UNIVERSIDADE ESTADUAL DE SANTA CRUZ

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

GENÔMICA ESTRUTURAL E FUNCIONAL DA RESISTÊNCIA DO CACAUEIRO A *PHYTOPHTHORA PALMIVORA* BUTLER (BUTLER).

JONATHAN JAVIER MUCHERINO MUÑOZ

ILHÉUS – BAHIA – BRASIL
Fevereiro de 2019

JONATHAN JAVIER MUCHERINO MUÑOZ

GENÔMICA ESTRUTURAL E FUNCIONAL DA RESISTÊNCIA DO CACAUEIRO A *PHYTOPHTHORA PALMIVORA* BUTLER (BUTLER).

Dissertação apresentada à Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Genética e Biologia Molecular.

Área de concentração: Genética e Biologia Molecular

Orientador: Prof. Ronan Xavier Corrêa
Coorientadora: Edna Dora Martins Newman Luz

ILHÉUS – BAHIA – BRASIL
Fevereiro de 2019

M942 Mucherino Muñoz, Jonathan Javier.
Genômica estrutural e funcional da resistência do cacaueiro a Phytophthora palmivora Butler (Butler). / Jonathan Javier Mucherino Muñoz. – Ilhéus, BA: UESC, 2019.
viii, 72f. : il.

Orientador: Ronan Xavier Corrêa .
Dissertação (Mestrado) – Universidade Estadual de Santa Cruz. Programa de Pós-Graduação em Genética e Biologia Molecular.
Inclui referências e apêndices.

1. Mapeamento cromossômico. 2. Cacau – Melhoramento genético. 3. Bioinformática. 4. Marcadores genéticos. 5. Podridão parda (Fitopatologia). I. Título.

CDD 572.8633

JONATHAN JAVIER MUCHERINO MUÑOZ

GENÔMICA ESTRUTURAL E FUNCIONAL DA RESISTÊNCIA DO CACAUEIRO A *PHYTOPHTHORA PALMIVORA* BUTLER (BUTLER).

Dissertação apresentada à Universidade Estadual de Santa Cruz, como parte das exigências para obtenção do título de Mestre em Genética e Biologia Molecular.

Área de concentração: Genética e Biologia Molecular

26 de fevereiro de 2019

Profa. Dra. Silvia Nietsche
(UFMG)

Profa. Dra. Fabienne Micheli
(UESC)

Dr. Isamire Silva Andrade
(UESC)

Ronan Xavier Corrêa
(UESC-Orientador)

DEDICATÓRIA

A **Leonel A. Mucherino C.** “Ponchera”, por me fazer o que eu sou, isto é para você Pai.

Dedico

AGRADECIMENTOS

À Universidad de Los Andes, pela oportunidade de continuar crescendo academicamente.

À Organização dos Estados Americanos, pela oportunidade de representar a Venezuela.

Ao Grupo Coimbra de Universidades Brasileiras, pelo espaço no universo acadêmico Brasileiro.

À Universidade Estadual de Santa Cruz, pela infinita lista de oportunidades concedidas e por se tornar uma das minhas *Alma Mater*.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela concessão da bolsa de mestrado.

Ao Programa de Pós-Graduação em Genética e Biologia Molecular, por se tornar lar e família.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, pelo suporte financeiro ao projeto de pesquisa(Processo 481074/2013-9).

Ao Professor Ronan Xavier Corrêa, pela confiança, os ensinamentos e recomendações sempre oportunas.

Ao Dr. Cláusio Antônio Ferreira de Melo, pela paciência e constante apoio laboratorial e acadêmico.

Ao Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas, por facilitar a infra-estrutura e equipamentos computacionais.

Ao Centro de Biotecnologia e Genética, por facilitar a infra-estrutura e equipamentos laboratoriais.

Ao Grupo BIOPREA, pelo constante incentivo de superação e ímpeto na vanguarda.

"*Take the first step in faith. You don't have to see the whole staircase, just take the first step*".

Martin Luther King Jr.

