JUAN CARLOS FLÓREZ VARON

TRANSCRIPTOMA DO CAFEEIRO (*Coffea arabica* L.) DURANTE A INTERAÇÃO COM *Hemileia vastatrix* Berk. & Br

Tese apresentada à Universidade Federal de Viçosa, como parte do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
Minas GERAIS – BRASIL
2017

Ficha catalográfica preparada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

F634t
2017

Florez Varon, Juan Carlos, 1980-
Transcriptoma do cafeeiro (Coffea arabica L.) durante a interação com Hemileia vastatrix Berk. & Br / Juan Carlos Florez Varon. - Vicosa, MG, 2017.
vii, 56f. : il. (algumas color.) ; 29 cm.

Inclui anexos.
Inclui apêndices.
Orientador: Eveline Teixeira Caixeta.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Cafeeiro. 2. Hemileia vastatrix. 3. Genética vegetal. I. Universidade Federal de Viçosa. Genetica e Melhoramento. Programa de Pós-graduação em Genética e Melhoramento. II. Título.

CDD 22 ed. 633.73

JUAN CARLOS FLÓREZ VARON

TRANSCRIPTOMA DO CAFEEIRO (*Coffea arabica* L.) DURANTE A INTERAÇÃO COM *Hemileia vastatrix* Berk. & Br

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*

APROVADA: 16 de fevereiro de 2017.

_____	_____
Sávio de Siqueira Ferreira	Marcelo Falsarella Carazole
_____	_____
Renato Domiciano Silva Rosado	Laércio Zambolim (Coorientador)

_____

Eveline Teixeira Caixeta
(Orientadora)

AGRADECIMENTOS

A Deus, por ter abençoado todos os dias da minha vida, por iluminar meu caminho e me dar forças para seguir sempre em frente e cumprir meus sonhos.

A toda minha família (pais, irmãos, cunhados e sobrinhos) pois deles sempre recebi apoio, torcida e confiança; pelos momentos que não estivemos juntos e souberam entender.

À Universidade Federal de Viçosa e o Programa de Pós-Graduação em Genética e Melhoramento pela oportunidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos concedida durante a realização deste doutorado.

A minha orientadora Eveline, pela ajuda, orientação e paciência. Porque com cada um de seus conselhos me mostro sempre o melhor caminho para crescer como um profissional íntegro.

Ao professor Dr. Laércio Zambolim, pela oportunidade de fazer parte da equipe do BioCafé.

A Moema Souza e todos meus colombianos aqui em Viçosa, porque de alguma forma se tornaram a minha família.

A todos meus colegas do laboratório BioCafé (Dani, Rejane, Júlia, Francielle, Emilly, Rafaela, Letícia, Tiago, Geleta, Sávio, Pedro, Laura, Alexandra, Dênia, Samila, Rose e Sr. Geraldo) pela ajuda e pelo carinho que me brindaram durante estes quatro anos.

Ao professor da fitopatológica Jorge Badel pelo tempo dedicado para trocar ideias e enriquecer meu trabalho.

A Marcelo, Luciana, Juliana, Sheila, Gustavo, Pablo e Cid do Laboratório de Genômica e expressão da Unicamp, pela amizade e ajuda.

Ao Brasil, um pais que me acolheu e abriu as portas para que eu pudesse crescer como profissional e como pessoa.

SUMÁRIO

RESUMO

Flórez, VARON, Juan Carlos, D.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Transcriptoma do cafeeiro (*Coffea arabica* L.) durante a interação com *Hemileia vastatrix* Berk. & Br.** Orientadora: Eveline Teixeira Caixeta. Coorientadores: Moysés Nascimento e Laércio Zambolim.

O cafeeiro é uma das culturas mais importantes no mundo, porém, é atacada por diversas doenças, sendo a ferrugem, causada pelo fungo *Hemileia vastatrix,* considerada a principal doença em todas as regiões produtivas. O estudo do transcriptoma do cafeeiro durante a interação com *H. vastatrix* pode auxiliar no controle dessa doença pelo entendimento de quais mecanismos de defesa a planta esta ativando. Em um patossistema, entender a expressão gênica é essencial para a identificação dos genes relacionados com os mecanismos de defesa e resistência da planta, que são ativados pelos genes de patogenicidade do agente causal. Esses genes identificados podem ser utilizados para obtenção de cultivares resistentes. Para estudar o transcriptoma de um organismo, o método que tem sido utilizado é o sequenciamento de mRNA, denominado RNA-Seq. Esta abordagem permite monitorar a expressão de genes da planta e do patógeno em diferentes etapas ao longo do processo infeccioso. Desta forma, o objetivo deste trabalho foi estudar o transcriptoma do cafeeiro durante a interação com *H. vastatrix*, fungo causador da ferrugem, a fim de identificar os genes que são ativados ou reprimidos em resposta à infecção. Folhas dos cafeeiros *C. arabica* cv. Caturra CIFC 19/1 (suscetível) e Híbrido de Timor CIFC 832/1 (resistente) foram inoculadas com urediósporos da raça XXXIII de *H. vastatrix* e coletadas em diferentes tempos após a inoculação (12, 24, 96 horas e 17 dias). Como controle, foram utilizadas folhas dos dois genótipos de café inoculados com água. Após a extração do RNA das amostras, foram obtidas as bibliotecas de cDNA, que foram sequenciadas usando a plataforma Illumina Miseq. As análises dos dados foram realizadas utilizando ferramentas de bioinformática e consistiram em: i) avaliação da qualidade das sequências com o programa *FastQC*; ii) limpeza dos dados e sobreposição de *reads* de acordo com os critérios de qualidade, usando *Pear* e *Clean Solexa*; iii) mapeamento dos transcritos com o genoma de referência de *Coffea canephora* e montagem *de novo* como

estratégia complementar; iv) analise de expressão diferencial de genes; v) anotação; e vi) validação de genes candidatos por PCR em tempo real. Um total de 43.159 *contigs* foram obtidos após o mapeamento contra *C. canephora* e a montagem *de novo*. Os resultados sugerem que durante a infecção inicial (12 e 24 hai), o Híbrido de Timor (HdT) foi mais responsivo ao ataque de *H. vastatrix* em relação ao Caturra por apresentar maior número de genes *up-regulated*. Foram selecionados treze genes (*up-regulated*) exclusivos do HdT, para avaliar o padrão de expressão com o uso de qPCR. As estratégias utilizadas para a montagem do transcriptoma foram eficientes e permitiram a obtenção de um banco de dados com qualidade, o qual poderá ser utilizado para mineração de genes expressos no patossistema *C. arabica-H. vastatrix* durante a infecção.

ABSTRACT

Flórez, VARON, Juan Carlos, D.Sc., Universidade Federal de Viçosa, February, 2017. **Transcriptome profile of coffee (C*offea arabica* L.) during an interaction with *Hemileia vastatrix* Berk. & Br.** Adviser: Eveline Teixeira Caixeta. Co-advisers: Moysés Nascimento and Laércio Zambolim.

Coffee is one of the most important crops in the world; however, several diseases, among which coffee rust, caused by the fungus *Hemileia vastatrix*, is considered the main disease in all producing regions. The study of coffee transcriptome during the interaction with *H. vastatrix* can help in the control of this disease as it reveals the mechanism of activation of genes involved in defense. In a pathosystem, understanding gene expression is essential for the identification of genes related to plant defense and resistance mechanisms, which are activated by the presence of the pathogen as well as the pathogenicity genes of the microorganism. To study the transcriptome of an organism, the method that has been used is the sequencing of next generation sequencing of mRNA, called RNA-Seq. This approach allows monitoring the gene expression of an organism at different stages throughout the infection process. Thus, the objective of this work was to study the coffee transcriptome during interaction with *Hemileia vastatrix*, rust-causing fungus, in order to identify genes that are activated or repressed in response to infection. Leaves of two Arabica coffee cultivars, Caturra CIFC 19/1 (susceptible) and Híbrido de Timor CIFC 832/1 (resistant), were inoculated with *H. vastatrix* race XXXIII and collected at different times after inoculation (12, 24, 96 hours and 17 days). As a control, leaves of the two coffee cultivars were inoculated with water. After extraction of the RNA from the samples, cDNA libraries were constructed and sequenced using the Illumina Miseq platform. Data analyzes were performed using bioinformatics tools and consisted of: i) evaluation of sequence quality with FastQC program; ii) data cleaning and overlapping of reads according to the quality criteria, using Pear and Clean Solexa; iii) mapping of transcripts with the reference genome of *Coffea canephora* and *de novo* assembly as a complementary strategy; iv) analysis of differential gene expression; v) annotation; and vi) validation of candidate genes by real-time PCR. A total of 43,159 contigs were obtained

after mapping against *C. canephora* and *de novo* assembly. The results suggest that during the initial infection (12 and 24 hai), Híbrido de Timor (HdT) was more responsive to *H. vastatrix* attack than Caturra because it had a higher number of up-regulated genes. Thirteen (up-regulated) genes unique to HdT were selected to evaluate the expression pattern by qPCR. The strategies used to assemble the transcriptome were efficient and allowed to obtain a high quality database, which will be used for mining of genes expressed in the *C. arabica-H. vastatrix* interaction.

1 INTRODUÇÃO GERAL

O cafeeiro pertence à família Rubiaceae, gênero Coffea, no qual se encontram descritas mais de 100 espécies. A produção comercial de café atualmente se baseia nas espécies *Coffea arabica* L. (60%) e *Coffea canephora* Pierre e Frohner (40%) (Anthony *et al*., 2002). Porém, a produtividade dessas duas espécies podem ser afetadas por diversas doenças. Dentre as doenças que ocorrem no cafeeiro, a ferrugem é a mais importante por causar grandes prejuízos (Chalfoun, 1997). A ferrugem do cafeeiro tem como agente causal o fungo *Hemileia vastatrix* Berk. et Br., pertencente à ordem Uredinales e família Pucciniaceae, um parasita biotrófico exclusivo do gênero *Coffea* (Silva *et al*., 2006).

A espécie *C. canephora* Pierre e Frohner apresenta variedades com resistência à doença, enquanto que a maioria das variedades comerciais da espécie *C. arabica* L. são suscetíveis à doença (Carvalho *et al*., 2010). O uso eficiente dos recursos genéticos disponíveis nas espécies selvagens diploides através de híbridos interespecíficos é essencial para o melhoramento contínuo das variedades de café arábica (Vossen, 2001; Lashermes *et al*., 2009). A incorporação de novas metodologias de introgressão de genes de outras espécies em *C. arabica* são particularmente valiosos para os programas de melhoramento (Marraccini *et al*., 2011). Nesse sentido, o estudo do transcriptoma do cafeeiro durante a interação com *H. vastatrix* permite identificar genes expressos na planta com potenciais para serem introduzidos em programas de melhoramento do cafeeiro.

1. 1 Estudo do transcriptoma na interação planta-patógeno

Estudar o perfil do transcriptoma é essencial para compreender plenamente as vias biológicas que são ativadas em varias condições fisiológicas ou estádio de desenvolvimento de um organismo (Wang *et al*., 2009; Ozsolak and Milos, 2010). Uma destas condições pode ser a interação planta-patógeno, a qual inicia com uma cascata de eventos dinâmicos que culmina com padrões de expressão de genes alterados na interação dos

dois organismos. O perfil do transcriptoma do hospedeiro e do patógeno pode fornecer um novo entendimento sobre este processo, além de permitir a identificação de novos genes de virulência no patógeno, ou novas vias na célula hospedeira que respondem à exposição a patógenos (Westermann *et al.*, 2012; Boyd *et al.*, 2013). A montagem e anotação do transcriptoma permite ainda ampliar o conhecimento dos elementos funcionais do genoma e especificidades moleculares de células e tecidos (Wang *et al.*, 2009; Martin and Wang, 2011).

Dentro das técnicas utilizadas para estudar o transcriptoma destacam-se, o sequenciamento de ESTs (*Expressed Sequence Tags*), o SAGE (*Serial analysis of gene expression*), *Differential display* e o SSH (*Supression subtrative hybridization*). Há ainda os Microarranjos e *Nothern blot* que durante muito tempo foram utilizados para a determinação dos níveis de transcrição em larga escala (Goldsmith and Dhanasekaran, 2004). Essas duas metodologias apesar de permitir análise de muitos RNAs em um mesmo ensaio, necessitam do conhecimento prévio do genoma (Ansorge, 2009; Martinez and Nelson, 2010). Mais recentemente, com o advento e avanço das tecnologias de sequenciamento de nova geração, surgiu a metodologia conhecida como RNA *deep sequencing* ou RNA-Seq (Wang *et al.*, 2009). Nessa metodologia, o transcriptoma é obtido por meio do sequenciamento em larga escala de bibliotecas de cDNA, o que permite estudar o perfil do transcriptoma com maior profundidade e precisão.

Diferentes estudos de transcriptoma que têm sido realizados a partir do RNA-Seq na interação planta-patógeno. Yazawa *et al.* (2013) analisaram simultaneamente o transcriptoma de *Sorghum bicolor* (L.) Moench na interação com o patógeno *Bipolaris sorghicola*. Neste experimento foi sequenciada a cultivar de sorgo SIL-05 (resistente) inoculada com o fungo. Os genes encontrados no fungo codificaram proteínas intracelulares das hifas, fatores de transcrição *CpcA* e *HacA*, além de enzimas que degradam a parede celular vegetal. Os genes que foram encontrados na planta codificaram dois receptores da família de proteínas de domínio *Leucine-rich repeat* (LRR), fatores de transcrição (*OsWRKY45* e *OsWRKY28*) e uma peroxidase da classe III envolvida na resistência à Mancha alvo.

Outro trabalho desenvolvido que inclui o uso do RNA-Seq foi feito em

cacau e *Moniliophthora perniciosa* durante a sua interação biotrófica. Um total de 34,997 genes foram encontrados na planta, deste um total 1269 genes foram considerados *up-regulated* e 698 *down-regulated*. A infecção de *M. perniciosa* desencadeou uma massiva reprogramação metabólica nos tecidos infectados, ativando a expressão de genes relacionados com a modificação da parede celular, metabolismo secundário e degradação de lipídeos (Teixeira *et al.*, 2014). Que *et al.* (2014) inocularam *Sporisorium scitamineum* em plantas resistentes e suscetíveis de cana de açúcar em diferentes tempos (24, 48 e 120 hrs) e sequenciaram as bibliotecas usando plataforma da Illumina e analisaram o transcriptoma da planta e do patógeno. Os resultados mostraram os genes encontrados estão associados a mecanismos de transdução de sinal, produção de energia e de conversão, transporte de íon inorgânico e metabolismo, e mecanismos de defesa.

1. 2 Sequenciamento de RNA (RNA-Seq)

A metodologia de RNA-Seq possibilita monitorar com precisão a expressão do conjunto genes da planta e do patógeno quando eles estão em interação. Além disso, a técnica permite sequenciar o transcriptoma de um pequeno número de células no local inicial da infecção (Kim *et al.*, 2011; Westermann *et al.*, 2012; Chen *et al.*, 2015). A vantagem de utilizar o RNA-Seq é a possiblidade de identificar transcritos sem a necessidade do conhecimento prévio do genoma do patógeno. Além da descoberta de novos transcritos, essa tecnologia pode ser empregada para medir a abundância de transcritos, identificar regiões não traduzidas, variantes de *splicing* e identificar polimorfismos de uma única base (SNPs) (Nowrousian *et al.*, 2010; Nowrousian, 2010; Han *et al.*, 2014; Chang *et al.*, 2014). Outras vantagens incluem minimizar os erros causados por ambiguidades dos *reads* mapeados com o genoma de referência e maior sensibilidade para as regiões de baixa e alta expressão gênica (Nagalakshmi *et al.*, 2010).

A técnica do RNA-Seq consiste em converter o RNA em uma biblioteca de fragmentos de cDNA com adaptadores ligados em ambas extremidades. Cada fragmento é então sequenciado para se obter sequências curtas (*reads*) de uma extremidade (*single-end*) ou de ambas

extremidades (*pair-end*). O tamanho dos *reads* gerados pelas plataformas de sequenciamento tipicamente está entre 30 a 400 pares de bases (Oshlack *et al.*, 2010). Os dados brutos gerados consistem em uma longa lista de sequências curtas com classificações de qualidade associadas (Oshlack *et al.*, 2010). No início, as tecnologias de sequenciamento para o RNA-Seq mais utilizadas eram *Illumina* e *Roche 454 system* (Wang *et al.*, 2009; Auer and Doerge, 2010). Atualmente, a plataforma *Roche* não é mais comercializada, portanto, a *Illumina* tem sido a mais utilizada. A geração de dados de RNA-Seq é um processo em constante evolução, que inclui o desenvolvimento da tecnologia de sequenciamento e o desenvolvimento de algoritmos para a análise dos dados (Han *et al.*, 2015).

1. 2. 1 Análise de dados

Após o sequenciamento, os milhões de *reads* gerados são submetidos a análises de bioinformática para a identificação e quantificação do transcriptoma. Resumidamente, as etapas da análise consistem em remover sequencias contaminantes (RNA ribossomal, adaptadores, etc), montagem do transcriptoma usando um genoma de referência ou *de novo*, seguido pelo mapeamento dos *reads*. Depois os *reads* mapeados para cada amostra são subsequentemente indexados a nível de gene ou transcrito para avaliar a sua abundância. Posteriormente, estes dados são avaliados por modelos estatísticos para identificar que genes são diferencialmente expressos. Finalmente, são realizados analises de anotação para categorizar funcionalmente o conjunto de genes identificados durante o processo de montagem (Han *et al.*, 2015).

O passo inicial é a avaliação por qualidade dos milhões de *reads* que são gerados pelo sequenciamento. De forma geral, o valor da qualidade é a probabilidade de erro na identificação de uma determinada base (Auer and Doerge, 2010). O processo de sequenciamento pode conter erros originados pela presença de contaminantes que afeta a qualidade dos resultados ou simplesmente erro de sequenciamento. Assim, nesta fase são removidos fragmentos de baixa qualidade e sequências como *primers*, vetores, adaptadores e sequências longas de bases repetidas que não são de

interesse, ou que de alguma forma possam afetar as análises das próximas fases (Han *et al.*, 2015). Diferentes ferramentas estão disponíveis para esta finalidade, e permitem visualizar graficamente a qualidade dos *reads.* Uma destas ferramentas é o *software* FastQC (http://www.bioinformatics.bbsrc.ac.uk/projects/ fastqc) (Han *et al.*, 2015).

A próxima etapa inclui mapear o conjunto de *reads* contra um genoma de referência ou utilizar a estratégia conhecida como montagem *de novo*, caso não exista um organismo próximo completamente sequenciado (Wang *et al.*, 2009; Martin and Wang, 2011; Grabherr *et al.*, 2011; Lu *et al.*, 2013), Figura 1. A ideia básica da primeira estratégia é alinhar o conjunto de *reads* a um genoma de referência. Quando existe uma referência de boa qualidade, esta estratégia é altamente sensível, e tem se convertido em um método básico usado em muitos estudos de sequenciamento de RNA (RNA-Seq). No entanto, a precisão da montagem do transcriptoma baseado em um genoma de referência depende de um correto alinhamento dos *reads* (Grabherr *et al.*, 2011). Diversos são os *softwares* que podem ser utilizados, entre eles, *Blat* (Kent, 2002), *Tophat* (Trapnell *et al.*, 2009), *SpliceMap* (Au *et al.*, 2010), *MapSplice* (Wang *et al.*, 2010) e *Bowtie* (Langmead *et al.*, 2009). Para a montagem do transcriptoma com genoma de referência é preciso buscar a sobreposição do conjunto de *reads* alinhados com a construção de grafos até obter os transcritos (Chang *et al.*, 2014).

