

IDENTIFICAÇÃO DE PROVÁVEIS GENES *R* CLASSE 1 E 2 DE *Coffea arabica* NO BANCO BRASILEIRO GENOMA FUNCIONAL DE CAFÉ (CafEST)

Marília S. SILVA¹, E-mail: marilia@cpac.embrapa.br; Érika V. S. ALBUQUERQUE²; Cristiane C. TEIXEIRA²; Magnólia A. CAMPOS³; Ângela MEHTA²; Natália F. MARTINS²; Maria Fátima GROSSI-DE-SÁ²

¹Embrapa Cerrados, Planaltina-DF; ²Embrapa Recursos Genéticos e Biotecnologia, Brasília-DF; ³Universidade Federal de Lavras, Lavras-MG.

Resumo:

O café é um produto agrícola tradicional da economia brasileira que representa 2,4% do valor total das exportações e gera cerca de 8 milhões de empregos. O Brasil produz aproximadamente 40% do café comercializado no mercado internacional, além de ser o segundo maior consumidor mundial desse produto, particularmente da espécie *Coffea arabica*. Apesar da produção e consumo de *C. arabica* corresponder à cerca de 70% do mercado de café, essa espécie é altamente susceptível a nematóides, pragas e doenças. Portanto, há um crescente interesse de programas de melhoramento genético de cafeeiro em desenvolver variedades de *C. arabica* com resistência a nematóides, pragas e doenças. Um grande número de genes de resistência (genes *R*) de plantas já foi isolado e classificado em seis grupos denominados classe 1-classe 6. A maioria dos genes *R* conhecidos pertence à classe 2 e codifica proteínas que contêm os seguintes domínios em sua sequência: *nucleotide binding site* (NBS), *leucine-rich repeat* (LRR) e, no N-terminal, um domínio *leucine-zipper* (LZ) ou outra sequência *coiled-coil* (CC). A classe 1 compreende genes *R* homólogos ao membro tipo dessa classe, que é o gene *R* denominado *pto*, o qual codifica uma proteína que apresenta o domínio catalítico de serina/treonina quinase e regiões de miristilação em sua porção N-terminal. Sequências aminoacídicas bem descritas correspondentes a genes *R* das classes 1 e 2 foram usadas para rastrear o Banco Brasileiro *Genoma Funcional de Café* (CafEST) por sequências de ESTs homólogas de genes *R* classes 1 e 2 em *C. arabica*. Os ESTs selecionados nessa busca foram agrupados em *clusters* (*contigs* ou *singlets*), os quais foram subsequentemente analisados quanto a homologia com genes *R* disponíveis em banco de dados públicos. Alinhamentos múltiplos entre sequências de aminoácidos deduzidas a partir das sequências consenso dos *contigs* do tipo genes *R* do CafEST e sequências homólogas foram usados para gerar filogramas. Os filogramas gerados indicam ambas as classes 1 e 2 de prováveis genes *R* de *C. arabica* estão consideravelmente representadas no CafEST, um vez que alto número total de ESTs foi recuperado quando do rastreamento desse banco, ou seja, 525 ESTs homólogos a genes *R* classe 1, agrupados em 262 *clusters* (118 *contigs* e 144 *singlets*), e 449 ESTs homólogos a genes *R* classe 2 agrupados em 190 *clusters* (82 *contigs* e 108 *singlets*), perfazendo um total de 973 ESTs e 452 *clusters* (200 *contigs* e 252 *singlets*). Ademais, os filogramas mostram que as sequências de aminoácido deduzidas a partir de *contigs* de *C. arabica* do CafEST podem ser agrupadas em quatro grupos de prováveis genes *R* classe 1 e quatro grupos de prováveis genes *R* classe 2, além de mostrar que as sequências desses *contigs* são consideravelmente homólogas às sequências conhecidas de genes *R* usadas para rastrear o CafEST. Análises de domínios típicos de proteínas *R* estão em andamento e poderão adicionar novas informações a estes dados. Esse estudo vai auxiliar o futuro desenvolvimento de marcadores moleculares correlacionados com marcadores genéticos de resistência em cafeeiro e o futuro isolamento de sequências completas de genes *R* de *C. arabica* que poderão ser usados em transgenia de plantas.