ÍNDICE

GENÔMICA ESTRUTURAL E FUNCIONAL DA RESISTÊNCIA DO CACAUEIRO A *PHYTOPHTHORA PALMIVORA* BUTLER (BUTLER).

EXTRATO

O Estado da Bahia é atualmente o principal produtor de cacau no Brasil, sendo as sementes o principal produto para o abastecimento da indústria nacional e internacional de chocolate. Este estudo validou mediante mapeamento por associação, usando a abordagem de Modelo Linear Generalizado (GLM), 29 marcadores moleculares SSR flanqueando a QTLs (*Quantitative Trait Loci*) associados à resistência a *Phytophthora palmivora* Butler (PP), que é um dos três principais patógenos que atacam a cultura internacionalmente. A validação foi realizada em amostras de três variedades locais antigas da Bahia (Comum, Pará e Maranhão), variedades que possuem um alto potencial na produção de chocolate gourmet. Quatro locos SSR associados à resistência a PP foram detectados, dois no cromossomo 8, explicando 7,43% e 3,72% da variação fenotípica, um no cromossomo 2 explicando 2,71% da variação fenotípica e um no cromossomo 3 explicando 1,93% da variação fenotípica. Uma anotação funcional baseada em domínios funcionais foi realizada em dois genomas de referência CRIOLLO e MATINA, de 20 regiões QTL associadas à resistência ao patógeno. Foram identificados 250 genes candidatos (no genoma CRIOLLO) e 170 (no genoma MATINA), com alta probabilidade de estarem envolvidos no reconhecimento e ativação de respostas na interação com o patógeno. Regiões genômicas ricas em genes com domínios *Coiled-coils* (CC), *Nucleotide binding sites* (NBS) e *Leucine-rich repeat* (LRR) foram identificados nos cromossomos 1, 3, 6, 8 e 10 com alta probabilidade de estarem envolvidos luta contra o patógeno, da mesma forma, foram identificadas regiões ricas em genes com domínios *Receptor-like Kinase domain* (RLK) e *Ginkbilobin2* (GNK2) nos cromossomos 4 e 6, que também têm alta probabilidade de estarem envolvidas na luta contra o patógeno.

Palavras-chave: Mapeamento por associação. Bioinformática. Anotação funcional. *Theobroma cacao* L. *Podridão parda*.

STRUCTURAL AND FUNCTIONAL GENOMICS OF THE COCOA RESISTANCE TO *PHYTOPHTHORA PALMIVORA* BUTLER (BUTLER).

ABSTRACT

The state of Bahia is currently the main producer of cocoa in Brazil, with seeds being the main product for supplying the national and international chocolate industry. This study validated by association mapping, using the General Linear Model approach, 29 SSR molecular markers flanking to QTL (Quantitative Trait Loci) associated with *Phytophthora palmivora* Butler (Butler) (PP) resistance, which is one of the three main pathogens that attack the crop internationally. The validation was carried out in three local ancient varieties of the Bahia (Comum, Pará and Maranhão), varieties that have a high potential in the production of gourmet chocolate. Four SSR loci associated with resistance to PP were detected, two on chromosome 8, explaining 7.43% and 3.72% of the phenotypic variation, one on chromosome 2 explaining 2.71% of the phenotypic variation and one on chromosome 3 explaining 1.93 % of the phenotypic variation. A functional annotation based on functional domains was carried out, in two CRIOLLO and MATINA reference genomes, of 20 QTL regions associated with resistance to the pathogen. Were identified 250 (genome CRIOLLO) and 170 (genome MATINA) candidate genes, with high probability of being involved in the recognition and activation of responses in the interaction with the pathogen. Genomic regions rich in genes with Coiled-coils (CC), nucleotide binding sites (NBS) and Leucine-rich repeat (LRR) domains were identified on chromosomes 1, 3, 6, 8 and 10 with a high probability of being involved in the fight against the pathogen, likewise, regions rich in Receptor-like Kinase domain (RLK) and Ginkbilobin2 (GNK2) domains were identified in chromosomes 4 and 6, which also have a high probability of being involved in the fight against the pathogen.