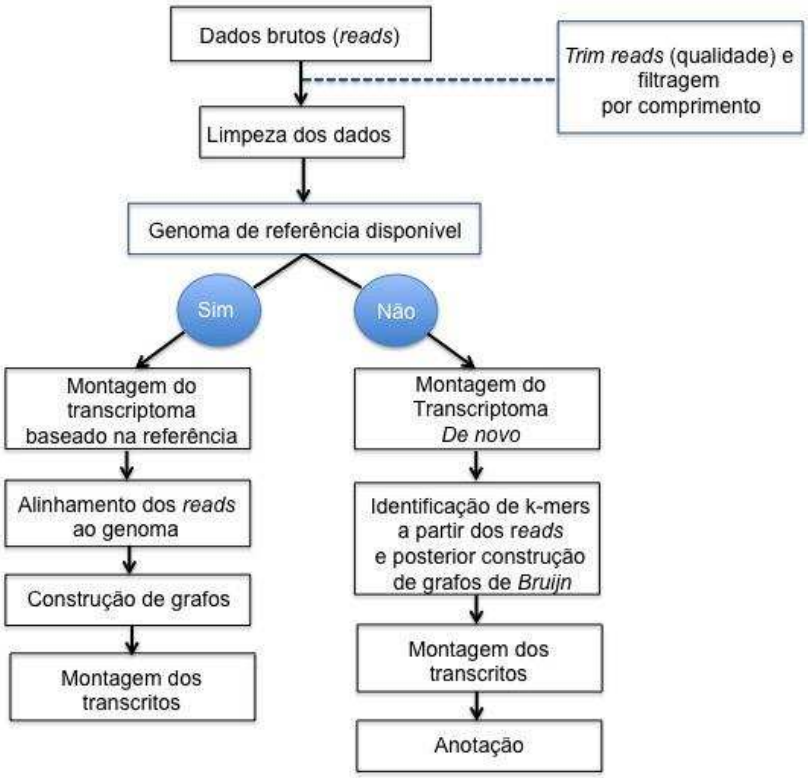


Figura 1. *Pipeline* para a montagem do transcriptoma (Martin and Wang, 2011)**.**

Na montagem *de novo* também vários *softwares* tem sido desenvolvidos, dentre eles *Oases* (Schulz *et al.*, 2012), *Trans-ABySS* (Robertson *et al.*, 2010), *SOAPdenovo-Trans* (Xie *et al.*, 2014) e *Trinity* (Grabherr *et al.*, 2011). Os *softwares* utilizados na montagem *de novo* utilizam geralmente uma estratégia que envolve a construção de grafos de *Bruijn* (Grabherr *et al.*, 2011; Schulz *et al.*, 2012). Um grafo de *Bruijn* é criado a partir de k-mers (comprimento de subsequências encontradas nos *reads*). Uma subcadeia de k-mers é deslocada por uma base de sequência, e ela se sobrepõe a outro k-mer. Finalmente uma cadeia linear k-mers é construída para obter os transcritos (Martin and Wang, 2011), Figura 2.

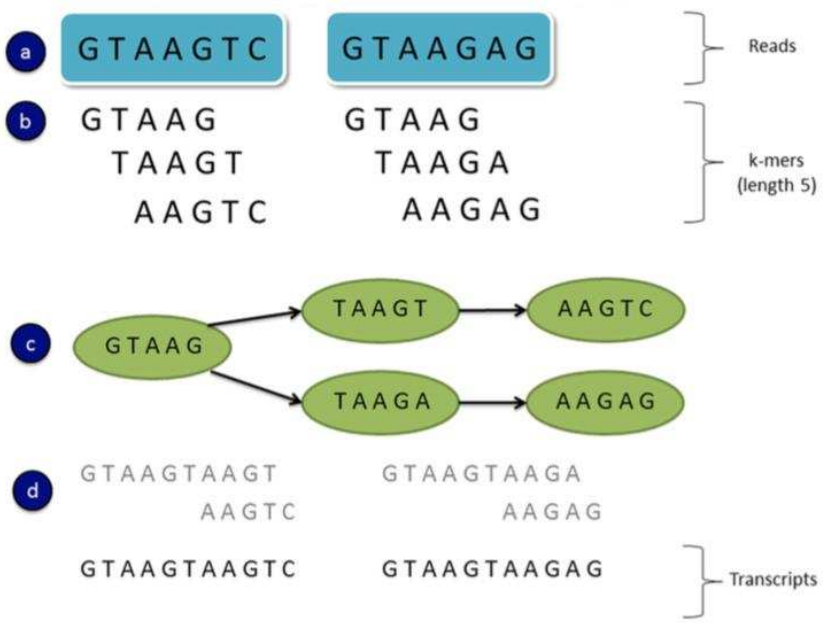


Figura 2. Grafo de Bruijn na montagem *de novo*. **a** Sequência dos *reads***, b** subsequências de *reads* (k-mers), **c** Grafos de Bruijn construídos a partir dos k-mers**, d** Transcritos montados (Moreton *et al.*, 2016).

Uma vez os *reads* estando alinhados, a seguinte etapa é a normalização, o que é crucial na análise de dados de RNA-Seq e tem um grande impacto sobre a detecção de genes diferencialmente expressos. Nos últimos anos, várias estratégias de normalização têm sido propostas para corrigir diferenças entre as amostras na contagem dos *reads*, tais como a profundidade de sequenciamento e comprimento do gene (Bullard *et al.*, 2010). *Reads per kilobase per million mapped reads* (RPKM) é o método mais utilizado para normalização de dados provenientes de plataformas de nova geração. Esta metodologia foi inicialmente proposta para facilitar a comparação entre genes dentro de uma amostra, já que tem a capacidade de redimensionar a contagem de genes para corrigir as diferenças de tamanho das bibliotecas e o comprimento do gene (Mortazavi *et al.*, 2008).

Uma vez que os *reads* são mapeados e os transcritos são identificados e normalizados, existem diferentes ferramentas que podem ser utilizadas para quantificar a expressão diferencial genes tais como *Cufflinks* (Trapnell *et al.*, 2010), *DESeq2* (Love *et al.*, 2014) ou *EdgeR* (Robinson *et al.*, 2010). Na etapa de analise de expressão diferencial de genes é possível identificar mudanças no nível de expressão de um conjunto de genes em

diferentes condições. No caso mais simples, o objetivo é comparar os níveis de expressão entre duas condições, por exemplo, selvagem e mutante. Desenhos experimentais mais complexos podem incluir fatores adicionais, potencialmente com vários níveis (por exemplo, vários mutantes). Um componente crucial dessa análise é o método estatístico utilizado para identificar os genes diferencialmente expressos (Anders *et al.*, 2012).

Finalmente, a etapa de anotação é o processo de procurar informação relevante dos transcritos obtidos na fase do mapeamento, devidamente interpretadas, para extrair seu significado biológico visando a compreensão dos processos biológicos (Stein, 2001). A maioria das estratégias para a anotação de transcritos usa uma ou mais abordagens baseadas em homologia para identificar sequências relacionadas de função conhecida, e portanto, transferem essa anotação para o novo transcrito (Emes, 2008). Algumas ferramentas utilizadas para desenvolver este processo são Pfam (Finn *et al.*, 2014), Interpro (Mitchell *et al.*, 2015) e Gene Ontology (Anon, 2015).

No contexto da interação planta-patógeno, a partir da análise por bioinformática dos dados provenientes de plataformas de sequenciamento, é possível identificar mecanismos de defesa que as plantas ativam quando são atacados por patógenos, principalmente genes *R* e outras proteínas que estão envolvidas no reconhecimento do patógeno. Essas proteínas desencadeiam cascatas de sinalização para ativar uma resposta de defesa específica na planta. Esse sistema de defesa da planta tem sido estudado em diferentes patossistemas modelo, e os conhecimentos gerados são utilizados para estudos de genômica dos organismos de interesse agronômico.

1. 3 Mecanismos de defesa das plantas contra fungos fitopatogênicos

Os mecanismos de defesa de uma planta podem ser morfológicos, estruturais e bioquímicos, ambos pré e/ou pós-formados em relação à tentativa de penetração do patógeno no hospedeiro. Os mecanismos estruturais constituem-se em barreiras físicas à penetração e/ou colonização do patógeno. Enquanto que os mecanismos bioquímicos englobam

substâncias capazes de inibir o desenvolvimento do patógeno ou gerar condições adversas para a sobrevivência nos tecidos do hospedeiro. Tais substâncias devem estar presentes em concentração adequada nas partes invadidas e em forma acessível ao patógeno, de tal maneira que mudanças na concentração da(s) substância(s) impliquem mudanças na expressão da doença (Schwan-Estrada *et al*., 2008).

Nos estágios iniciais da infecção são acionados padrões moleculares associados a patógenos em geral (*pathogen-associated molecular patterns*-PAMPs), que desencadeiam respostas de defesa na planta conhecidos como imunidade desencadeada por PAMPs (PTI - *PAMP-triggered immunity*). Se o agente patogênico suprime os diferentes componentes da PTI, uma segunda linha de defesa da planta é acionada. Esse mecanismo de defesa envolve o reconhecimento de proteínas efetoras específicas do patógeno, agora referidas como proteínas de avirulência (*AVR*), por genes de resistência (*R*) da planta. O resultado dessa interação é uma resposta de resistência conhecida como imunidade ativada por efetores (ETI) (Zhang and Zhou, 2010; Boyd *et al.*, 2013). Esse tipo de imunidade é patógeno-específica e desencadeia resposta de hipersensibilidade (HR) e resistência sistêmica adquirida (SAR) (Fu and Dong, 2013). Na interação planta-patógeno, a ativação de HR e outros mecanismos de defesa das plantas são desencadeados principalmente por meio da transcrição de numerosos genes, que codificam proteínas que exibem amplas atividades biológicas. Estas mudanças transcricionais podem ser monitoradas nas primeiras horas após a infecção do patógeno (Ganesh *et al.*, 2006).

Dentre os vários tipos de classificações usados para diferenciar as respostas de defesa, um deles classifica a resistência de plantas a patógenos em duas categorias: resistência qualitativa e resistência quantitativa (Dmitriev, 2003; Kou and Wang, 2010). A resistência qualitativa é mediada por um ou poucos genes de resistência (*R*), que permite o reconhecimento de raças distintas de patógenos e desencadeiam reações de defesa no sitio de infecção (Eulgem, 2005). A resistência quantitativa ou poligênica permite à planta utilizar-se da combinação de mecanismos de defesa pré-existentes ou induzidos para defender-se parcialmente. Por isso, quando as condições para infecção e o desenvolvimento de uma doença são

favoráveis, o genótipo da planta e o do patógeno serão determinantes para ocorrência (em estádio moderado ou grave) ou não da doença (Kushalappa *et al*., 2016).

1. 3. 1 Mecanismos de defesa de *Coffea arabica* na interação com *Hemileia vastatrix*

Estudos citológicos e bioquímicos têm demostrado que o cafeeiro ativa diversos mecanismos de defesa quando é atacado por *H. vastatrix*. Um desses mecanismo é a resposta de hipersensibilidade (HR), que está associada com a deposição de calose, compostos fenólicos (ácido clorogênico e flavonoides) e lignificação da parede celular (Silva *et al.*, 2002). Além de produzir enzimas como lipoxigenase e peroxidase que estimulam a rota dos fenilpropanoides e proteínas PR como β-1,3-glucanases e quitinases, que também estão associadas com a resistência de alguns cultivares de cafeeiro (Silva *et al.*, 2002; Fernandez *et al.*, 2004).

Em outro estudo foram inoculados esporos de *H. vastatrix* raça XXXIII em plantas de *C. arabica* cv. Caturra Vermelho (CIFC 19/1) e Híbrido de Timor (CIFC 832/1), para estabelecer interação compatível e incompatível. O processo de colonização do fungo e as respostas de defesa da planta foram avaliados em diferentes tempos após a inoculação: 10, 17, 24, 48, 72 e 96 horas; 17 e 21 dias. As observações citológicas indicaram que a resistência do Híbrido de Timor CIFC 832/1 à raça XXXIII de *H. vastatrix* é pré-haustorial, ao contrário da resistência pós-haustorial que é geralmente descrita para interações cafeeiro-*H. vastatrix* (Freitas *et al*., 2014).

A nível molecular, as técnicas de SSH (*suppression subtractive hybridization*), pirosequenciamento 454 e qRT-PCR têm permitido a identificação de vários genes putativamente envolvidos com a resistência no hospedeiro. Mais de um quarto das proteínas preditas encontradas utilizando essas técnicas em estudos de interação cafeeiro-*H. vastatrix* correspondiam a proteínas de resistência, proteínas de estresse e defesa, e componentes de vias de transdução de sinal (por exemplo, quitinases, beta-1,3-glucanases, PR10, lipoxigenase, tipo-AP2, e fatores de transcrição WRKY). A atividade das enzimas oxidativas (lipoxigenase, peroxidase e superóxido

dismutase), fenilalanina amónia-liase, quitinases e glucanases foram detectadas na reação de resistência (Fernandez *et al.*, 2004, 2012; Ganesh *et al.*, 2006; Diniz *et al.*, 2012).

Usando inoculações com diferentes raças de *H. vastatrix* foram realizados estudos de herança e resistência em algumas variedades de *C. arabica*, onde foram identificados pelo menos nove genes maiores dominantes (S_H1-S_H9) (Rodrigues *et al.*, 1975; Bettencourt and Rodrigues, 1988). Dos nove genes conhecidos, quatro foram identificados em *C. arabica*, S_H1, S_H2, S_H4 e S_H5; a maioria das cultivares arábicas comerciais possuem o gene S_H5. No entanto, a resistência dessas cultivares foi suplantada pelo fator de virulência *v5* que está presente na raça II de *H. vastatrix*, e em pelo menos outras trinta raças deste patógeno, onde o gene *v5* se encontra associado a outros genes de virulência (Rodriguez *et al*., 1993; Várzea and Marques, 2005).

Até o ano 2005, 50 raças de *H. vastatrix* tinham sido descritas no mundo (Várzea and Marques, 2005). No Brasil, foram identificadas 15 raças: I, II, III, VII, X, XIII, XV, XVI, XVII, XXI, XXII, XXIII, XXIV, XXV ou XXXI, e XXXVII (Cabral *et al*., 2009; Zambolim, 2016). As raças de *H. vastatrix* mais recentemente descobertas têm sido capazes de infectar derivados de Híbrido de Timor (HdT), a principal fonte de resistência utilizada nos programas de melhoramento genético (Gichuru *et al.*, 2012). Por exemplo, a cultivar Oeiras MG 6851, originada do cruzamento entre *C. arabica* cv. Caturra (CIFC 19/1) e HdT (CIFC 832/1), foi lançada como resistente à ferrugem (Pereira *et al*., 2000). No entanto, 12 anos após sua liberação, a resistência foi suplantada pela raça XXXIII de *H. vastatrix* (Capucho *et al.*, 2012).

A obtenção de cultivares de café com resistência vertical e completa à ferrugem (*H. vastatrix*) tem sido um constante desafio para os melhoristas devido ao surgimento contínuo de raças fisiológicas do patógeno, as quais tem suplantado a resistência de algumas dessas cultivares. A transcriptômica poderá auxiliar no entendimento do mecanismo genético da resistência a esta doença, onde estão envolvidos inúmeros genes que são ativados e/ou reprimidos de forma interativa na planta e no patógeno.

2. OBJETIVOS

2.1 Objetivo Geral

Estudar o transcriptoma do cafeeiro durante a interação com *Hemileia vastatrix*, a fim de identificar genes que são ativados ou reprimidos em resposta à infecção.

2.1. Objetivos específicos

- Obter o transcriptoma do cafeeiro na interação com *H. vastatrix,* considerando diferentes tempos após a infecção;
- Avaliar a expressão diferencial dos genes candidatos identificados nos diferentes tempos de infecção de *H. vastatrix* em plantas de Caturra (interação compatível) e de Híbrido de Timor (interação incompatível);
- Validar a expressão de genes candidatos de resistência por RT-qPCR (*Real Time Quantitative PCR*) em resposta à infecção.

3 REFERÊNCIAS BIBLIOGRÁFICAS

Anders, S. and Huber, W. (2012) Differential expression of RNA-Seq data at the gene level–the DESeq package. *EMBL, Heidelberg, Ger.* **11**, R106.

Anon (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **43**, D1049–D1056.

Ansorge, W.J. (2009) Next-generation DNA sequencing techniques. *N. Biotechnol.* **25**, 195–203.

Anthony, F., Combes, M., Astrorga, C., Bertrand, B., Graziosi, G., Lashermes, P. (2002) The origin of cultivated Coffea arabica L. varieties revealed by AFLP and SSR markers. *TAG Theor. Appl. Genet.* **104**, 894–900.

Au, K.F., Jiang, H., Lin, L., Xing, Y. and Wong, W.H. (2010) Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Res.* **38**, 4570–4578.

Auer, P.L. and Doerge, R.W. (2010) Statistical design and analysis of RNA sequencing data. *Genetics* **185**, 405–416.

Bettencourt, A. J and Rodrigues, C. (1988) Principles and practice of coffee breeding for resistance to rust and other diseases. *Elsevier Appl. Sci.* **3**, 199–234.

Boyd, L.A., Ridout, C., O'Sullivan, D.M., Leach, J.E. and Leung, H. (2013) Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends Genet.* **29**, 233–240.

Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94.

Cabral, P. G. C., Zambolim, E. M., Zambolim, L., Lelis T. P., Capucho, A.S. and and Caixeta, E.T. (2009) Identification of a new race of *Hemileia vastatrix* in Brazil. *Australas Plant Dis Notes* **4**, 129–130.

Carvalho, V., Chalfoun, S., Cunha, R. Manejo de doenças do cafeeiro. In: Reis, P., Cunha, R. (Eds.). Café arábica: do plantio à colheita. Lavras: U.R. EPAMIG SM, 2010. pp. 689 -755.

Capucho, A.S., Zambolim, E.M., Freitas, R.L., Haddad, F., Caixeta, E.T. and Zambolim, L. (2012) Identification of race XXXIII of *Hemileia vastatrix* on *Coffea arabica* Catimor derivatives in Brazil. *Australas. Plant Dis. Notes* **7**, 189–191.

Chalfoun, S., Carvalho, V. Efeito da produção e da composição química de folhas de cafeeiros sobre a intensidade de ataque de ferrugem (*Hemileia vastatrix* Berk. e Br.). In: Congresso brasileiro de pesquisas cafeeiras, 14., 1987, Campinas, Brasil. Resumos. Varginha: Fundação Procafé, 1987. pp. 121-122.

Chang, Z., Wang., Z. and Li, G. (2014) The impacts of read length and transcriptome complexity for de ovo assembly: A simulation study. *PLoS One* **9**, 1–8.