Palavras-chave: *Coffea arabica*, resistência, gene *R* classe 1, gene *R* classe 2, EST, NBS, LRR.

IDENTIFICATION OF PUTATIVE *Coffea arabica* CLASS 1 AND 2 *R* GENES WITHIN THE BRAZILIAN COFFEE FUNCTIONAL GENOME DATABASE (CafEST)

Abstract:

Coffee is a traditional agricultural product of the Brazilian economy which accounts for 2,4% of the total value of exportations and for the generation of 8 million jobs. Brazil produces ca. 40% of the coffee commercialized in the international market and is the world's second biggest coffee consumer, particularly of the species *Coffea arabica*. Although *C. arabica* production and consumption corresponds to approximately 70% of the coffee market, it is highly susceptible to nematodes, pests and diseases. Therefore, there is a raising interest of genetic breeding programs in developing *C. arabica* varieties with resistance to nematodes, pests and diseases. A large number of plant resistance genes (*R* genes) have already been isolated and classified into six categories denoted class 1-class 6. The majority of the *R* genes belong to the class 2 and encode proteins which contain the following domains in its sequence: *nucleotide binding site* (NBS), *leucine-rich repeat* (LRR) and an N-terminal putative *leucine-zipper* (LZ) or other *coiled-coil* (CC) sequence. Class 1 comprises the *R* genes homologous to the type member gene of this class, which is the gene denoted *pto*, that encodes a protein possessing a serine/threonine kinase catalytic region and a myristylation motif at its N-terminus. Well described aminoacid sequences corresponding to class 1 and 2 *R* genes were used to screen the *Coffee Functional Genome* Brazilian Database (CafEST) for class 1 and 2 *R* gene homologous EST sequences from *C. arabica*. The selected ESTs in this search were grouped into *clusters* (*contigs* ou *singlets*), which were subsequently analyzed in terms of homology with *R* genes available within public databases. Multi-alignments among the consensus deduced amino acid sequences of *contigs* representing probable CafEST *R* genes were used to generate phylograms. The phylograms generated indicate that both putative class 1 and 2 *C. arabica* *R* genes are considerably represented within the CafEST, since a high number of class 1 and 2 *R* gene related ESTs was retrieved upon screening of this database, this is, 525 ESTs homologous to class 1 *R* genes

clustered into 262 clusters (118 contigs and 144 singlets) and 449 ESTs homologous to class 2 *R* genes clustered into 190 clusters (82 contigs and 108 singlets), resulting in a total of 973 ESTs and 452 clusters (200 contigs and 252 singlets). Moreover, the phylograms show that the CafEST *C. arabica* contig sequences may be grouped in four groups of putative class 1 *R* genes and four groups of putative class 2 *R* genes, besides showing these contig sequences are considerably homologous to the sequences used to screen the CafEST. Undergoing analysis of typical R protein domains may add new information to the presented data. This study will help the future development of molecular markers correlated with genetic markers of resistance in coffee and the future isolation of complete *R* gene sequences from *C. arabica*, which may be used for plant genetic transformation.

Key words: *Coffea arabica*, plant resistance, class 1 *R* gene, class 2 *R* gene, EST, NBS, LRR.

Introdução

O cafeeiro é uma planta arbustiva lenhosa da família *Rubiaceae* e que apresenta um longo ciclo biológico. Dentre as mais de 80 espécies do sub-gênero *Coffea* conhecidas até o momento, *Coffea arabica* é a espécie comercial mais cultivada e consumida. Apesar da produção e consumo de *C. arabica* corresponder a cerca de 70% do mercado de café, essa espécie é altamente susceptível a nematóides, pragas e doenças. Portanto, há um crescente interesse de programas de melhoramento genético de cafeeiro em desenvolver variedades de *C. arabica* com resistência a nematóides, pragas e doenças.

Um grande número de genes de resistência (genes *R*) de plantas já foi isolado e foi categorizado em seis classes denominadas classe 1- classe 6. A maioria dos genes *R* conhecidos pertence à classe 2 e contém os seguintes domínios em sua sequência: *nucleotide binding site* (NBS), *leucine-rich repeat* (LRR) e, no N-terminal, um domínio *leucine-zipper* (LZ) ou outra sequência *coiled-coil* (CC). A classe 1 compreende genes *R* homólogos ao membro tipo dessa classe, o gene *R* denominado *pto*, que apresenta domínio catalítico de serina/treonina quinase e regiões de miristilação em sua porção N-terminal.

