC_039919.1_Chr_11e_nlr_77	-	-
Ca_NC_039919.1_Chr_11e_nlr_85	-	-
Ca_NC_039919.1_Chr_11e_nlr_90	-	-
Ca_NC_039919.1_Chr_11e_nlr_92	-	-
Ca_NC_039919.1_Chr_11e_nlr_93	-	-
Ca_NW_020848476.1_nlr_1	-	-
Ca_NW_020850474.1_nlr_1	-	-
Ca_NW_020850474.1_nlr_2	-	-
Ca_NW_020850474.1_nlr_4	-	-
Ca_NW_020850474.1_nlr_7	-	-
Ca_NW_020850474.1_nlr_8	-	-
Ca_NW_020850885.1_nlr_2	-	-
Ca_NW_020851092.1_nlr_1	-	-
Ca_NW_020851248.1_nlr_1	-	-
Ca_NW_020851248.1_nlr_3	-	-