Chen, Q., Guo, W., Feng, L., Ye, X., Xie, W., Huang, X. and Liu, J. (2015) Transcriptome and proteome analysis of Eucalyptus infected with *Calonectria pseudoreteaudii*. *J. Proteomics* **115**, 117–131.

Diniz, I., Talhinhas, P., Azinheira, H.G., et al. (2012) Cellular and molecular analyses of coffee resistance to *Hemileia vastatrix* and nonhost resistance to *Uromyces vignae* in the resistance-donor genotype HDT832/2. *Eur. J. Plant Pathol.* **133**, 141–157.

Dmitriev, A.P. (2003) Signal Molecules for Plant Defense Responses to Biotic Stress. *Russ. J. Plant Physiol.* **50**, 417–425.

Emes, R.D. (2008) Inferring Function from Homology. In pp. 149–168.

Eulgem, T. (2005) Regulation of the Arabidopsis defense transcriptome. *Trends Plant Sci.* **10**, 71–78.

Fernandez, D., Santos, P., Agostini, C., et al. (2004) Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). *Mol. Plant Pathol.* **5**, 527–536.

Fernandez, D., Tisserant, E., Talhinhas, P., et al. (2012) 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Mol. Plant Pathol.* **13**, 17–37.

Finn, R.D., Bateman, A., Clements, J., et al. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–D230.

Freitas, R.L., Zambolim, E., Silva, M.C., Caixeta, E.T., Lelis, D.T., Zambolim, L., Sakiyama, N. (2014) Cytological evaluation of the infection process of *Hemileia vastatrix* (race XXXIII) in resistant and susceptible coffee. In proceedings of the 25th International conference on coffee science (ASIC). pp. 42–46. Armenia, Colombia.

Fu, Z.Q. and Dong, X. (2013) Systemic Acquired Resistance: Turning Local Infection into Global Defense. *Annu. Rev. Plant Biol.* **64**, 839–863.

Ganesh, D., Petitot, A.S., Silva, M.C., Alary, R., Lecouls, A.C. and Fernandez, D. (2006) 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.* **170**, 1045–1051.

Gichuru, E.K., Ithiru, J.M., Silva, M.C., Pereira, A.P. and Varzea, V.M.P. (2012) Additional physiological races of coffee leaf rust (*Hemileia vastatrix*) identified in Kenya. *Trop. Plant Pathol.* **37**, 424–427.

Goldsmith, Z. and Dhanasekaran, N. (2004) The Microrevolution: Applications and impacts of microarray technology on molecular biology and medicine (Review). *Int. J. Mol. Med.* **13**, 483–495.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52.

Han, L., Vickers, K.C., Samuels, D.C. and Guo, Y. (2014) Alternative applications for distinct RNA sequencing strategies. *Brief. Bioinform.* **16**, 629–639.

Han, Y., Guo, S., Muegge, K., Zhang, W. and Zhou, B. (2015) Advanced Applications of RNA Sequencing and Challenges. *Bioinform. Biol. Insights*, 29.

Kent, W.J. (2002) BLAT--The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656–664.

Kim, K.H., Kang, Y.J., Kim, D.H., Yoon, M.Y., Moon, J.-K., Kim, M.Y., Van, K. and Lee, S.-H. (2011) RNA-Seq Analysis of a Soybean Near-Isogenic Line Carrying Bacterial Leaf Pustule-Resistant and -Susceptible Alleles. *DNA Res.* **18**, 483–497.

Kou, Y. and Wang, S. (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. *Curr. Opin. Plant Biol.* **13**, 181–185.

Kushalappa, A.C., Yogendra, K.N. and Karre, S. (2016) Plant Innate Immune Response: Qualitative and Quantitative Resistance. *CRC. Crit. Rev. Plant Sci.* **35**, 38–55.

Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Lashermes, P., Bertrand, B., Etienne, H. Breeding coffee (*Coffea arabica*) for sustainable production. In: Jain, S. M., Priyadarshan, P. M. (Ed.). Breeding plantation tree crops. New York: Springer, 2009. pp. 525-544.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Lu, B., Zeng, Z. and Shi, T. (2013) Comparative study of de novo assembly and genome-guided assembly strategies for transcriptome reconstruction based on RNA-Seq. *Sci. China Life Sci.* **56**, 143–155.

Martin, J.A. and Wang, Z. (2011) Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**, 671–682.

Marraccini, P., Vinecky, F., Alves, G.S.C., et al. (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of Coffea canephora. *J. Exp. Bot.* **63**, 4191–4212.

Martinez, D.A. and Nelson, M.A. (2010) The Next Generation Becomes the Now Generation Richardson, P.M., ed. *PLoS Genet.* **6**, e1000906.

Mitchell, A., Chang, H.-Y., Daugherty, L., et al. (2015) The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **43**, D213–D221.

Moreton, J., Izquierdo, A. and Emes, R.D. (2016) Assembly, Assessment, and Availability of De novo Generated Eukaryotic Transcriptomes. *Front. Genet.* **6**.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628.

Nagalakshmi, U., Waern, K. and Snyder, M. (2010) RNA-Seq: A Method for Comprehensive Transcriptome Analysis. In Current Protocols in Molecular Biology., Hoboken, NJ, USA: John Wiley & Sons, Inc.

Nowrousian, M. (2010) Next-Generation Sequencing Techniques for Eukaryotic Microorganisms: Sequencing-Based Solutions to Biological Problems. *Eukaryot. Cell* **9**, 1300–1310.

Nowrousian, M., Stajich, J.E., Chu, M., et al. (2010) De novo Assembly of a 40 Mb Eukaryotic Genome from Short Sequence Reads: Sordaria macrospora, a Model Organism for Fungal Morphogenesis Richardson, P.M., ed. *PLoS Genet.* **6**, e1000891.

Oshlack, A., Robinson, M.D. and Young, M.D. (2010) From RNA-seq reads to differential expression results. *Genome Biol.* **11**, 220.

Ozsolak, F and Milos, P. (2010) RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98.

Pereira, A. A; Zambolim, L; Chaves, G.M; Sakiyama, N.S. (2000) Cultivar de café resistente à Ferrugem: Oeiras-MG 6851. *Rev. Ceres* **46**, 121–124.

Que, Y., Su, Y., Guo, J., Wu, Q. and Xu, L. (2014) A global view of transcriptome dynamics during Sporisorium scitamineum challenge in sugarcane by RNA-seq. *PLoS One* **9**.

Robertson, G., Schein, J., Chiu, R., et al. (2010) De novo assembly and analysis of RNA-seq data. *Nat. Methods* **7**, 909–12.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140.

Rodrigues, C.J., Bettencourt, A.J. and Rijo, L. (1975) Races of the Pathogen and Resistance to Coffee Rust. *Annu. Rev. Phytopathol.* **13**, 49–70.

Rodriguez, C., Varzea, I., Godinho, S.P. y R. and R (1993) New physiologic races of *Hemileia vastatrix*. In 15° Colloque Scientifique International sur le café., pp. 318–321. Montpellier, France.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**, 1086–1092.

Schwan-Estrada, K. R. F; Stangarlin, J. R; Pascholati, S.F. (2008) Mecanismos bioquímicos de defesa vegetal. In Interação Planta Patógeno – fisiologia, Bioquímica e Biologia Molecular. (Pascholati, S.F.; Leite, B.; Stangarlin, J.., ed), pp. 227–248. Piracicaba.

Silva, M.C., Nicole, M., Guerra-Guimarães, L. and Rodrigues, C.J. (2002) Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. *Physiol. Mol. Plant Pathol.* **60**, 169–183.

Silva, M. do C., Várzea, V., Guerra-Guimarães, L., Azinheira, H.G., Fernandez, D., Petitot, A.-S., Bertrand, B., Lashermes, P. and Nicole, M. (2006) Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian J. Plant Physiol.* **18**, 119–147.

Stein, L. (2001) Genome annotation: from sequence to biology. *Nat. Rev. Genet.* **2**, 493–503.

Teixeira, P.J.P.L., Toledo Thomazella, D.P. de, Reis, O., et al. (2014) High-resolution transcript profiling of the atypical biotrophic interaction between Theobroma cacao and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell* **26**, 4245–4269.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M.J. van, Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.

Várzea, V. M. P; Marques, D. V. (2005) Population variability of *Hemileia vastatrix* vs coffee durable resistance. In: Zambolim L, Zambolim E, Várzea VMP (eds) Durable resistance to coffee leaf rust. UFV Press, Viçosa-Brazil, pp 53–74.

Vossen, H. A. M. van der. Coffee breeding practices. In: Clarke, R. J., Vitzthum, O. G. (Ed.). Coffee: recent developments, agronomy. London: Blackwell Science, 2001. v. 1, p. 184-201.

Wang, K., Singh, D., Zeng, Z., et al. (2010) MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Res.* **38**, e178–e178.

Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63.

Westermann, A.J., Gorski, S. a. and Vogel, J. (2012) Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630.

Xie, Y., Wu, G., Tang, J., et al. (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* **30**, 1660–1666.

Yazawa, T., Kawahigashi, H., Matsumoto, T. and Mizuno, H. (2013) Simultaneous Transcriptome Analysis of Sorghum and *Bipolaris sorghicola* by Using RNA-seq in Combination with De Novo Transcriptome Assembly. *PLoS One* **8**.

Zambolim, L. (2016) Current status and management of coffee leaf rust in Brazil. *Trop. Plant Pathol.* **41**, 1–8.

Zhang, J. and Zhou, J.M. (2010) Plant immunity triggered by microbial molecular signatures. *Mol. Plant* **3**, 783–793.

4 COMPARATIVE GENE EXPRESSION PROFILE OF RESISTANT AND SUSCEPTIBLE COFFEE IN RESPONSE TO RUST DISEASE INFECTION

SUMMARY

Coffee rust disease is one of the major diseases in coffee throughout the world. To determine how the coffee transcriptome is regulated in response to Coffee rust infection, we present a comparative gene expression profile of Caturra (susceptible) and Hibrido de Timor (HdT, resistant) due to biotrophic interaction with *Hemilia vastatrix* during 0, 12, 24, 96 hours after inoculation (hai) and 17 days after inoculation (dai). This approach allows monitoring the gene expression in the coffee plant at different stages throughout the infectious process. The main objectives were to obtain a global overview of transcriptionally up-regulated and down-regulated genes in these coffee genotypes and analyze up-regulated HdT specific genes with potential exploratory characteristics for coffee breeding to develop new varieties with resistance to specific *H. vastatrix* races. To assemble the transcriptome, we used *Coffea canephora* as a reference genome and *de novo* assembly approach as a complement, resulting in a total of 43,159 transcript sequences. The results suggested that at early infection events (12 and 24 hai), HdT responded to the attack of *H. vastatrix* with a larger number of up-regulated genes than Caturra. We selected thirteen differentially expressed HdT exclusive genes to evaluate their expression pattern using qPCR. A set of resistant genotype genes showed higher level of expression than in susceptible at early stage of *H. vastatrix* infection. These genes could be candidates for coffee breeding programs. Collectively, our results provide understanding of expression profiles in *coffee – H. vastatrix* interaction over a time course in susceptible and resistant coffee plants.

Keywords: Coffee rust, Transcriptome, biotrophic interaction, resistance.

INTRODUCTION

Coffee is a worldwide-appreciated beverage and an important commodity for several countries. *Coffea arabica* and *Coffea canephora* are the two most economically important species (Davis *et al.*, 2011). *C. arabica*, a natural alotetraploid (2n = 4x = 44) hybrid between the two diploids species *C. canephora* and *C. eugenioides* (Lashermes *et al.*, 1999), accounts for 70% of the world production (Davis *et al.*, 2011). One of the major problems in coffee production is the coffee rust disease, which can cause yield losses up to 50% (Zambolim, 2016). The causal agent of this disease, *Hemileia vastatrix*, is a biotrophic fungus, which it is entirely dependent on the cells of living plants for its growth and reproduction. The fungus produces several specialized structures to colonize the plant. After spore germination, the germ tube differentiates to apressoria, which, in turn, became penetration hyphae at the stomata. This structure will reach the substomatic chamber where it will differentiate to haustoria that will invade adjacent cells altering plant metabolism to meet its nutritional needs and, thus, completing the fungus lifecycle. This mode of interaction involves a prolonged and effective suppression of the host immune system and, at the same time, the induction of host-specific genes for its establishment as biotrophic fungus (Schulze-Lefert and Panstruga, 2003; Voegele and Mendgen, 2003; Guerra-Guimarã*es et al*., 2015). Therefore, in coffee - *H. vastatrix* pathosystem, identification of genes involved in the plant defense mechanism triggered by the presence of the pathogen is indispensable to find new defense genes to develop resistant cultivars.

Plant defense against pathogens starts with the recognition of pathogen-associated molecular patterns (PAMPs) by transmembrane proteins named pattern recognition receptors, PRRs (Jones and Dangl, 2006). This step activates the PAMP-triggered imunity (PTI). Pathogen can overcome this initial defense by secreting small extracellular proteins called effectors (Wit *et al.*, 2009). However, plants can recognize these effectors direct or indirectly by a resistance (*R*) protein, and induce a second defense phase, named Effector-triggered immunity (ETI), which is more efficient than PTI (Wit *et al.*, 2009). Each *R* gene recognize a specific effector (called *Avr* gene, for Avirulent) (Flor, 1942), therefore, plant immunity is related to the presence of specific alleles of

R genes. Some pathogens have evolved their effectors proteins either by losing an *Avr* gene, or diversifying its genes into different effector protein and thus, it can suppress ETI and successfully colonize the host (Jones and Dangl, 2006).

Cytological and biochemical studies have shown coffee activates some defense mechanisms when attacked by *H. vastatrix*. One such mechanism is the hypersensitive response (HR), a type of ETI, which is associated with the deposition of callose, phenolic compounds (flavonoid and chlorogenic acid), and cell wall lignification (Silva *et al*., 2002). Other types of responses involve enzymes such as lipoxygenase and peroxidase that stimulate the pathway of phenylpropanoids. PR proteins (pathogen related proteins), as β-1, 3-glucanase and chitinase, are also found in association with resistance of some coffee cultivars to *H. vastatrix* race II (Silva *et al*., 2002; Fernandez *et al*., 2012). Molecular methodologies as suppression subtractive hybridization (SSH), 454 pyrosequencing and RT-qPCR have been used to identify several genes putatively involved in host resistance. Some of the expressed sequence tags (ESTs) found in these studies encode for proteins involved in resistance, stresses, defense and signal transduction pathways (i.e. chitinases, beta-1, 3 glucanases, PR10, lipoxygenase type-AP2 and WRKY transcription factors). Also, activity of oxidative enzymes (lipoxygenase, peroxidase and superoxide dismutase), phenylalanine ammonia lyase, chitinase and glucanase were detected in the resistance reaction to *H. vastatrix* race II (Fernandez *et al*., 2004; 2012; Ganesh *et al*., 2006; Diniz *et al*., 2012). Coffee resistance against *H. vastatrix* is governed by at least nine major dominant genes (S_H1-S_H9) that have the corresponding virulence genes (v_1–v_9) in the pathogen (Rodrigu*es et al*., 1975; Bettencourt, A. J and Rodrigues, 1988; Várzea and Marques, 2005). However, resistance breakdown by virulence factor *v5,* which is present in the race II of *H. vastatrix* and at least thirty other races, has been reported in most commercial coffee cultivars with S_H5 gene (Rodriguez *et al*., 1993).

Until 2005, more than 50 races of *H. vastatrix* had been broadly described worldwide (Várzea and Marques, 2005; Zambolim, 2016). In Brazil 15 races were identified: I, II, III, VII, X, XIII, XV, XVI, XVII, XXI, XXII, XXIII, XXIV, XXV or XXXI, and XXXVII (Cabral *et al*., 2009; Capucho *et al*., 2012; Zambolim, 2016). The *H. vastatrix* races recently discovered have been able to infect derivatives of Híbrido de Timor (HdT), a natural hybrid originated from *C.*

arabica and *C. canephora* crossing (Bettencourt, 1973; Rodrigues Júnior *et al*., 2004). HdT is the main source of resistance used in breeding programs throughout the world (Gichuru *et al.*, 2012; Avelino *et al.*, 2015; McCook and Vandermeer, 2015; Vossen *et al.*, 2015; Zambolim, 2016). The presence of such *H. vastatrix* virulent races in the field represents a serious risk to coffee production worldwide. An example is the cultivar Oeiras MG 6851, originated from a crossing between *C. arabica* cv. Caturra (CIFC 19/1) and HdT (CIFC 832/1) and released as an important rust resistant cultivar (Pereira *et al*., 2000). Twelve years after its release, the resistance was broken by race XXXIII of *H. vastatrix* (Capucho *et al*., 2012). Thus, to study the dynamics of the interaction between the pathogen and the plant through transcriptome profiling will help to understand the supplanting of plant resistance by new physiological races of the fungus.

Despite all the efforts, there is limited information available on the transcriptome analysis to reveal which defense genes are involved in response to *H. vastatrix* infection. Transcriptome profiling studies are essential to fully understand the biological pathways that are activated in various physiological conditions or stages of an organism development (Wang *et al.*, 2009; Ozsolak and Milos, 2010). Expression profiling of host and pathogen can provide a new understanding of this interaction, and allow the identification of virulence genes in the pathogen or defense pathways in the host cells (Westermann *et al*., 2012; Boyd *et al*., 2013). Here, we report a time-course high throughput transcriptome analysis of a susceptible and resistant genotypes challenged by *H. vastatrix* infection. In this approach, we could identify genes related to both compatible (susceptible) and incompatible (resistant) interactions that are important for disease development and resistance. This knowledge will help breeding programs to develop new varieties with resistance to specific *H. vastatrix* races.

RESULTS AND DISCUSSION

Read quality filtering, mapping and assembly

Transcriptome of coffee during compatible (Caturra) and incompatible (HdT) interactions with *H. vastatrix* race XXXIII was analyzed using various

bioinformatics tools. Firstly, a total of 103,031,664 2X250 pair-end reads were generated from 10 libraries sequenced by the Illumina Miseq platform (Table 1). The overlap between paired reads was performed producing 70,450,936 overlapped reads with average length of 250 bp. This approach generated assembled (overlapped) and unassembled (R1 and R2) 300-bp reads, which were concatenated generating a total of 110,135,873 single-end reads (Fig. 6 and Table 1). As such, the quality and average length of reads was increased improving the chance to obtain good transcriptome assembly.

Since *C. arabica* reference genome has not made available to date, we used a genome from a closely related species, i.e. *C. canephora*, as a reference for read mapping. In the susceptible genotype, read mapping against *C. canephora* genome showed a decrease of alignment along the infection time course, starting from 48% (0 hai) and reducing to 33% (17 dai) (Fig. 1a). In contrast, resistant library alignment was almost constant at all time points (Fig 1b). On the contrary, when mapping was made against *H. vastatrix* partial genome (Cristancho *et al*., 2014), the mapping of reads in the susceptible genotype during the early hours of infection (12 and 24) was 3%, then started to increase at 96 hai, up to 8% in the 17 dai library (Fig. 1a). This final library showed a high percentage of reads mapped against *H. vastatrix* genome due to increased fungus biomass at this time point. Approximately 20 days after the start of the infection process, hyphae forming a large mycelial density, appearing on the outside through the stomata, a uredosporic serum in a "bouquet" occurs in the area of penetration (Silva *et al*., 2006). In resistant genotype the effect was the opposite; the amount of the fungus reads aligned began to decrease from 12 hai and onwards (Fig. 1b).