No presente estudo, é descrita a identificação de sequências de *C. arabica* possivelmente codantes de genes *R* classe 1 e 2 presentes no Banco Brasileiro *Genoma Funcional de Café* (CafEST). Análises mais aprofundadas dessas sequências fornecerão dados importantes sobre os mecanismos de resistência em cafeeiro. A identificação e validação da função dos potenciais genes *R* classe 1 e 2 de *C. arabica* do CafEST apoia o futuro desenvolvimento de marcadores moleculares para assistir programas de melhoramento de cafeeiro, além de apoiar o futuro isolamento de genes *R* de café, visando a geração de plantas transgênicas resistentes a doenças.

Material e Métodos

Todas as sequências de *C. arabica* usadas no presente trabalho correspondem a ESTs sequenciados e *clusters* (*contigs* e *singlets*) obtidos no Banco Brasileiro *Genoma Funcional de Café* (CafEST: <http://cafe.lge.ibi.unicamp.br/>) e são oriundas de bibliotecas de cDNA específicas para diferentes genótipos, órgãos (folha, caule, fruto, folha e raiz), crescimento ou condições de estresse (Vieira et al, 2006). O CafEST foi rastreado por genes *R* classes 1 e 2, separadamente, usando-se sequências homólogas conhecidas de genes *R* classes 1 e 2 através de buscas por *Basic Local Alignment Tool* (BLAST) tBLASTn (Altschul et al., 1997). A sequência do gene *R pto* (gi|557882) de *Lycopersicon pimpinellifolium* foi usada como sequência homóloga de genes *R* classe 1. As seguintes sequências, que representam subclasses da classe 2 de genes *R*, foram também usadas para rastrear o CafEST: Bs2 (gi|6456755) de *Capsicum chacoense*; Dm3 (gi|4106975) de *Lactuca sativa*; Gpa2 (gi|6164969) de *Solanum tuberosum*; Hero (gi|26190258) de *Lycopersicon esculentum*; HRT (gi|7110565) de *Arabidopsis thaliana*; I2 (gi|4689223) de *L. esculentum*; Mi1.1 (gi|3449378) de *L. esculentum*; Mi1.2 (gi|3449380) de *L. esculentum*; Pib (gi|37777009) de *Oryza sativa*; Pi-ta (gi|12642090) de *O. sativa*; R1 (gi|17432423) de *Solanum demissum*; RP1 (gi|5702196) de *Zea mays*; RPM1 (gi|963017) de *A. thaliana*; Rpp8 (gi|29839585) de *A. thaliana*; Rpp13 (gi|7229451) de *A. thaliana*; Rps2 (gi|15236112) de *A. thaliana*; Rps5 (gi|3309619) de *A. thaliana*; Rx1 (gi|8515762) e Rx2 (gi|5911745) de *S. tuberosum*; Xa1 (gi|2943742) de *O. sativa* e Sw5 (gi|15418714) de *L. esculentum*. Os ESTs de *C. arabica*, recuperados pelo rastreamento, foram agrupados em *contigs* e *singlets* para cada classe de gene *R* separadamente usando-se o *Contig Assembly Program* (CAP3). As sequências de aminoácidos deduzidas e correspondentes a *Open Read Frame* (ORF) dos *contigs* de *C. arabica* foram obtidas usando o programa *NCBI ORF Finder* (<http://www.ncbi.nlm.nih.gov/projects/gorf/>). As sequências aminoacídicas de *contigs* de *C. arabica* e das sequências homólogas mencionadas acima foram analisadas, separadamente para cada classe de gene *R*, por multi-alinhamentos feitos pelo programa *EMBL-EBI ClustalW* (<http://www.ebi.ac.uk/clustalw/>). Os multi-alinhamentos demonstrando as similaridades entre sequências aminoacídicas foram usados para gerar representações em filograma, separadamente para cada classe de gene *R*, utilizando-se o programa *EMBL-EBI ClustalW*. Os filogramas resultantes foram visualizados utilizando-se o programa *TreeView* (<http://darwin.zoology.gla.ac.uk/~rpage/treeviewx/>).