Table 1 Total raw and concatenated reads from coffee - *H. vastatrix* interaction libraries.

Source	Library	Total Paired end Reads	Total Overlapping Reads	Total concatenated Single-end Reads
Susceptible	0 hai	6 129 221	2 845 924	7 525 660
	12 hai	12 468 894	10 226 867	11 323 418
	24 hai	9 574 685	6 896 300	9 621 706
	96 hai	6 708 346	4 750 954	6 695 765
	17 dai	12 285 892	9 483 293	11 679 975
Resistant	0 hai	13 647 518	8 071 594	16 351 660
	12 hai	11 164 309	8 925 786	11 478 694
	24 hai	15 530 942	8 253 540	19 494 216
	96 hai	6 179 620	3 030 221	7 471 180
	17 dai	9 342 237	7 966 457	8 493 599
	Total	103 031 664	70 450 936	110 135 873

hai, hours after inoculation., dai, days after inoculation

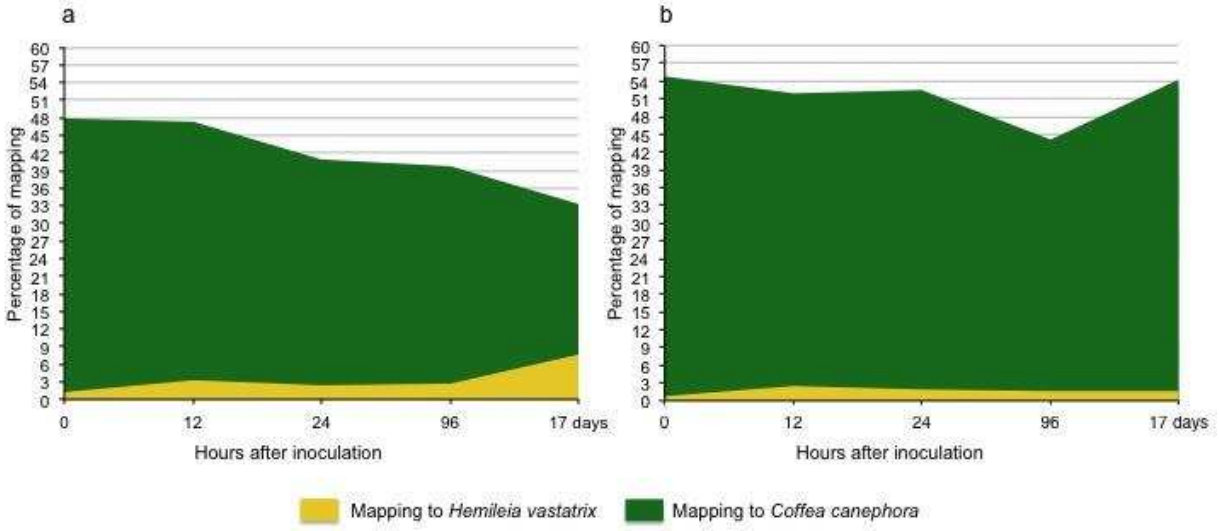


Fig. 1 Percentage of reads mapping against *Coffea canephora* (green) and *Hemileia vastatrix* (yellow) reference genomes during infection course. (a) Susceptible libraries (b) Resistant libraries.

The transcriptome assembly using *C. canephora* reference genome produced 28,119 contigs with an N50 value of 1,626 (Table 2). Approximately 50% of the total reads of each HdT' libraries mapped to *C. canephora* genome, which is higher than observed for the susceptible libraries (Fig. 1). This higher mapping percentage in the resistant libraries is probably due to HdT origin, since it is a natural hybrid between *Coffea arabica* and *C. canephora* (Bettencourt, 1973; Rodrigues Júnior *et al*., 2004), therefore a larger part of its genome derives from *C. canephora*. As a complement to the strategy using *C. canephora* as a reference, and in order to increase the number of contigs to have a greater coverage of the transcriptome, *de novo* assembly of unaligment reads was adopted in this study. We obtained a total of 15,040 contigs with N50 of 1,317 (Table 2). A total of 43,159 contigs were produced using *C. canephora* genome reference and *de novo* assembly. This number is much greater than reported by Fernandez *et al* (2012) in a transcriptome analysis of *C. arabica* CIFC H147/1 (resistant) during the interaction with *H. vastatrix* using the 454-pyrosequencing platform, the only published study reporting transcriptome profiling of this pathosystem. The authors produced 352,146 reads assembled into 13,951 contigs with an average size of 631 bp. The functions of the majority of the genes were related to caffeine or chlorogenic acid biosynthesis, and genes associated to plant defense response such as *WRKY* transcription factor, pathogenesis-related (PR) proteins (1,3-b-glucanases, PR1b, PR-5 of the thaumatin-like protein family and chitinases). Similar genes were founded in our experiment; however, other HdT exclusive genes such as NAC, MYB, Beta-glucosidase, F-box protein and peroxidase were found involved in defense mechanisms at early hours (12 and 24) of *H. vastatrix* infection.

In addition to obtaining a greater number of contigs, we analyzed both compatible and incompatible coffee - *H. vastatrix* interaction with the aim of providing a better understanding of gene expression profile. Moreover, here, we used Illumina sequencing which has less low error rate, compared to 454-pyrosequencing platform (Loman *et al*., 2012). Recently, other transcriptome studies of coffee, however not involving plant-pathogen interactions, showed similar number of contigs assembled as in this present study. Leaf and fruit transcriptome analysis of *Coffea eugenioides* produced 36,935 contigs using Illumina HiSeq platform (Yuyama *et al*., 2016). Also Mofatto *et al* (2016)

obtained a total of 41,512 contigs from *C. arabica* transcriptome, comparing the molecular responses to drought in two commercial cultivars using 454 - pyrosequencing and Sanger platforms. The expression profiles (read counts) in all libraries were submitted for PCA analysis in order to evaluate the intrinsic variation between libraries. A distinction between the group of resistant and susceptible genotype libraries was observed when *C. canephora* was used as a reference for transcriptome assembly (Fig. 2). The results from *de novo* assembly (data not shown) are similar to the result elucidated by *C. canephora* assembly. The PCA clearly revealed that the greatest changes in host transcriptome were caused by the fungal infection following the interaction process. In particular, in susceptible genotype 17 dai library was well separated from all other libraries.

Table 2 Statistical assembly parameters in *C. canephora* and the *de novo* assembly

Assembly	Total contigs	Mean size (bp)	N50	# Total base pair
C. canephora	28 119	1296	1626 (7148 contigs)	36 468 733
de novo	15 040	1274	1317 (5370 contigs)	19 161 987

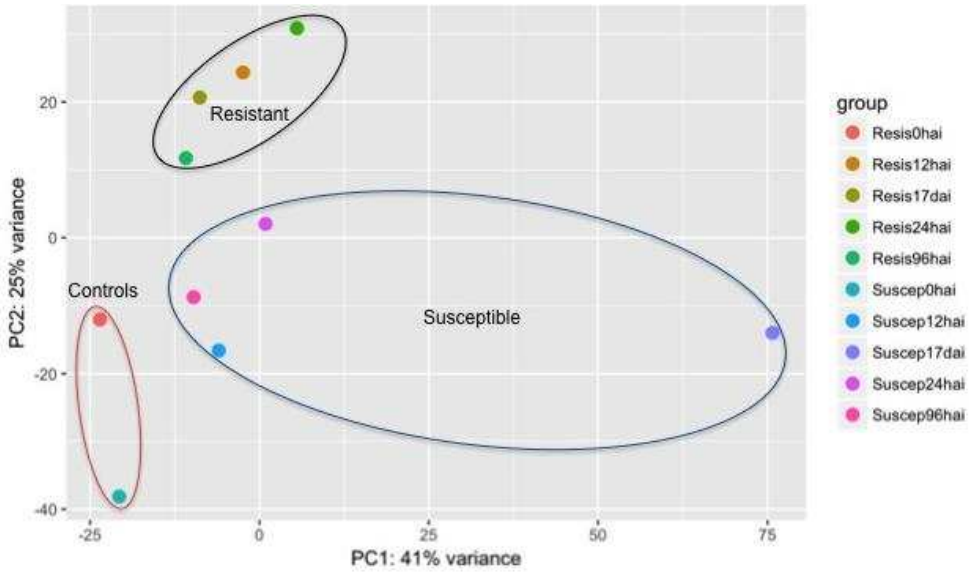


Fig. 2 Principal component analysis (PCA) of susceptible and resistant genotype libraries during the interaction with *H. vastatrix*.

Profiling of differentially expressed genes (DEGs)

Plants are in a constant struggle with pathogens, and use different defense mechanisms to preserve their integrity, such as PTI and ETI. In coffee - *H. vastatrix* interaction, the pathogen establishes a biotrophic relation with the host within the first hours after infection (<24 hai) (Ramiro *et al.*, 2009; Guerra-Guimarães *et al*., 2015; Maia *et al.*, 2016; Talhinhas *et al.*, 2016). Here, at early infection events (12 and 24 hai), HdT responded to the attack of *H. vastatrix* with a larger number of up-regulated genes than Caturra, which is indicated as a higher red intensity in heatmap (Fig. 3a) and total number of up-regulated genes (Fig. 3b), which shows around 500 more up-regulated genes than in Caturra. This markedly gene up-regulation in HdT may reflect the exploitation of cellular resources and / or the activation of defense responses (Grenville-Briggs and West, 2005). In fact, it has been reported that coffee plants with complete resistance to *H. vastatrix,* fungus growth ceases in the early stages of infection where there is a disruption of the cytoplasmic contents of the fungal infection structures (Diniz *et al.*, 2012). Therefore, this overall up-regulation of transcription might be a direct consequence of fungus attack and it is responsible for a coordinate and effective defense.

Generally, sets of genes in plants are induced against a biotic or abiotic stress condition. This response may be early or late, but the first response is key to induce signaling cascades for expression of genes involved in defense mechanisms (Kumar and Kirti, 2011). For the majority of pathosystems, all this process starts with the recognition or perception, signal transduction, followed by the activation of defense genes (Oliveira *et al*., 2016). By inducing a hierarchical network of genes that regulate the expression of resistance genes, the advancing pathogen within the plant is suppressed (Kushalappa *et al*., 2016). Here, we found several DEGs encoding receptors in the transcriptome of HdT such as receptor-like kinases (RLKs) and receptor-like proteins (RLPs) in the initial stage of infection (data not shown). These receptors could be related to pathogen perception, such as PAMPs recognition by PRRs, which leads to a signaling cascade that alerts the plant to the presence of a pathogen and induces PTI (Ishiga *et al.*, 2013). This PTI response frequently involves ion fluxes, the production of reactive oxygen species, protein phosphorylation,

ethylene biosynthesis, and callose deposition (Boller and Felix, 2009). Linked to such functions during the early hours of HdT-*H. vastatrix* interaction, we found genes such as calcium-transporting, peroxidase, ethylene-responsive transcription and callose synthase.

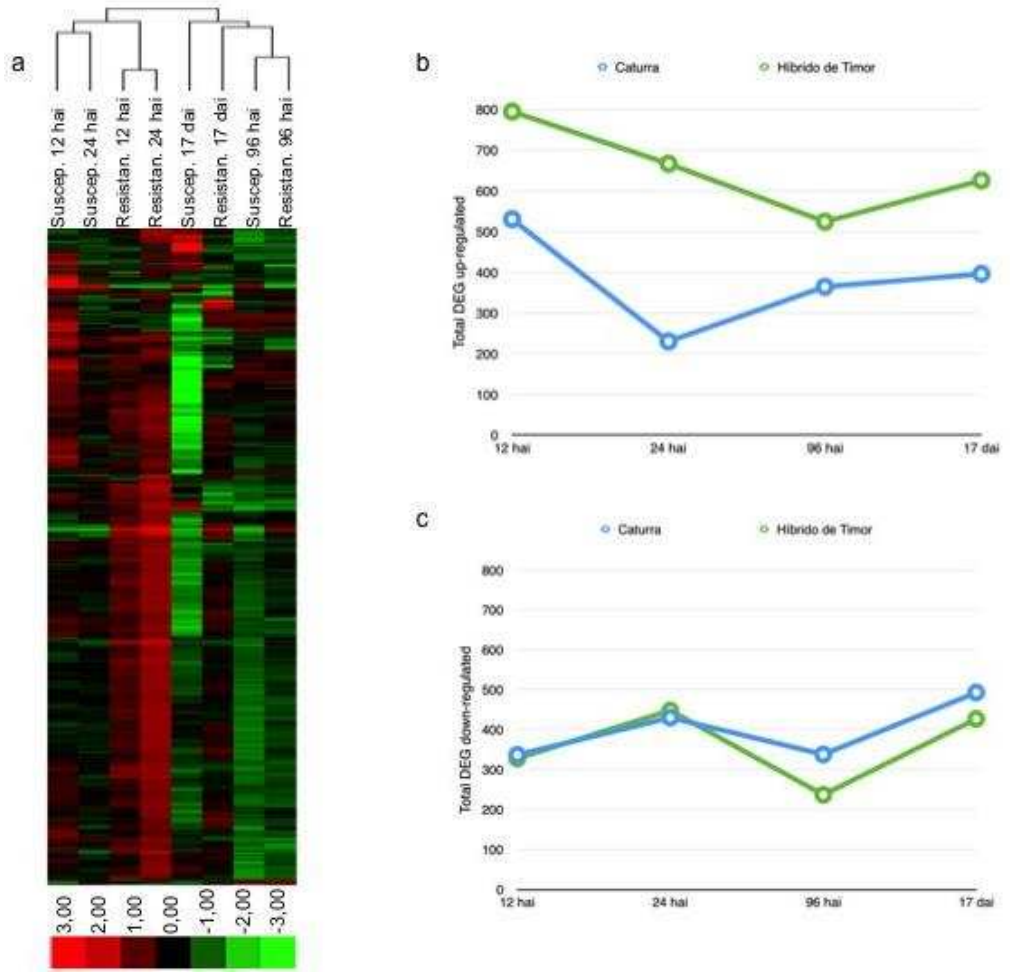


Fig. 3 Expression profile in Híbrido de Timor (resistant) and Caturra (susceptible) inoculated with *H. vastatrix*. (a) Hierarchical clustering according to the changes in expression level along the time course. Cluster analysis was performed using Cluster 3.0 software. Each column represents the Log_2 fold change in transcript levels in coffee at the indicated times. Up-regulated and down-regulated genes are shown in red and green, respectively. The intensity color scale indicates the level of expression (b) and (c) Total up and down-regulated genes, respectively.

Additionally, we confirmed that the early response of HdT to the *H. vastatrix* attack was related to pre-haustorial resistance as suggested (Heath, 1977; Mellersh and Heath, 2003; Niks and Rubiales, 2002; Freitas *et al*., 2014). Usually the haustoria formation of *H. vastatrix* begins after 24 hai in the host (Ramiro *et al.*, 2009; Maia *et al.*, 2016). Although, in coffee is also possible to found post haustorial resistance, the type of response depend of the coffee genotype (Silva *et al.*, 2006). The pre-haustorial response was also observed in the barley-*Puccinia graminis* f. sp. *tritici* pathosystem, during this interaction the resistance gene *RPG1* was expressed a few minutes after inoculation with stem rust races (Nirmala *et al.*, 2010). More recently, the same pathogen it was observed early responses to avirulent stem rust races in wheat carrying resistance genes *Sr5* and *Sr36* involved callose deposition in stomatal guard cells (Wang *et al.*, 2015). It is possible that the genes identified here as induced in early hours in HdT may trigger a rapid and efficient response for the plant defense, and these mechanisms may involve pre-haustorial resistance in coffee-*H. vastatrix* interaction, in the same way found for barley and wheat pre-haustorial defense to *P. graminis*. Niks and Rubiales (2002) suggest that combining different resistance mechanisms acting at different stages of the infection process could provide multiple barriers that are not easily overcome by simple race-type changes of the pathogen. In addition, different members of a host plant species contain different *R* gene complements. Rust infections can therefore lead to either a resistant or a susceptible outcome, depending upon the plant and pathogen genotypes involved (Bettgenhaeuser *et al.*, 2014).

In the case of a group of genes in susceptible plant, more up-regulated genes at 12 hai (Fig. 3a) may be associated with basal resistance of the plant (Jones and Dangl, 2006). In the following time points, however, the level of gene expression was suppressed (Fig. 3a). Note that in all time points, HdT showed a higher gene expression, which can be noticed by the color intensity in heatmap (Fig. 3a). This shows that transcription programming in the resistant genotype is completely different from susceptible genotype. On the other hand, the number of down-regulated genes is comparable in both genotypes (Fig. 3b). Plant–pathogen interactions are complex processes that trigger a series of molecular responses at several expression levels. While resistant plants initiate responses in incompatible interactions, susceptible plants can also launch a

series of basal defense responses in compatible interactions. Although they present similar expression profiles, defense gene induction in compatible interactions occurs later than that in incompatible interactions (Balaji *et al*., 2008; Lara-Ávila *et al*., 2012).

BLAST2GO analysis was performed for genes up-regulated at 12 and 24 hai in both resistant and susceptible genotypes (Fig. S1). These time points are when actual biotrophic interactions are assumed to occur. The result presented for the biological process category, the two genotypes showed the existence of up-regulated genes associated with programmed cell death but at different time points (12hai in susceptible and 24 hai in resistant genotype). It could be inferred that programmed cell death is a delayed response to efficiently deter the ingression and establishment of the invading pathogen. A microscopic study revealed that in coffee and *H. vastatrix* race XXXIII pathosystem, 30% of sites with HR were observed at 12hai in caturra and HdT. However, this same study showed an increase HR to 70% in HdT at 24hai and 30% in caturra (Freitas *et al*., 2014). On the other hand, for the molecular function category for the majority of the genes (70%) were represented by genes involved in catalytic activity in resistant genotype. Of remarkable difference that can be noticed is the higher proportion of up-regulated genes predicted as cell wall component under the cellular component category in resistant genotype as part of structural resistance.

On the other hand, analysis of differentially expressed genes using *C. canephora* genome reference, represented using venn diagrams in all time courses between Caturra and HdT, allowed to identify one set of up-regulated genes unique to HdT for each time of infection (Fig. 4a-d), the results from *de novo* assembly (data not shown) are similar to the result elucidated. However, within each time it was possible to observe that both genotypes have some genes in common (Fig. 4a-d), some of which could be associated with basal resistance in both genotypes, considering that Caturra does not show vertical resistance against *H. vastatrix* (Silva *et al.*, 2006). Basal resistance is a kind of defense response mechanism common in plants when they are attacked by biotrophic fungi (Niks *et al.*, 2015), and that is initiated during the early phases of pathogen infection (Gill *et al.*, 2015). Finally, one more intersection of this pool of genes allowed us to choose 13 genes exclusively up-regulated in HdT,

which showed importance within the plant defense mechanisms. This selection was made from 219, 129 and 118 genes corresponding to 12, 24 and 96 hours after infection in HdT (Fig. 4e).

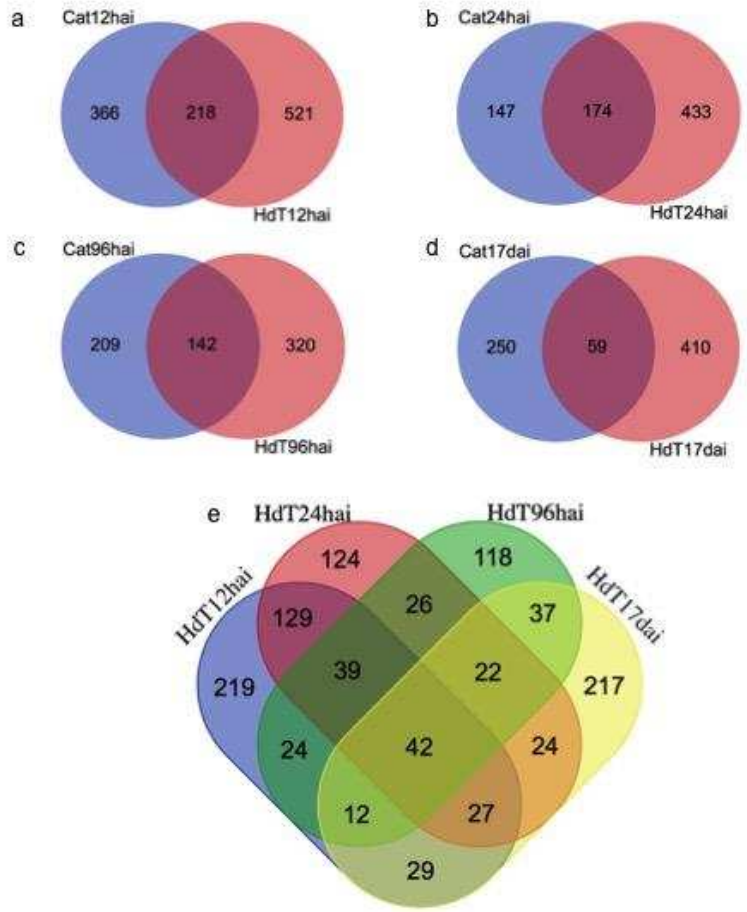


Fig. 4 Number of up-regulated genes (DEGs) of the *C. canephora* assembly in Caturra and HdT in response to *H. vastatrix* infection (a) Venn diagram displaying the number of up-regulated DEGs between Caturra (Cat) and HdT at 12hai (b) Up-regulated DEGs between Cat and HdT at 24hai (c) Up-regulated DEGs between Cat and HdT at 96hai (d) Up-regulated DEGs between Cat and HdT at 17dai (e) Up-regulated DEGs exclusive to HdT at all times investigated.

The qPCR analysis of 13 different types of genes related to plant defense mechanisms were conducted at 0, 12, 24 and 96 hai in susceptible and resistant coffee genotypes. Expression pattern in qPCR was similar to transcriptome analysis for most of the studied genes (Fig. S2). A similar correlation between qPCR and transcriptome profiling was reported by Rubio *et*

al (2015). Both analyzes have their own strengths and limitations, the different expression levels between transcriptome and qPCR could be caused by the bioinformatics process performed in the transcriptome analysis, which includes several factors that can affect the reproducibility of quantitative expression profiles, including alignment choices, estimation of transcript expression, etc. (Labaj *et al.*, 2011). The first group of analyzed genes was transcription factors (*Putative basic helix-loop-helix bHLH DNA-binding superfamily protein,* and *Ethylene-responsive transcription factor 1B*), which showed early expression in HdT in contrast to Caturra (Fig. 5a-b). The *bHLH* (Fig. 5a) presented a high peak expression at 12hai in the incompatible interaction. The results suggest that this gene may be involved in regulating the expression of some resistance genes or defense mechanisms only in resistant genotype. Kundu *et al* (2015) evidenced that the expression profiling of transcription factor of this same family (bHLH) may play a complementary and/or overlapping role in enhancing expression of downstream components of the defense pathway.

The second transcription factor, *Ethylene-responsive transcription factor 1B* (ERF) presented two expression peaks (at 12 and 24 hai) in resistant genotype and also significantly higher when compared to susceptible genotype (Fig. 5b). Usually, the transcription factors are activated by external stimuli and subsequently regulate the expression of genes involved in plant defense. A number of ERF genes confer tolerance to various biotic stresses when expressed in different plants under different conditions (Licausi *et al*., 2013). There is evidence that ERFs induce transcription resistance genes, pathogenesis-related (PR) genes, osmotin, chitinase and b-1,3-glucanase (Licausi *et al*., 2013). This kind of gene has been selected through evolution to regulate a series of stress-response pathways and could be used in genetic engineering for the breeding of plants with specific traits related to stress tolerance (Licausi *et al*., 2013). Ganesh *et al* (2006) evaluated the expression of three defense related genes during coffee - *H. vastatrix* interaction, including a transcription factor. These genes showed that they were transiently induced during early stages (12–24 hai) of pathogen infection. They reported that two of the three genes (*CaR111,* gene encoding a protein of unknown function and *CaWRKY1*, a transcription factor) were up-regulated in incompatible samples, which suggest their involvement in defense mechanism.

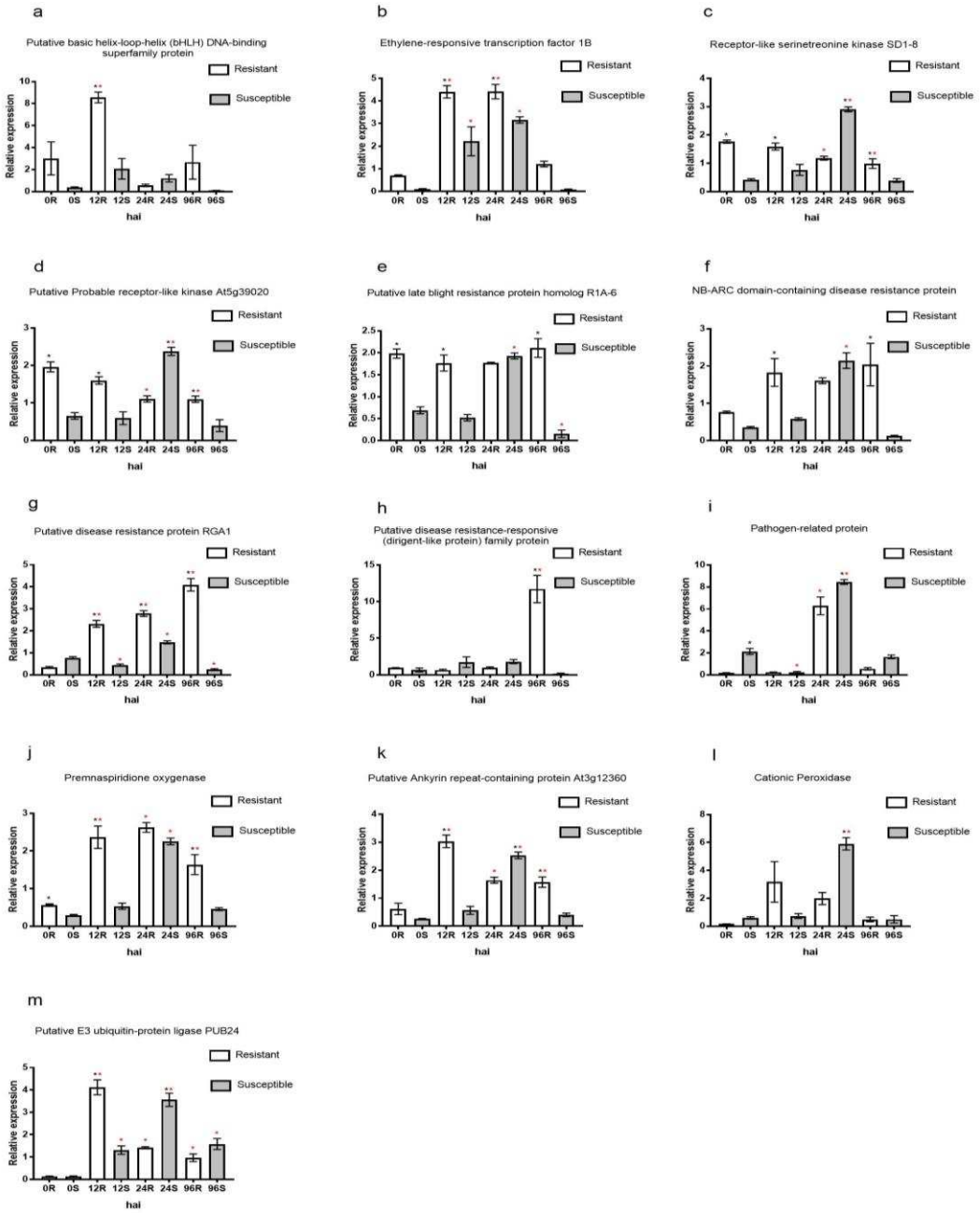


Fig. 5 Validation of candidate genes by qPCR. Y- axis represent the relative expression at various times after *H. vastatrix* inoculation in susceptible and resistant coffee genotypes. Error bars = SEM, n = 3 independent biological replicates. *Significantly up-or down-regulated relative to uninoculated samples (0 hai) within interaction as defined by one-way ANOVA followed by Dunnett test ($p<0.05$); *Significant difference in expression level between the same hai across interactions defined by Tukey test ($p<0.05$)

The group of genes related to recognition of pathogen elicitor proteins (*Serine-threonine protein-receptor-like kinase SD1-8* and *Probable putative*

protein receptor-like kinase (At5g39020 homolog)), exhibited a constant expression in HdT during infection process while the expression in Caturra was delayed (24 hai) (Fig. 5c-d). Such behavior are associated with signal perception through their extracellular domain and propagate the signal through the intracellular kinase domain to activate *R* genes (Afzal *et al*., 2008). A similar expression profiling was found with the resistance gene group: *Putative late blight resistance protein homolog R1A-6*, *NB-ARC domain-containing disease resistance protein* and *Putative disease resistance protein RGA1* (Fig. 5e-g). These genes are most likely involved in the activation of plant defense mechanisms and their expression during all time course is possibly associated with controlling the differentiation of several fungal structures. In coffee, Diniz *et al* (2012) evaluated the expression of *CaRLK*, which is involved in recognition, signaling and defense, showed that this gene was activated in the early events of the HdT CIFC 832/2 – *H. vastatrix* interaction. *CaRLK* showed expression peak at 6 to 12 hai, a period when appressoria and penetration hypha differentiate. According to this report, the second highest expression peak was when the anchors and haustoria mother cells differentiate (21–24 hai). In the present work, *RLK* was continuously expressed along all the time courses studied in resistant genotype unlike in the susceptible genotype in which the expression was delayed (Fig. 5c-d). The result suggests that this gene is one of the key defense components which is in accordance to the work of (Diniz *et al.*, 2012).

In the case of *Putative disease resistance-response* (dirigent-like protein) family protein, it presented a unique and high relative expression at 96 hai (Fig. 5h). This might be a case of an *R*-genes that is highly genotype specific (Narusaka *et al*., 2013) and downstream activated by signaling cascades (Hammond-Kosack and Jones, 1997). The PR protein showed high expression in both genotypes at 24 hai, however the susceptible was higher than resistant genotype (Fig. 5i), which suggests that in general these types of genes are important in the defense of plants against pathogens, mainly in susceptible plants. Several types of PR proteins in resistant and susceptible wheat genotypes were also reported as up-regulated (Xin *et al.*, 2012). The PR protein expression is often triggered by pathogen infection, limiting the pathogen progression (Silva *et al.*, 2006).

Another group of genes analyzed by qPCR were *Premnaspiridione oxygenase*, *Putative Ankyrin repeat-containing protein* At3g12360, *Cationic peroxidase* and *Putative E3 ubiquitin-protein ligase PUB24*. The first gene encodes solavetivone, a potent antifungal phytoalexin (Takahashi *et al.*, 2007). Generally, phytoalexins are organic compounds, which possess antimicrobial or repellent activities (Jeandet *et al*., 2013). Their production starts a few hours after the pathogen or pest attack (Pedras *et al.*, 2011). The results found here are in agreement with the literature, since in the resistant genotype *Premnaspiridione oxygenase* up-regulation starts at 12hai while in the susceptible genotype it only starts at 24 hai (Fig. 5j). The putative ankyrin repeat-containing protein followed a similar expression pattern (Fig. 5k). Functional characterization of *Ankyrin repeats proteins* across species have indicated a conserved role for them in protection against pathogen and disease resistance, as promoter of Systemic Acquired Resistance (SAR) (Sharma and Pandey, 2016). This protein is important for transducing the Salicylic acid (SA) signal. In some *R-avr*–mediated interactions, SA is required for the *R* gene–dependent host programmed cell death (called the hypersensitive response, HR) and/or for disease resistance (Lu, 2003). Perhaps, the induction of *Ankyrin repeat–containing protein* in the resistant plant, within the first hours after infection, promotes the production of SA, important compound to trigger HR.

The *Cationic peroxidase gene* is related with the accumulation of lignin-like compounds and reduction in pathogen multiplication in leaves and onset of the HR. We found this gene up-regulated at 12 hai in HdT and at 24 hai in Caturra (Fig. 5l). Its rice homologue is also induced at 24 hai during incompatible interaction with *Xanthomonas oryzae* pv oryzae pathogen (Young *et al*., 1995) and during compatible interaction at 48 hai. The early expression of cationic peroxidase in HdT may be involved in most important responses in coffee HR (Silva *et al*., 2002). *Putative E3 ubiquitin-protein ligase PUB24* is related to different positive and negative functions in different steps of plant defense such as regulation of RLKs (Duplan and Rivas, 2014), involved in recognition of pathogen elicitor proteins (Kenn, 1999; Ebel, 1997; Shibuya and Minami, 2001; Peck, 2003). Wang *et al* (2006) demonstrated that the interaction of *E3 Ubligase* gene with a *RLKs* type protein triggered cascades of defense gene activation in rice plant against *Xanthomonas oryzae* pv oryzae. Here, we

found a high gene expression of *Putative E3 ubiquitin-protein ligase PUB24* at 12 hai (Fig. 5m) in the resistant plant at the same time of an up-regulation of an *Serine-threonine protein-receptor-like kinase SD1-8* and *Probable putative protein receptor-like kinase* (At5g39020 homolog) genes (Fig. 5 c-d). Therefore, due to their expression pattern, it is reasonable to hypothesize that these three genes might be involved in a defense mechanism similar to that found in rice against *X. oryzae* pv *oryzae*, where coffee E3 ubiquitin-ligase acts together with these two receptor kinases to activate the signaling cascade leading to defense response.

Taken together, the qPCR expression analysis between resistant and susceptible genotypes confirmed that genes such as *putative basic helix-loop-helix bHLH DNA-binding superfamily protein*, *ethylene-responsive transcription factor 1B, putative disease resistance protein RGA1, putative disease resistance-response (dirigent-like protein) family protein*, and *premnaspiridione oxygenase* showed higher expression at early stage upon challenge with the biotrophic pathogen *H. vastatrix* in resistant plant. In contrast, no such up-regulation in early hours was detected in susceptible genotype. These genes could represent potential candidates for various biotechnological and effective management tools for operational application in coffee breeding programs. In general, the regulation of gene expression is a dynamic process. We confirmed that early expression patterns of up-regulated genes in HdT are directly related to prehaustorial resistance. This transcriptome study paves a way for more detailed work on coffee resistance to *H. vastatrix*, which will be important for future development of plants with certain adaptive characteristics to coffee rust.

EXPERIMENTAL PROCEDURES

Plant inoculation, library preparation and RNA Sequencing

Two coffee genotypes were used in this study: *C. arabica* cv. caturra vermelho CIFC 19/1 (susceptible) and Híbrido de Timor CIFC 832/1 (resistant). They are the parents of cultivar Oeiras, the one that have the resistance supplanted by the race XXXIII of *H. vastatrix*. Plants (juvenile leaves from second pair) were inoculated with spores of *H. vastatrix* race XXXIII and sampled at 0, 12, 24, 96 hours after inoculation (hai) and 17 days after

inoculation (dai), totaling 10 different samples (5 time points for each of the 2 genotypes). Total RNA was extracted with RNeasy Plant Mini Kit (Qiagen) and cDNA synthesis was performed using the Mint-2 cDNA Synthesis kit (Evrogen), according to the manufacturer's instructions. After this step, the samples were normalized with TRIMMER – cDNA Normalization kit (Evrogen), to reduce the abundant transcripts, especially rRNAs. RNA sequencing was performed using MiSeq platform (Illumina) at University of North Carolina Chape Hill, NC, USA, with TruSeq DNA Sample Preparation protocol (Illumina). This sequencing produced 2x250pb paired-end reads for each genotype and time sampled (Freitas, 2015), and these 10 libraries were sequenced in 10 MiSeq runs, each run comprising all libraries in a multiplex reaction, using a different barcode for each library.

Quality assessment and overlapping paired-end reads

Read quality was assessed with *FastQC* software version 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). The strategy used to improve read quality and length was overlapping the forward (R1) and reverse (R2) reads using the PEAR software (Zhang *et al*., 2014), with T-30 parameter. Reads which did not overlap (*unassembled*) were filtered using Clean Solexa software, to remove low-quality regions (Q<20). Finally, Q>20 assembled and unassembled reads were concatenated and submitted to transcriptome assembly as single-end reads (Fig. 6).

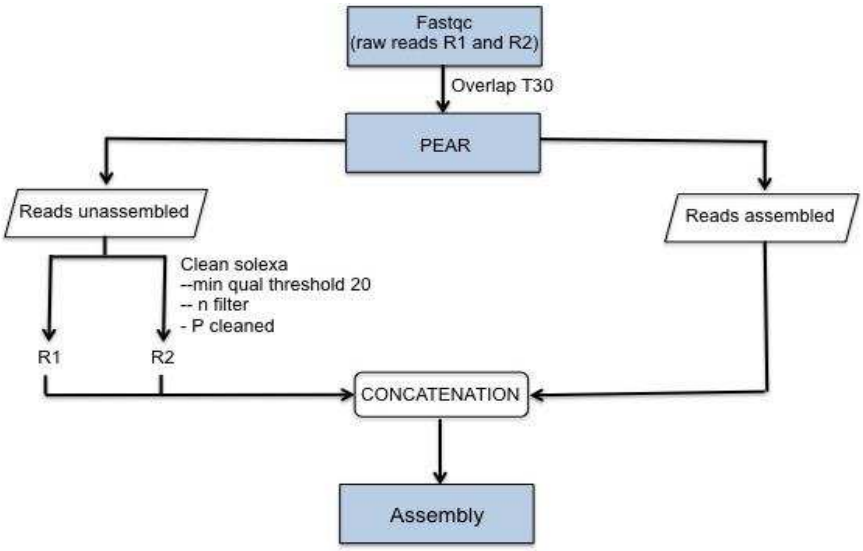


Fig. 6 Pipeline of filtering raw reads by quality using the PEAR and Clean Solexa software. R1 = *Read forward*; R2 = *Read reverse*

Transcriptome assembly

The 10 libraries were mapped against a reference genome. Despite *C. arabica* genome is not available to date, *C. canephora* had its genome recently made available (Denoeud *et al.*, 2014). *C. canephora* is a very closely related species to coffee, and it is believed that *C. arabica* is an interspecific hybrid from *C. canephora* and *C. eugenioides* (Lashermes *et al.*, 1999). Therefore, we used *C. canephora* as a reference genome. An index file of *C. canephora* was created with Bowtie2 version 2.2.5 software, and mapped against the set of reads with each library using Tophat2 version 2.0.13 (Trapnell *et al*., 2009). The assembly of each library was created with Cufflinks (Trapnell *et al*., 2010), pooled together with Cuffmerge software and annotated with Cuffcompare. For this latter process, we used the *C. canephora* GFF3 and CDS (*Coding DNA Sequence*) files, which contains structural and functional genes. The set of annotated transcripts was analyzed using RSEM (*RNA-Seq by Expectation Maximization*) version 1.2.20 software (Li and Dewey, 2011) to estimate the abundance of genes. The Fasta, GFF3 and CDS files of *C. canephora* genome were downloaded at http://coffee-genome.org/coffeacanephora.