Resultados e Discussão

A busca por homólogos de genes *R* classes 1 e 2 de *C. arabica* no CafEST, usando como critério *e-value* inferior a 1×10^{-4} , levou à identificação de 525 ESTs homólogos a genes *R* classe 1 e 449 ESTs homólogos a genes *R* classe 2, somando um total de 973 ESTs. O agrupamento das ESTs, realizado para cada classe separadamente, resultou em 118

contigs e 144 *singlets* de prováveis genes *R* classe 1 e 82 *contigs* e 108 *singlets* de prováveis genes *R* classe 2. Portanto, o rastreamento do CafEST por genes *R* classe 1 utilizando-se uma única sequência homóloga (i.e. *pto*) foi eficiente pois recuperou um grande número de prováveis genes *R* classe 1 de *C. arabica*. Vários *contigs* de prováveis genes *R* classe 1 apresentaram longos fragmentos de sequência de aminoácidos com alta homologia a sequência gênica de *pto*, como indicado no filograma na Figura 1. No filograma, as sequências de aminoácido deduzidas de *contigs* do CafEST homólogos a genes *R* classe 1 foram agrupadas em quatro grupos principais, sendo que no maior deles se encontra a sequência de *Pto* de tomate. A similaridade entre *contigs* de possíveis genes *R* classe 2 do CafEST, apresentada no filograma na Figura 2, demonstra que o CafEST possui ESTs representando todas as sub-classes rastreadas, ou seja, subclasses Bs2, Dm3, Gpa2, Hero, HRT, I2, Mi1.1, Mi1.2, Pib, Pi-ta, R1, RPM1, Rpp8, Rpp13, Rps2, Rps5, RX1, RX2, Sw5 e Xa1. No filograma de sequências de aminoácidos deduzidas de *contigs* de CafEST homólogos a genes *R* classe 2, observam-se quatro grupos principais, sendo um primeiro grupo homólogo às sequências de Pi-ta, Rps2, Rps5, Dm3, I2, RP1, HRT, Rpp8 e Rpm1, um segundo grupo formado por um *contig* sem homologia com outras proteínas R usadas para construir o filograma (C36), um terceiro grupo homólogo às sequências de Bs2, R1, Hero, Mi1.1, Mi1.2, Gpa2, RX1, RX2, Sw5, Pib e Rpp13 e um quarto grupo homólogo à sequência Xa1. Interessantemente, é possível que o *contig* C36 represente uma subclasse de genes *R* classe 2 de *C. arabica* distinta das atualmente conhecidas e bem descritas. Ademais, ambos filogramas indicam que os *contigs* gerados são consideravelmente homólogos às sequências conhecidas de proteínas R usadas para rastrear o CafEST.

Conclusões

No presente trabalho, é apresentada a busca no CafEST por prováveis genes *R* das classes 1 e 2 de *C. arabica* e a análise das suas homologias de sequência com outros genes *R* bem estudados. Demonstrou-se que prováveis genes *R* classes 1 e 2 de *C. arabica* estão vastamente representados dentro do CafEST e que, quanto à homologia de sequência aminocídica, tanto os genes *R* classe 1 quanto os genes *R* classe 2 se agrupam em quatro grupos distintos. O presente estudo subsidia o desenvolvimento de marcadores moleculares de resistência de plantas a serem empregados em programas de melhoramento genético de cafeeiro. Os dados aqui apresentados são igualmente fundamentais para o futuro isolamento de genes *R* completos de *C. arabica* a serem eventualmente usados na geração de plantas transgênicas resistentes a patógenos.

Agradecimentos

Os autores agradecem o apoio recebido da Embrapa Recursos Genéticos e Biotecnologia, da Embrapa Café e do Consórcio Brasileiro de Pesquisa dos Cafés do Brasil.

Referências Bibliográficas

Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W. & Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25:3389-3402.

Martin, G.B.; Bogdanove, A. J. & Sessa, G. (2003). Understanding the functions of plant disease resistance proteins. *Annual Review of Plant Biology*. 54:23-61.

Vieira, L.G. E., Andrade, A. C., Colombo, C. A., Araújo, A. H., Mehta, A. *et al.* (2006). Brazilian coffee genome project: an EST-based genomic resource. *Brazilian Journal of Plant Physiology* 18:95-108.