The next step was to get reads that mapped against *H. vastatrix* and subtracted from unmapped reads against *C. canephora,* using a partial genome

of *H. vastatrix* (http://bioinformatics.cenicafe.org/index.php/wiki/CoffeeRustHybridDraftAssembly_Contigs). Additionally, the unmapped reads against *H. vastatrix* were subtracted from the mapped *C. canephora* (Fig. S3). The purpose of this strategy was to get reads unique to *C. arabica*. Based on the result, we produced a *de novo* assembly using Trinity software (Grabherr *et al*., 2011; Haas *et al*., 2013). Subsequently, the obtained transcripts were analyzed with the Transdecoder (cut ORFs >250pb) and compared with the non-redundant databases (-NR-NCBI) (*e-value* 1e-5) by BLASTX. In addition, transcripts were submitted to BLASTX against the annotated genomes of *Arabidopsis thaliana*, *Vitis vinifera* and *Solanum lycopersicum* (http://www.phytozome.net). Similarly, the set of annotated transcripts obtained by *de novo* methodology was analyzed using RSEM (*RNA-Seq by Expectation Maximization*) version 1.2.20 software (Li and Dewey, 2011) to estimate the abundance of genes.

Identification and annotation of differentially expressed genes

Read counting and normalization (FPKM, *fragments per kilobase of exon per million mapped reads*) obtained in previous steps were used to analyze differential expression of genes using two statistical packages (DESeq, Anders and Huber, 2012 and EDGER, Robinson *et al*., 2010) to identify differentially expressed genes at time points 12 hai, 24 hai, 96 hai and 17 dai compared to 0 hai (control) for each genotype separately. The list of differentially expressed genes was filtered by Log_2 *fold change* >=0.5 and Log_2 *fold change* <= -0.5, with *q* value of 0.1. To assess the variability between samples, principal component analysis (PCA) plot was performed using the statistical R package, with reads count table from RSEM. Cluster analysis was performed using Cluster 3.0 software (Hoon *et al*., 2004), starting from normalized read counts. Functional annotation related to biological process, molecular function and extracellular component was performed using BLAST2GO (Conesa *et al*., 2005).

Candidate gene selection and validation by real time qPCR

Only differentially expressed genes identified by both statistical packages (DESeq and EDGER) were used for quantitative real time PCR validation. Initially, the up- and down-regulated genes within each genotype and at all contrasting times were identified (Fig. 7). Within the intersection of each group, a new intersection between the up-regulated and down-regulated genes was done in susceptible and resistant cultivars for all the time points monitored. Then with the unique genes of the HdT (resistant) in each time, a last intersection was separated in order to select important genes related to plant defense (Fig. 7).

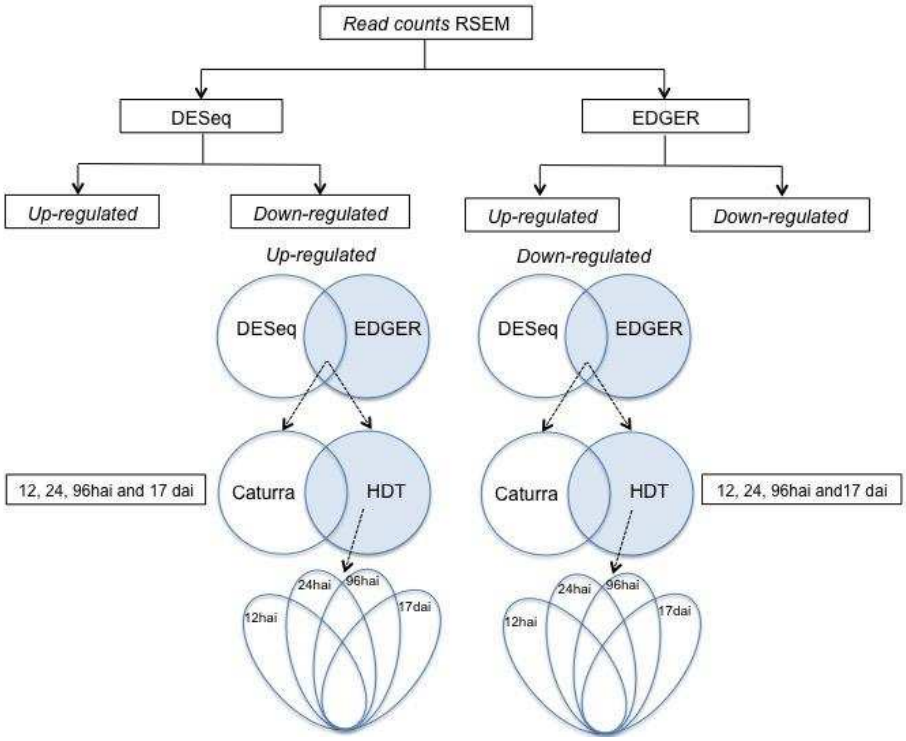


Fig. 7 Pipeline strategy to select candidate genes. HDT: Híbrido de Timor. hai: hours after inoculation. dai: days after inoculation.

The qPCR was conducted to compare the expression pattern of some selected genes during compatible (Caturra – *H. vastatrix*) and incompatible (Híbrido de Timor – *H. vastatrix*) interactions. Primers flanking the sequence of each candidate gene were designed (Table 1S) with *Primer-BLAST* software of NCBI (available at: http://www.ncbi.nlm.nih.gov/tools/primer-blast/) with the following parameters: Amplicon length between 90 and 150 bp, primer size: 20±

2 bp, annealing temperature (Ta) between 58℃ and 60℃, GC content of ± 50%. Reference genes used were *GADPH* (*Glyceraldehyde-3-phosphate dehydrogenase*), *UBQ10* (polyubiquitin 10) and *S24* (ribosomal protein) (Cruz *et al*., 2009), see Table 1S. Primer efficiency was tested by developing a standard curve of five dilution points of cDNA (1:5), and primer efficiency (E) was calculated from the slope (a) of standard curve ($E = 10^{(-1/a)} - 1$) with the Cts obtained for each dilution.

Using a different set of plants of those used for transcriptome study, total RNA was extracted from leaves of Caturra and Híbrido de Timor genotypes infected with *H. vastatrix* at 0, 12, 24 and 96 hai. Three biological and three technical replicates were performed for each sample. Reverse transcription was conducted using the ImProm-II™ Reverse Transcriptase system (*Promega, Madison, USA*), with 1µg of total RNA. The qPCR reactions were carried out in 7500 Real Time PCR System (Applied Biosystems, Thermo Fisher Scientific, USA), in final volume of 10µl with 50ng/µl of cDNA and 100 nM of forward and reverse primers in 1x *GoTaq® qPCR Master Mix* (*Promega, Madison, USA*) as final concentrations. Reaction parameters were 95℃ for 10 minutes followed by 40 cycles of 95℃ for 15 seconds and 60℃ for 1 minute and melting cu rve stage was set to default conditions. The expression levels of the candidate genes were calculated using qBase software (Hellemans *et al*., 2007).

Acknowledgments

We thank Jorge Badel for his helpful suggestions during the preparation of the manuscript. We are also grateful to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the financial support and the scholarship, DTI (Diretoria de Tecnologia da Informação) at Universidade Federal de Viçosa and LGE (Laborátorio de Genômica e Expressão) at Universidade Estadual de Campinas for providing bioinformatics facilities.

REFERENCES

Afzal, A.J., Wood, A.J. and Lightfoot, D.A. (2008) Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol. Plant. Microbe. Interact.* **21**, 507–517.

Anders, S. and Huber, W. (2012) Differential expression of RNA-Seq data at the gene level–the DESeq package. *EMBL, Heidelberg, Ger.* **11**, R106.

Avelino, J., Cristancho, M., Georgiou, S., et al. (2015) The coffee rust crises in Colombia and Central America (2008–2013): impacts, plausible causes and proposed solutions. *Food Secur.* **7**, 303–321.

Balaji, V., Mayrose, M., Sherf, O., et al. (2008) Tomato Transcriptional Changes in Response to Clavibacter michiganensis subsp. michiganensis Reveal a Role for Ethylene in Disease Development. *Plant Physiol.* **146**, 1797–1809.

Bettencourt, A.J. (1973) Considerações gerais sobre o "Hibrido de Timor." *Circ. No 23*, 20.

Bettencourt, A. J and Rodrigues, C. (1988) Principles and practice of coffee breeding for resistance to rust and other diseases. *Elsevier Appl. Sci.* **3**, 199–234.

Bettgenhaeuser, J., Gilbert, B., Ayliffe, M. and Moscou, M.J. (2014) Nonhost resistance to rust pathogens - a continuation of continua. *Front. Plant Sci.* **5**.

Boller, T. and Felix, G. (2009) A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annu. Rev. Plant Biol.* **60**, 379–406.

Boyd, L.A., Ridout, C., O'Sullivan, D.M., Leach, J.E. and Leung, H. (2013) Plant-pathogen interactions: Disease resistance in modern agriculture. *Trends Genet.* **29**, 233–240.

Cabral, P. G. C., Zambolim, E. M., Zambolim, L., Lelis T. P., Capucho, A.S. and and Caixeta, E.T. (2009) Identification of a new race of *Hemileia vastatrix* in Brazil. *Australas Plant Dis*. *Notes* **4**, 129–130.

Capucho, A.S., Zambolim, E.M., Freitas, R.L., Haddad, F., Caixeta, E.T. and Zambolim, L. (2012) Identification of race XXXIII of *Hemileia vastatrix* on *Coffea arabica* Catimor derivatives in Brazil. *Australas. Plant Dis. Notes* **7**, 189–191.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

Cristancho, M.A., Botero-Rozo, D.O., Giraldo, W., et al. (2014) Annotation of a hybrid partial genome of the coffee rust (*Hemileia vastatrix*) contributes to the gene repertoire catalog of the Pucciniales. *Front. Plant Sci.* **5**, 594.

Cruz, F., Kalaoun, S., Nobile, P., et al. (2009) Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. *Mol. Breed.* **23**, 607–616.

Davis, A.P., Tosh, J., Ruch, N. and Fay, M.F. (2011) Growing coffee: Psilanthus (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of Coffea. *Bot. J. Linn. Soc.* **167**, 357–377.

Denoeud, F., Carretero-Paulet, L., Dereeper, A., et al. (2014) The coffee

genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*. **345**, 1181–1184.

Diniz, I., Talhinhas, P., Azinheira, H.G., et al. (2012) Cellular and molecular analyses of coffee resistance to *Hemileia vastatrix* and nonhost resistance to *Uromyces vignae* in the resistance-donor genotype HDT832/2. *Eur. J. Plant Pathol.* **133**, 141–157.

Duplan, V. and Rivas, S. (2014) E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. *Front. Plant Sci.* **5**.

Ebel J, S.D. (1997) Signals in host-parasite interactions. In The Mycota., pp. 85–105.

Fernandez, D., Santos, P., Agostini, C., et al. (2004) Coffee (*Coffea arabica* L.) genes early expressed during infection by the rust fungus (*Hemileia vastatrix*). *Mol. Plant Pathol.* **5**, 527–536.

Fernandez, D., Tisserant, E., Talhinhas, P., et al. (2012) 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant-rust interaction. *Mol. Plant Pathol.* **13**, 17–37.

Flor, H. (1942) Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* **32**, 653–669.

Freitas, R.L., Zambolim, E., Silva, M.C., Caixeta, E.T., Lelis, D.T., Zambolim, L., Sakiyama, N. (2014) Cytological evaluation of the infection process of *Hemileia vastatrix* (race XXXIII) in resistant and susceptible coffee. In proceedings of the 25th International conference on coffee science (ASIC). pp. 42–46. Armenia, Colombia.

Freitas, R. L (2015) Análise citológica e perfil de expressão gênica de *Hemileia vastatrix* (raça xxxiii) na interação com o cafeeiro. Universidade Federal de Viçosa. pp. 21-22.

Ganesh, D., Petitot, A.S., Silva, M.C., Alary, R., Lecouls, A.C. and Fernandez, D. (2006) 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.* **170**, 1045–1051.

Gichuru, E.K., Ithiru, J.M., Silva, M.C., Pereira, A.P. and Varzea, V.M.P. (2012) Additional physiological races of coffee leaf rust (*Hemileia vastatrix*) identified in Kenya. *Trop. Plant Pathol.* **37**, 424–427.

Gill, U.S., Lee, S. and Mysore, K.S. (2015) Host Versus Nonhost Resistance: Distinct Wars with Similar Arsenals. *Phytopathology* **105**, 580–587.

Grabherr, M.G., Haas, B.J., Yassour, M., et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–52.

Grenville-Briggs, L.J. and West, P. van (2005) The Biotrophic Stages of Oomycete–Plant Interactions. *Adv. Appl. Microbiol*. **57**, 217–243.

Guerra-Guimarães, L., Tenente, R., Pinheiro, C., et al. (2015) Proteomic analysis of apoplastic fluid of *Coffea arabica* leaves highlights novel biomarkers for resistance against *Hemileia vastatrix*. *Front. Plant Sci.* **6**, 1–16.

Haas, B.J., Papanicolaou, A., Yassour, M., et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity plataform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512.

Hammond-Kosack, K.E. and Jones, J.D. (1997) Plant disease resistance genes. *Annu. Rev. Plant Physiol. Plant Mol. Biol* **48**, 575–607.

Heath, M.C. (1977) A comparative study of non-host interactions with rust fungi. *Physiol. Plant Pathol.* **10**, 73–88.

Hellemans, J., Mortier, G., Paepe, A. De, Speleman, F. and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **8**, R19.

Hoon, M.J.L. de, Imoto, S., Nolan, J. and Miyano, S. (2004) Open source clustering software. *Bioinformatics* **20**, 1453–1454.

Ishiga, Y., Upplapapti, S.R. and Mysore, K.S. (2013) Expression analysis reveals a role for hydrophobic or epicuticular wax signals in pre-penetration structure formation of *Phakopsora pachyrhizi*. *Plant Signal. Behav.* **8**, e26959.

Jeandet, P., Clement, C., Courot, E., Cordelier, S. (2013) Modulation of phytoalexin biosyn- thesis in engineered plants for disease resistance. *Int. J. Mol. Sci.* **14**, 14136–14170.

Jones, J.D.G. and Dangl, J.L. (2006) The plant immune system. *Nature* **444**, 323–329.

Kenn, N.T. (1999) Plant Disease Resistance: Progress in basic understanding and practical application. In Advances in Botanical Research., pp. 291–328.

Kumar, K.R.R. and Kirti, P.B. (2011) Differential gene expression in Arachis diogoi upon interaction with peanut late leaf spot pathogen, *Phaeoisariopsis personata* and characterization of a pathogen induced cyclophilin. *Plant Mol. Biol.* **75**, 497–513.

Kundu, A., Patel, A., Paul, S. and Pal, A. (2015) Transcript Dynamics at Early Stages of Molecular Interactions of MYMIV with Resistant and Susceptible Genotypes of the Leguminous Host, Vigna mungo Hong, J.K., ed. *PLoS One* **10**, e0124687.

Kushalappa, A.C., Yogendra, K.N. and Karre, S. (2016) Plant Innate Immune Response: Qualitative and Quantitative Resistance. *CRC. Crit. Rev. Plant Sci.* **35**, 38–55.

Labaj, P.P., Leparc, G.G., Linggi, B.E., Markillie, L.M., Wiley, H.S. and Kreil, D.P. (2011) Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling. *Bioinformatics* **27**, i383–i391.

Lara-Ávila, J.P., Isordia-Jasso, M.I., Castillo-Collazo, R., Simpson, J. and Alpuche-Solís, Á.G. (2012) Gene Expression Analysis during Interaction of Tomato and Related Wild Species with Clavibacter michiganensis subsp. michiganensis. *Plant Mol. Biol. Report.* **30**, 498–511.

Lashermes, P., Combes, M.-C., Robert, J., Trouslot, P., D'Hont, A., Anthony, F. and Charrier, A. (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Mol. Gen. Genet. MGG* **261**, 259–266.

Li, B. and Dewey, C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323.

Licausi, F., Ohme-Takagi, M. and Perata, P. (2013) APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *New Phytol.* **199**, 639–649.

Loman, N.J., Misra, R. V, Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J. and Pallen, M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.* **30**, 434–439.

Lu, H. (2003) ACD6, a Novel Ankyrin Protein, Is a Regulator and an Effector of Salicylic Acid Signaling in the Arabidopsis Defense Response. *Plant cell online* **15**, 2408–2420.

Maia, T., Badel, J.L., Marin-Ramirez, G., Rocha, C. de M., Fernandes, M.B., Silva, J.C.F. da, Azevedo-Junior, G.M. de and Brommonschenkel, S.H. (2016) The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the *SH1* rust resistance gene. *New Phytol.* **213**, 1315-1329.

McCook, S. and Vandermeer, J. (2015) The Big Rust and the Red Queen: Long-Term Perspectives on Coffee Rust Research. *Phytopathology* **105**, 1164–1173.

Mellersh, D.G. and Heath, M.C. (2003) An Investigation into the Involvement of Defense Signaling Pathways in Components of the Nonhost Resistance of *Arabidopsis thaliana* to Rust Fungi Also Reveals a Model System for Studying Rust Fungal Compatibility. *Mol. Plant-Microbe Interact.* **16**, 398–404.

Mofatto, L.S., Carneiro, F. de A., Vieira, N.G., et al. (2016) Identification of candidate genes for drought tolerance in coffee by high-throughput sequencing in the shoot apex of different *Coffea arabica* cultivars. *BMC Plant Biol.* **16**, 94.

Narusaka M, Kubo Y, Hatakeyama K, Imamura J, Ezura H, Nanasato Y, Tabei Y, Takano Y, Shirasu K, N.Y. (2013) Interfamily transfer of dual NB-LRR genes confers resistance to multiple pathogens. *PLoS One* **8**, e55954.

Niks, R.E., Qi, X. and Marcel, T.C. (2015) Quantitative Resistance to Biotrophic Filamentous Plant Pathogens: Concepts, Misconceptions, and Mechanisms. *Annu. Rev. Phytopathol.* **53**, 445–470.

Niks, R.E. and Rubiales, D. (2002) Potentially durable resistance mechanisms in plants to specialised fungal pathogens. *Euphytica* **124**, 201–216.

Nirmala, J., Drader, T., Chen, X., Steffenson, B. and Kleinhofs, A. (2010) Stem Rust Spores Elicit Rapid RPG1 Phosphorylation. *Mol. Plant-Microbe Interact.* **23**, 1635–1642.

Oliveira, M.D.M., Varanda, C.M.R. and Félix, M.R.F. (2016) Induced resistance during the interaction pathogen x plant and the use of resistance inducers. *Phytochem. Lett.* **15**, 152–158.

Ozsolak, F and Milos, P. (2010) RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87–98.

Peck, S.C. (2003) Early phosphorylation events in biotic stress. *Curr. Opin. Plant Biol.* **6**, 334–338.

Pedras, M.S.C., Yaya, E.E. and Glawischnig, E. (2011) The phytoalexins from cultivated and wild crucifers: Chemistry and biology. *Nat. Prod. Rep.* **28**, 1381.

Pereira, A. A; Zambolim, L; Chaves, G.M; Sakiyama, N.S. (2000) Cultivar de café resistente à Ferrugem: Oeiras-MG 6851. *Rev. Ceres* **46**, 121–124.

Ramiro, D.A., Escoute, J., Petitot, A.S., Nicole, M., Maluf, M.P. and Fernandez, D. (2009) Biphasic haustorial differentiation of coffee rust (*Hemileia vastatrix* race II) associated with defence responses in resistant and susceptible coffee cultivars. *Plant Pathol.* **58**, 944–955.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene

expression data. *Bioinformatics* **26**, 139–140.
Rodrigues, C.J., Bettencourt, A.J. and Rijo, L. (1975) Races of the Pathogen and Resistance to Coffee Rust. *Annu. Rev. Phytopathol.* **13**, 49–70.
Rodrigues Júnior, C.J., Gonçalves, M.M. and V. and V.M.P (2004) Importância do Híbrido de Timor para o território e para o melhoramento da cafeicultura mundial. *Rev. Ciências Agrárias* **27**, 203–213.
Rodriguez, C., Varzea, I., Godinho, S.P. y R. and R (1993) New physiologic races of Hemileia vastatrix. In 15° Colloque Scientifique International sur le café., pp. 318–321. Montpellier, France.
Rubio, M., Rodríguez-Moreno, L., Ballester, A.R., Moura, M.C. de, Bonghi, C., Candresse, T. and Martínez-Gómez, P. (2015) Analysis of gene expression changes in peach leaves in response to Plum pox virus infection using RNA-Seq. *Mol. Plant Pathol.* **16**, 164–176.
Schulze-Lefert, P. and Panstruga, R. (2003) Establishment of biotrophy by parasitic fungi and reprogramming of host cells for disease resistance. *Annu. Rev. Phytopathol.* **41**, 641–667.
Sharma, M. and Pandey, G.K. (2016) Expansion and Function of Repeat Domain Proteins During Stress and Development in Plants. *Front. Plant Sci.* **6**, 1218.
Shibuya, N. and Minami, E. (2001) Oligosaccharide signalling for defence responses in plant. *Physiol. Mol. Plant Pathol.* **59**, 223–233.
Silva, M.C., Nicole, M., Guerra-Guimarães, L. and Rodrigues, C.J. (2002) Hypersensitive cell death and post-haustorial defence responses arrest the orange rust (*Hemileia vastatrix*) growth in resistant coffee leaves. *Physiol. Mol. Plant Pathol.* **60**, 169–183.
Silva, M. do C., Várzea, V., Guerra-Guimarães, L., Azinheira, H.G., Fernandez, D., Petitot, A.-S., Bertrand, B., Lashermes, P. and Nicole, M. (2006) Coffee resistance to the main diseases: leaf rust and coffee berry disease. *Brazilian J. Plant Physiol.* **18**, 119–147.
Takahashi, S., Yeo, Y.-S., Zhao, Y., O'Maille, P.E., Greenhagen, B.T., Noel, J.P., Coates, R.M. and Chappell, J. (2007) Functional Characterization of Premnaspirodiene Oxygenase, a Cytochrome P450 Catalyzing Regio- and Stereo-specific Hydroxylations of Diverse Sesquiterpene Substrates. *J. Biol. Chem.* **282**, 31744–31754.
Talhinhas, P., Batista, D., Diniz, I., et al. (2016) The Coffee Leaf Rust pathogen *Hemileia vastatrix* : One and a half centuries around the tropics. *Mol. Plant Pathol.*
Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Baren, M.J. van, Salzberg, S.L., Wold, B.J. and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515.
Várzea, V. M. P; Marques, D. V. (2005) Population variability of *Hemileia vastatrix* vs coffee durable resistance. In: Zambolim L, Zambolim E, Várzea VMP (eds) Durable resistance to coffee leaf rust. UFV Press, Viçosa-Brazil, pp 53–74.
Voegele, R.T. and Mendgen, K. (2003) Rust haustoria: nutrient uptake and beyond. *New Phytol.* **159**, 93–100.

Vossen, H. van der, Bertrand, B. and Charrier, A. (2015) Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. *Euphytica* **204**, 243–256.

Wang, X., McCallum, B.D., Fetch, T., Bakkeren, G. and Saville, B.J. (2015) *Sr36* - and *Sr5* -Mediated Resistance Response to *Puccinia graminis* f. sp. tritici Is Associated with Callose Deposition in Wheat Guard Cells. *Phytopathology* **105**, 728–737.

Wang, Y.-S., Pi, L.-Y., Chen, X., et al. (2006) Rice *XA21* Binding Protein 3 Is a Ubiquitin Ligase Required for Full Xa21-Mediated Disease Resistance. *PLANT CELL ONLINE* **18**, 3635–3646.

Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63.

Westermann, A.J., Gorski, S. a. and Vogel, J. (2012) Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630.

Wit, P.J.G.M. de, Mehrabi, R., Burg, H.A. Van Den and Stergiopoulos, I. (2009) Fungal effector proteins: past, present and future. *Mol. Plant Pathol.* **10**, 735–747.

Xin, M., Wang, X., Peng, H., Yao, Y., Xie, C., Han, Y., Ni, Z. and Sun, Q. (2012) Transcriptome Comparison of Susceptible and Resistant Wheat in Response to Powdery Mildew Infection. *Genomics. Proteomics Bioinformatics* **10**, 94–106.

Young, S. a, Guo, a, Guikema, J. a, White, F.F. and Leach, J.E. (1995) Rice cationic peroxidase accumulates in xylem vessels during incompatible interactions with *Xanthomonas oryzae* pv oryzae. *Plant Physiol.* **107**, 1333–1341.

Yuyama, P.M., Reis Junior, O., Ivamoto, S.T., Domingues, D.S., Carazzolle, M.F., Pereira, G.A., Charmetant, P., Leroy, T. and and Pereira, L.P. (2016) Transcriptome analysis in *Coffea eugenioides*, an Arabica coffee ancestor, reveals differentially expressed genes in leaves and fruits. *Mol. Genet. Genomics* **291**, 323–336.

Zambolim, L. (2016) Current status and management of coffee leaf rust in Brazil. *Trop. Plant Pathol.* **41**, 1–8.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614–620.

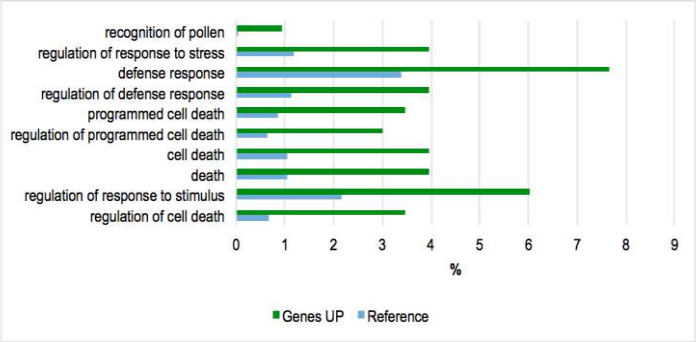
.

SUPPLEMENTAL MATERIAL

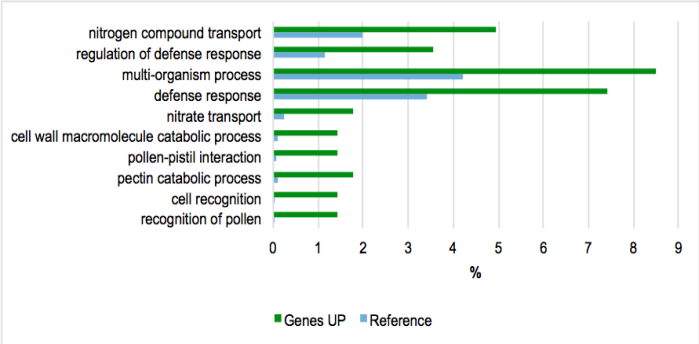
Supplemental Fig. 1 12 and 24 hai up-regulated genes annotation in susceptible and resistant genotypes by BLAST2GO, only top 10 GO classes are shown with the 3 main categories: Biological Process, Molecular Function and Cellular Component.