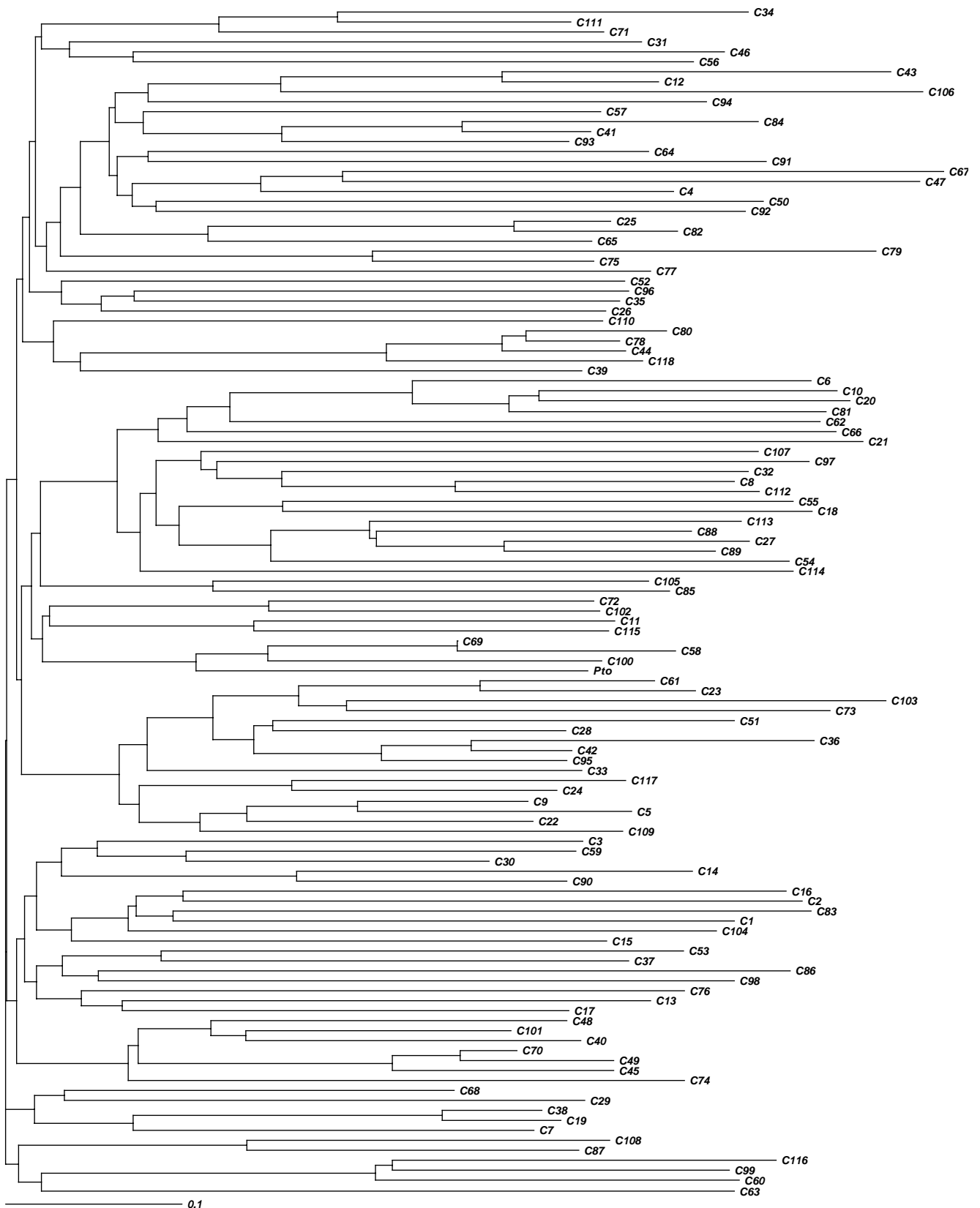


Figura 1. Representação em filograma da similaridade entre as seqüências de aminoácidos deduzidas a partir entre seqüências relacionadas com genes *R* classe 1. As seqüências analisadas são *contigs* de *C. arabica* do CafEST (C#) e uma seqüência conhecida de proteína *R* classe 1 usada para rastrear o CafEST (Pto).

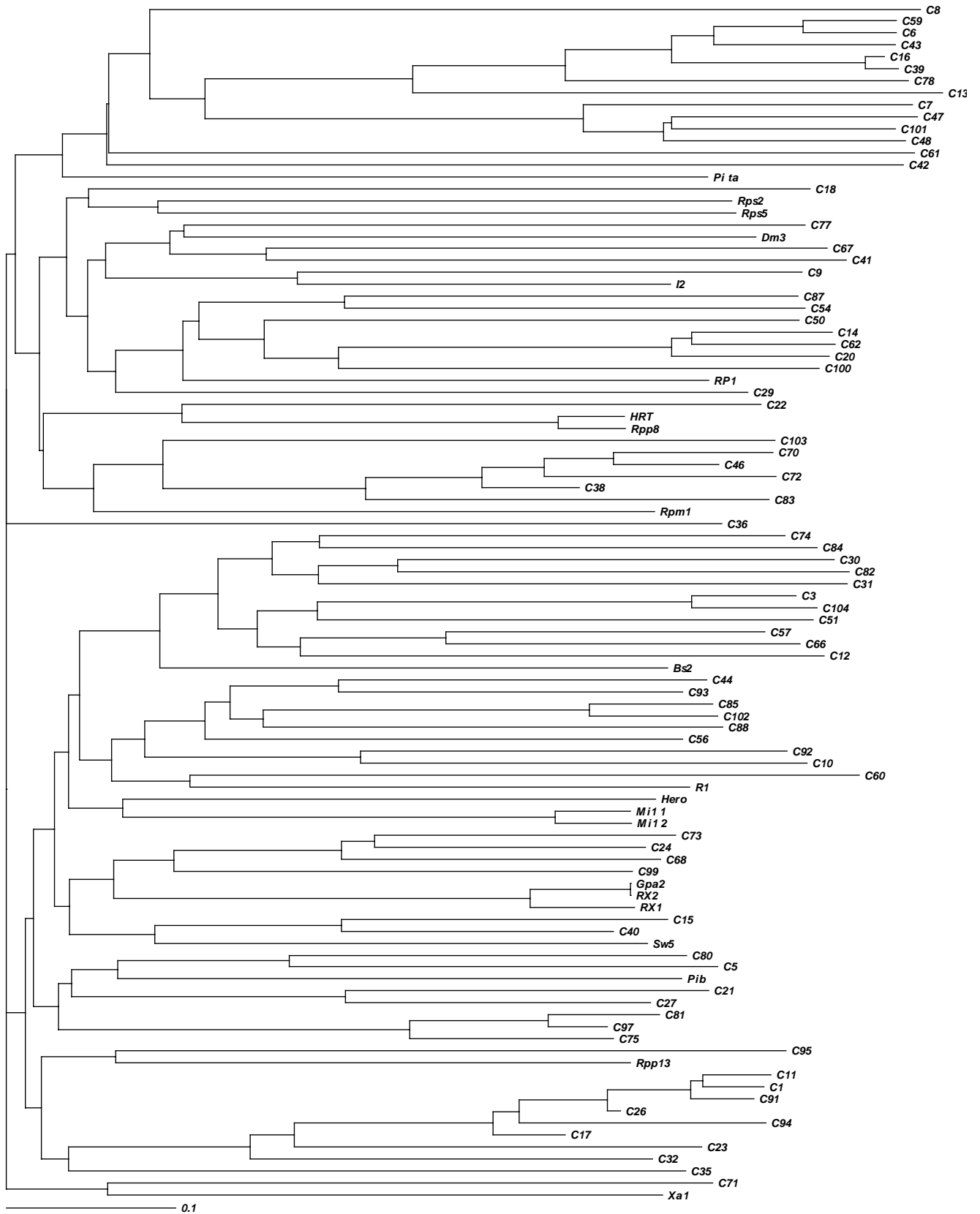


Figura 2. Representação em filograma de similaridade de sequência de aminoácidos entre sequências relacionadas com genes *R* classe 2. As sequências analisadas são *contigs* de *C. arabica* do CafEST (C#) e sequências conhecidas de proteínas *R* classe 2 usadas para rastrear o CafEST (Bs2, Dm3, Gpa2, Hero, HRT, I2, Mi1.1, Mi1.2, Pib, Pi-ta, R1, RP1, RPM1, Rpp8, Rpp13, Rps2, Rps5, Rx1, Rx2, Xa1, Sw5).