Enriched GO - Biological Process 12 hai

Susceptible

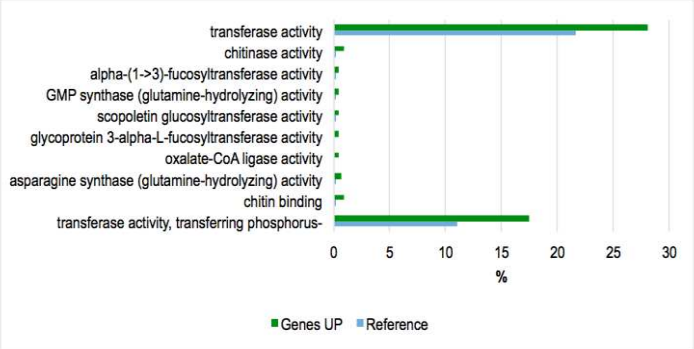


Resistant

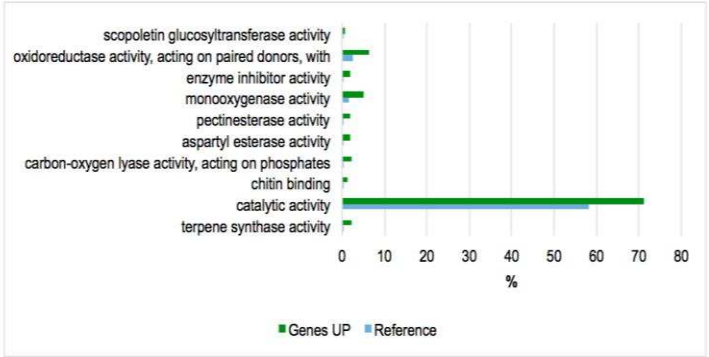


Enriched GO - Molecular Function 12 hai

Susceptible

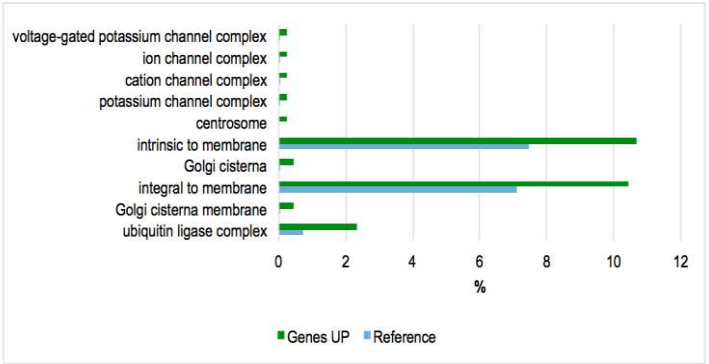


Resistant

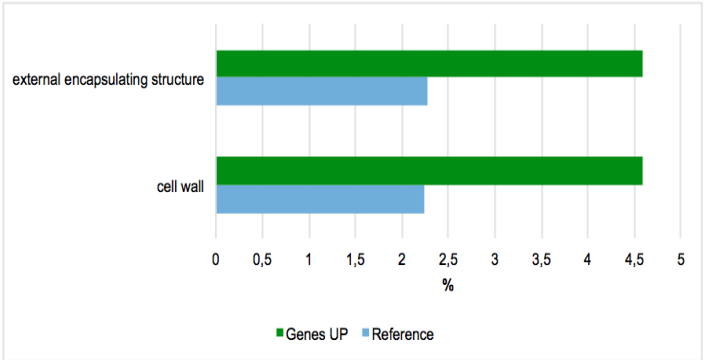


Enriched GO - Cellular component 12 hai

Susceptible

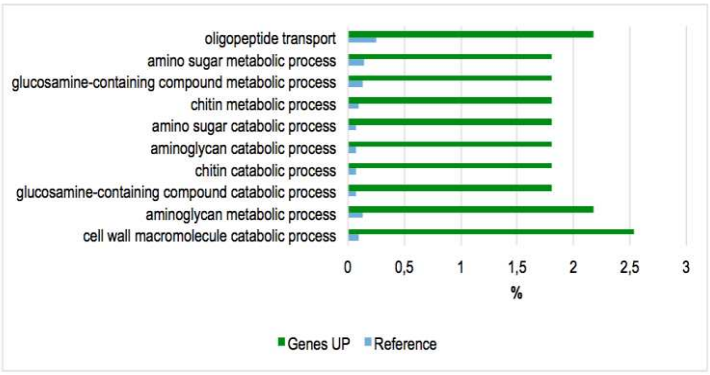


Resistant

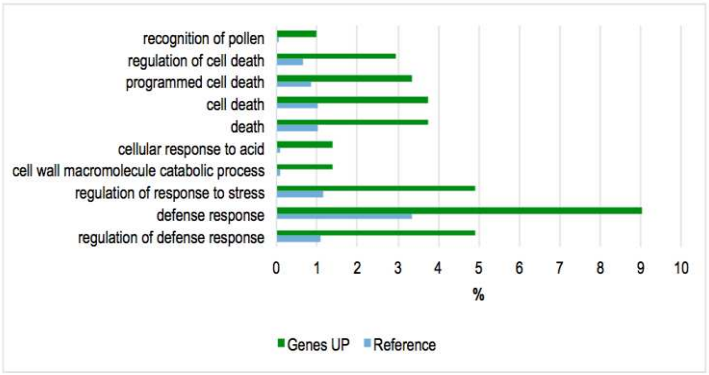


Enriched GO - Biological Process 24 hai

Susceptible

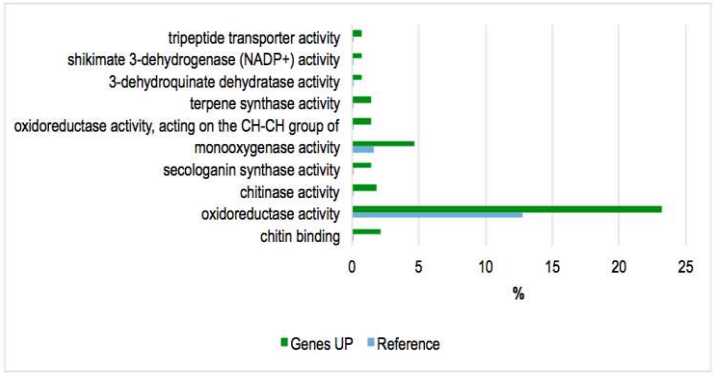


Resistant

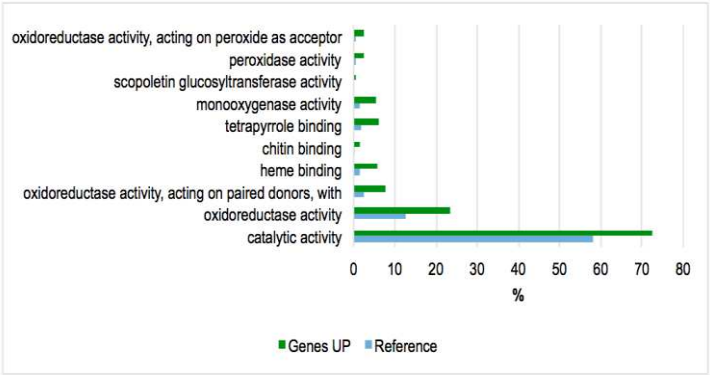


Enriched GO - Molecular Function 24 hai

Susceptible

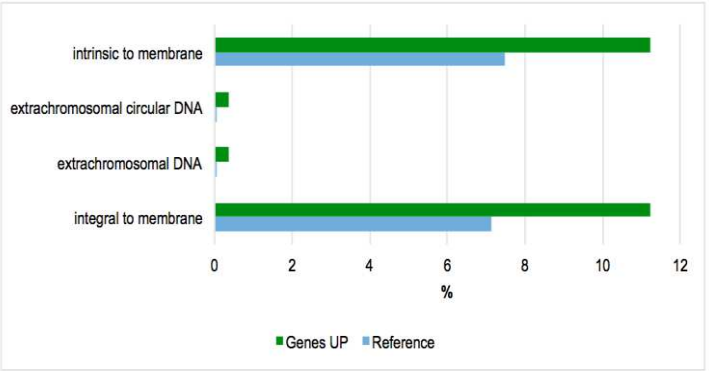


Resistant

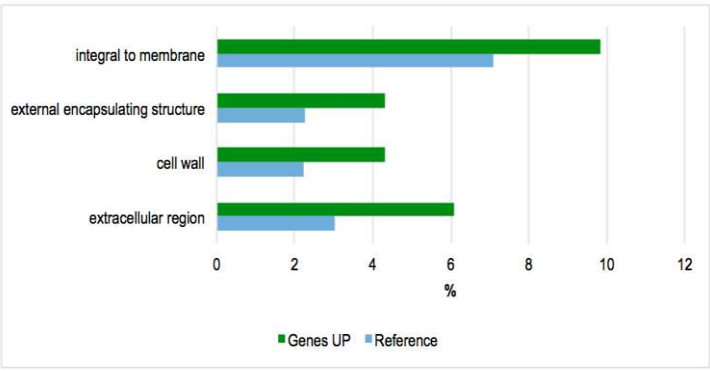


Enriched GO - Cellular component 24 hai

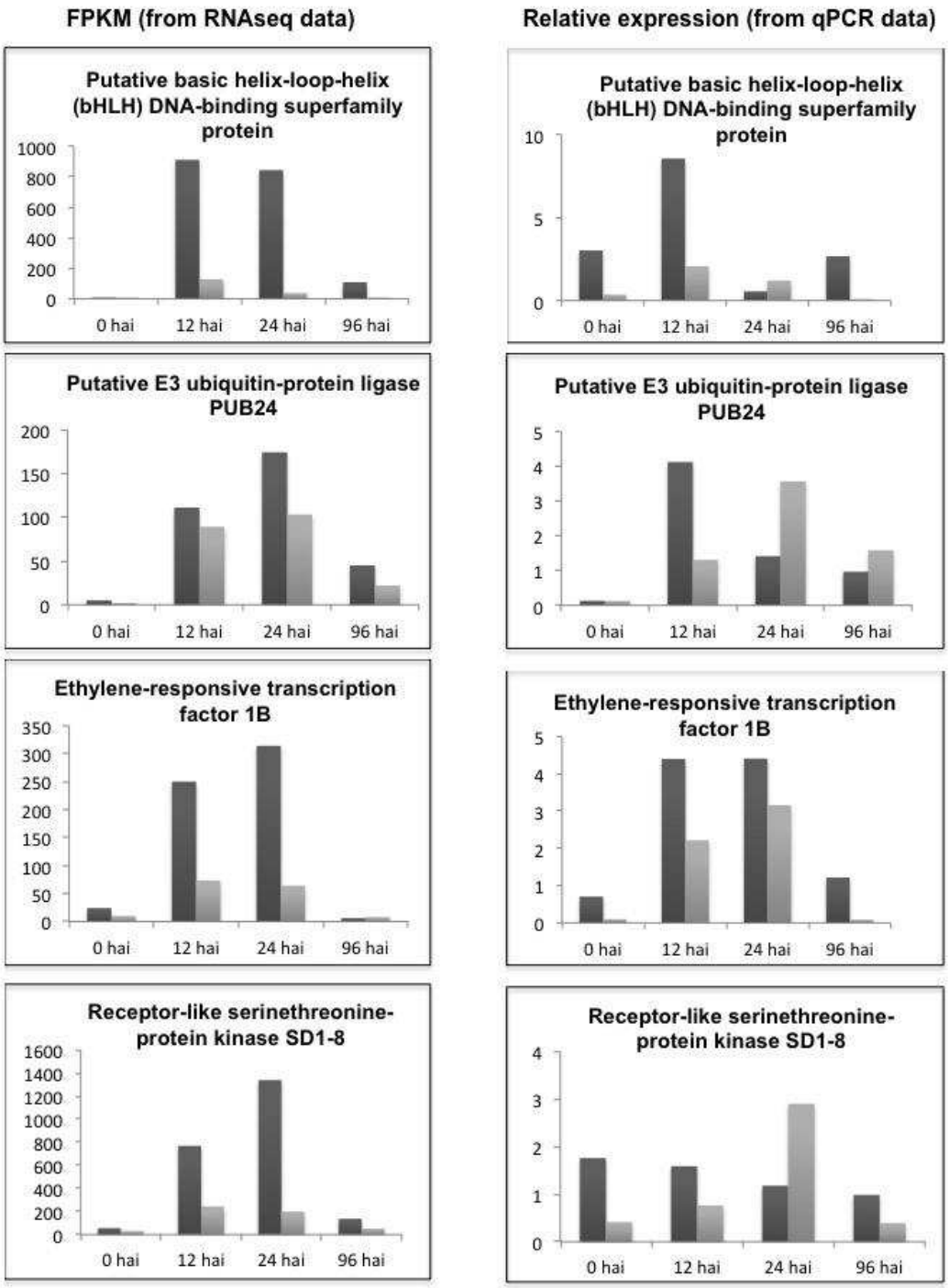
Susceptible

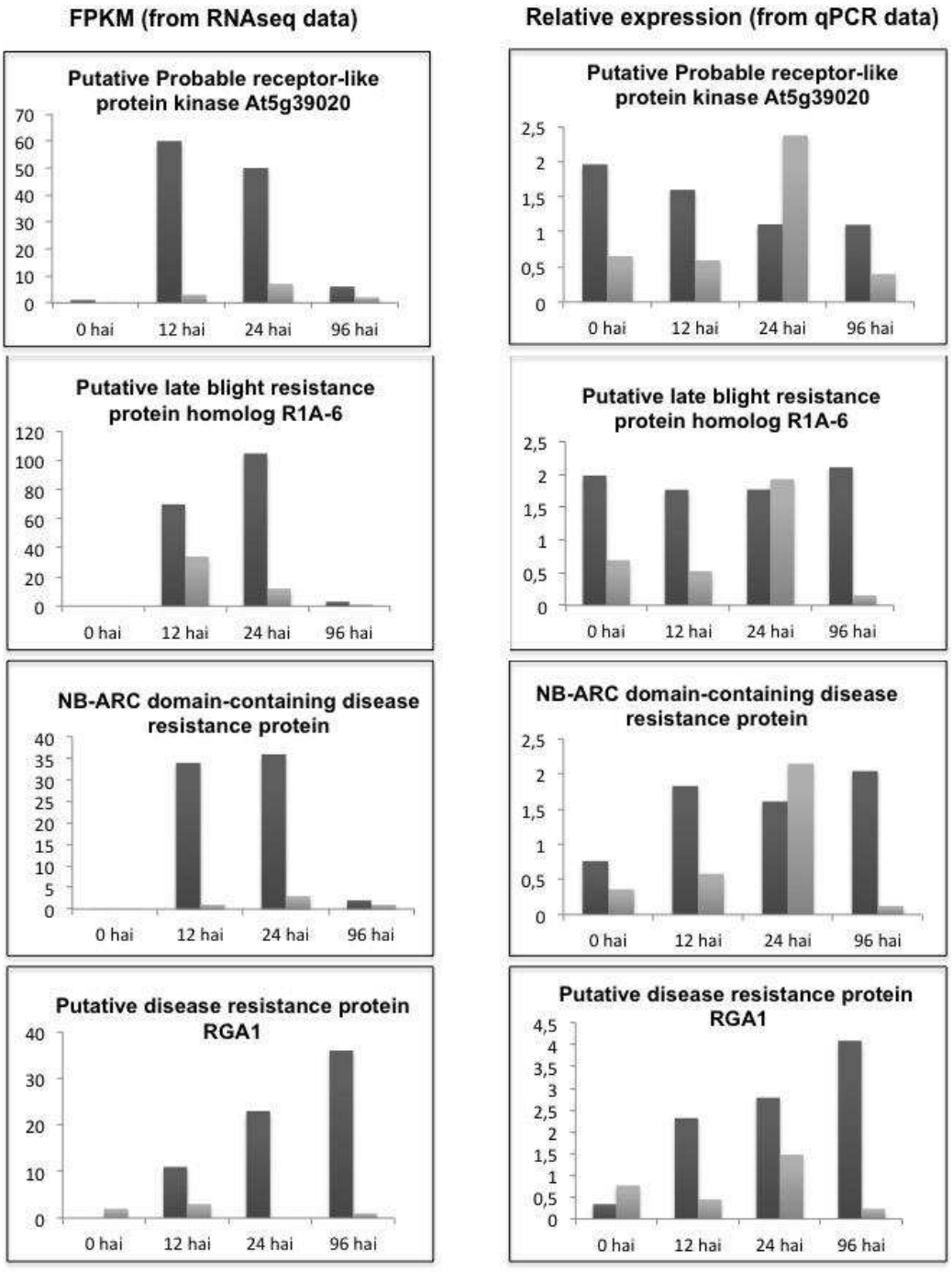


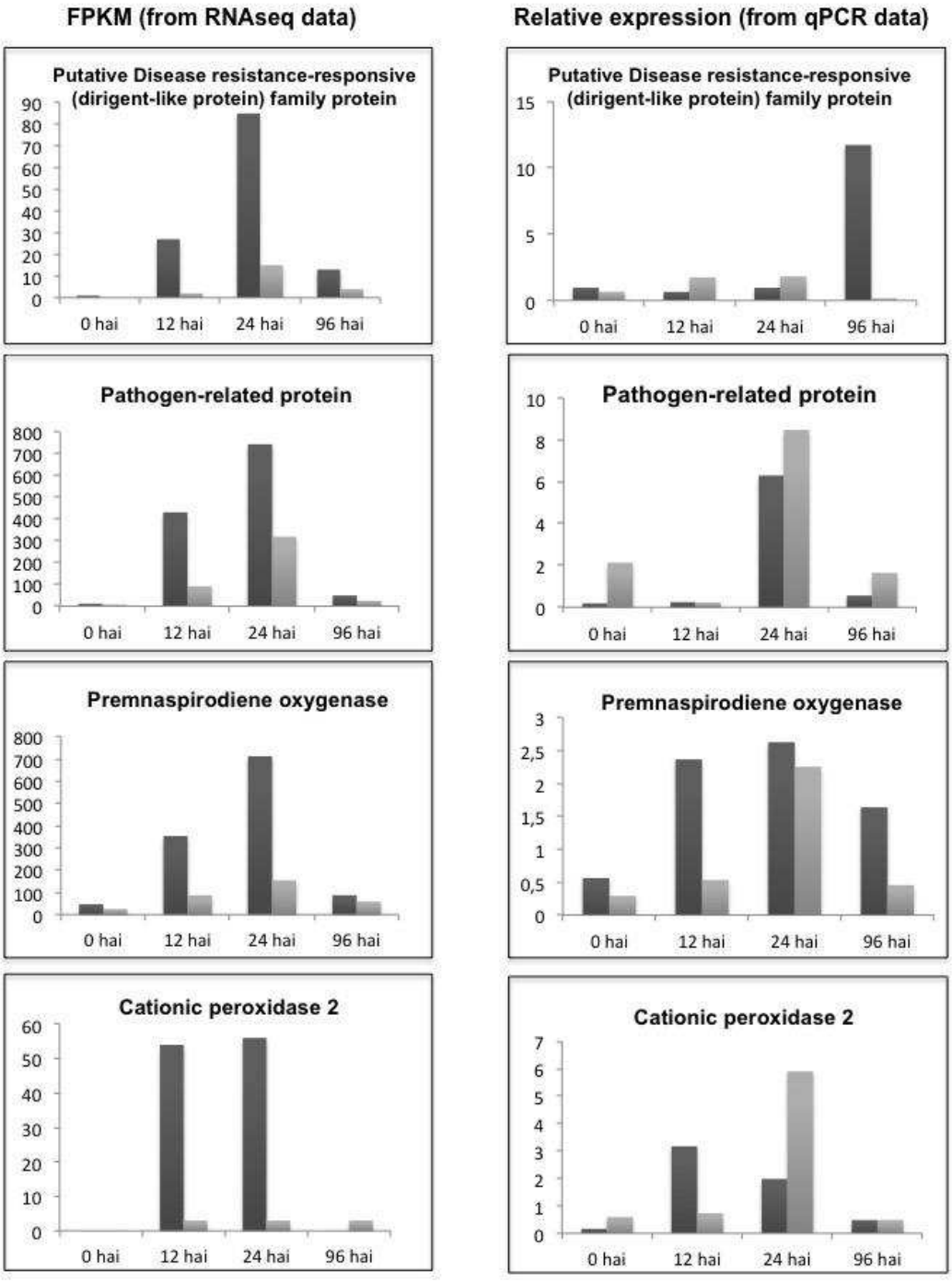
Resistant

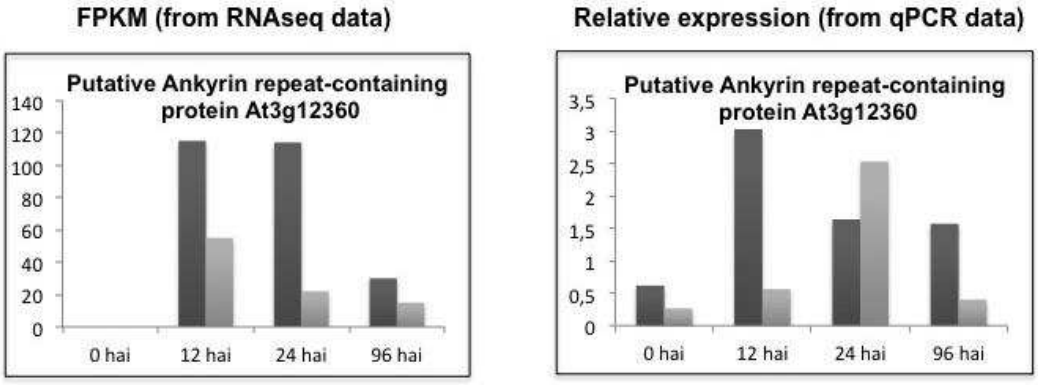


Supplemental Fig. 2 Correlation between transcriptome profiling and qPCR



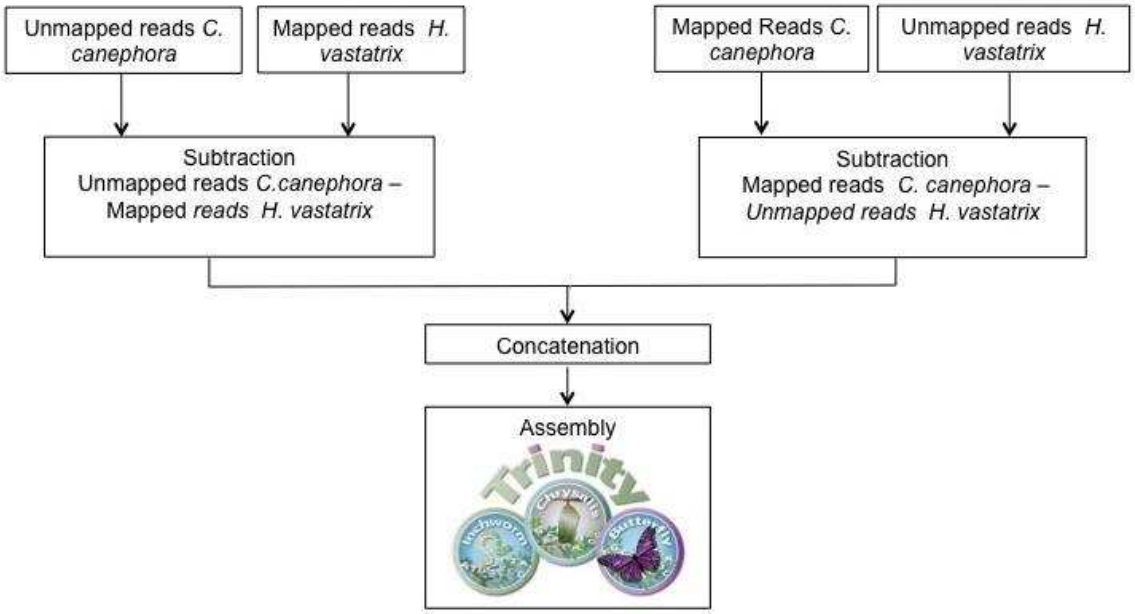

FPKM (from RNAseq data)
Relative expression (from qPCR data)
Putative Probable receptor-like protein kinase At5g39020
Putative late blight resistance protein homolog R1A-6
NB-ARC domain-containing disease resistance protein
Putative disease resistance protein RGA1
0 hai
12 hai
24 hai
96 hai


FPKM (from RNAseq data)
Relative expression (from qPCR data)
Putative Disease resistance-responsive (dirigent-like protein) family protein
Pathogen-related protein
Premnaspirodiene oxygenase
Cationic peroxidase 2
0 hai
12 hai
24 hai
96 hai



Dark gray: Resistant genotype
Light gray: Susceptible genotype

Supplemental Fig. 3 Strategy used for *de novo* assembly



Supplemental Table 1 Sequences of primers used for qPCR experiments

Gene	Name Sequence Primer	Annealing temperature (℃)	Product size (bp)	Publication
Receptor-like serinethreonine-protein kinase SD1-8	RLSKsd1 F: 5'- ACTCGGGGCAAGGAATAGAAGA-3'	60.8	92	This study
	RLSKsd1 R: 5'-GCAACAACCTAAAAGGCGAACTAAA -3'	60.7		
Putative late blight resistance protein homolog R1A-6	R1A-6 F: 5'-TCCGAAGGTTGTTATGGCTTTGG -3'	61.3	150	This study
	R1A-6 R: 5'- TGGCACCACTCGTAGTTCTTTGT-3'	62.3		
Putative E3 ubiquitin-protein ligase PUB24	E3ubiq F: 5'- AAATTCTCGGGTACAGTTGGGG-3'	60.2	112	This study
	E3ubiq R: 5'-TTGCCTTATCCTTGAGGTGCGA -3'	62.0		
Putative Probable receptor-like protein kinase At5g39020	RLKat5 F: 5'-GATACATGGCTCCTGAGTTGTTCT -3'	60.3	119	This study
	RLKat5 R: 5'-TGCATTTACATTTCTCCTCCTTCCT -3'	60.2		
Premnaspirodiene oxygenase	Premna F: 5'-ACGGGAAAGAGGACCATTGAAGA -3'	61.5	105	This study
	Premna R: 5'-AACTGATAAAGGGGCAGGAGGA -3'	60.8		
NB-ARC domain-containing disease resistance protein	NB-ARC F: 5'- ACGGGGAATTGTCGAAGGTGTT -3'	62.4	111	This study
	NB-ARC R: 5'- ATGCAGGGATTTCATGGTCCTC -3'	60.4		
Ethylene-responsive transcription factor 1B	Eth F: 5'- CCCTTCGTGATATGAAATGCGGT-3'	61.3	135	This study
	Eth R: 5'- CCGCCTCTCTTGCAATTTTGTG-3'	60.9		
Cationic peroxidase 2	Perox F: 5'-TTGGGGGACATACGATTGGAAC -3'	60.3	111	This study
	Perox R: 5'- GAGAAAGAAAAGAGGGACTGATGGA-3'	60.2		
Putative disease resistance protein RGA1	RGA1 F: 5'-TCCTTGTTCTTGATGATGTGTGGA -3'	60.2	124	This study
	RGA1 R: 5'-ACTGGTTGTAGACGAGTAGTGAGA -3'	60.5		
Putative Ankyrin repeat-containing protein At3g12360	Anky F: 5'-GGAAGGAACCCTCTTCATGTTGCT -3'	62.8	112	This study
	Anky R: 5'- TGGTCTCCCCACGTCTAGTCTTTT -3'	63.0		
Putative basic helix-loop-helix (bHLH) DNA-binding superfamily protein	bHLH F: 5'-CATATTCAGGCGCTACAGGTGAGA -3'	62.2	102	This study
	bHLH R: 5'- GTTCTTCCAGCCCATTTAGCAGA -3'	60.8		

Putative Disease resistance-responsive (dirigent-like protein) family protein	DRR F: 5'- CGGTCGAAGCCAAGGTCTTTATGT-3' DRR R: 5'- TCCAATGTGCTACCGTTGTATTGT-3'	63.1 60.8	108	This study
Pathogen-related protein	PRP F: 5'-TCTAGGCATCAACGGGAGGAAA -3' PRP R: 5'- TTCGCAGGATTGTAAACCCGAA-3'	61.1 60.5	120	This study
Coffea arabica cDNA clone CA00-XX-LV5-041-E04-QH, mRNA sequence	GAPDH F: 5'-AGGCTGTTGGGAAAGTTCTTC-3' GAPDH R: 5'- ACTGTTGGAACTCGGAATGC-3'	63.4 64.0	70	Cruz *et al*., 2009
S24 Ribosomal protein	S24 F 5'- GCCCAAATATCGGCTTATCA– 3' S24 R 5'- TCTTCTTGGCCCTGTTCTTC– 3'	63.5 63.3	92	Cruz *et al*., 2009
UBQ10 Poliubiquitin 10	UBQ10 F 5 5'- CAGACCAGCAGAGGCTGATT– 3' UBQ10 R 5 5'- AGAACCAAGTGAAGGGTGGA– 3'	64.6 63.5	100	Cruz *et a*l., 2009