Keywords: Association Mapping. Bioinformatics. Functional annotation. *Theobroma cacao* L. *Black Pod*.

1 INTRODUCTION

Cocoa (*Theobroma cacao* L.) is a perennial tree belonging to the Malvaceae Juss. (APG, 2016) family, with a natural distribution in the tropical forests of South America, an area that constitutes a central region of its genetic diversity (Schultes & Cuatrecasas, 1964). *T. cacao* L. is a diploid species (2n = 2x = 20) (Davie, 1935) with an estimated genome length of approximately 326.9 to 445 Mb (Argout *et al*., 2017; Motamayor *et al*., 2013). This plant has great, social and economic importance, representing a livelihood for more than 120 million people in 50 tropical countries, generating 12 billion of US dollars per year in revenues, mainly related to the production of chocolate (Green America, 2014; ICCO, 2017).

The Quarterly Bulletin of Cocoa Statistics (QBCE, 2014) projected that the 2014/2015 period would have a deficit in the global supply of cocoa of approximately 150,000 tons. However, the annual report of the INTERNATIONAL COCOA ORGANIZATION (ICCO, 2017) reported this deficit at approximately 196,000 tons, and additionally, FAOSTAT reported a 17% decrease in planted area in Brazil for the 2016/2017 period, and comparing the tons of cocoa bean produced in 2017 (235,809), the production has decreased 15% compared to the of 2015.

Brazil is the main producer of cocoa bean of America (ICCO, 2017) and the state of Bahia is the main producer of the country (IBGE, 2018.). In southern Bahia, nearly 70% of cocoa cultivation occurs in cabruca systems (a type of agroforestry system) (Araujo *et al*., 1998). Comum, Pará and Maranhão were the predominant cocoa varieties in Bahia, until at least 2003, representing approximately 50% of the land cultivated for cocoa in the state (Pires, 2003; Santos *et al*., 2015). These varieties were initially introduced in Bahia approximately in 1746 (Comum), 1874 (Pará) and 1876 (Maranhão) (Bartley, 2005; Vello & Garcia, 1971).

These varieties possess unique characteristics, and were naturalized as Bahian cocoa or Bahia local cocoa cultivars, this region is treated as a secondary

area of cocoa diversity (Bartley, 2005; Santos *et al.*, 2015). The Maranhão and Catongo (spontaneous mutants of Comum cocoa) varieties are currently used in the fine cocoa market, they produce less astringent and more flavorful chocolate (Leite *et al*., 2013). However these varieties have profiles of more susceptibility to diseases compared to new clonal cultivars produced in response to the outbreak of witches' broom disease (caused by the fungus *Moniliophthora perniciosa*) in Bahia in 1989 (Monteiro *et al.*, 2007)

Cocoa production has been affected by several factors, lack of varieties of high potential, pests and pathogens, among others. The black pod (BP) disease is one of the world's most devastating diseases in cocoa cultivation (Lanaud *et al.*, 2009) because of its direct effects on reduction of production (Akaza *et al.*, 2016). The BP in cocoa has an etiology represented by seven species of the genus *Phytophthora* (Luz *et al.*, 2004). The most important in the pantropic are *P. palmivora* Butler (Butler) and *P. capsici* Leonian (Guest, 2007). This disease is one of the most devastating that attack the cocoa (from the Greek, (phytón) "plant" and (phthorá) "destruction", "the destroyer of plants"). Losses varying of 10% up to 100%, depending on of the country, *Phytophthora* species and climatic zones are reported (Akrofi *et al*., 2003; Dakwa, 1987; Ploetz, 2007; Pokou *et al*., 2008).

Quantitative Trait Loci (QTL) mapping has been used for different quantitative traits, such as disease resistance (Clair, 2010; Kover & Caicedo, 2001). The mapping of QTLs associated with *Phytophthora* spp. resistance, was performed by several authors (Akaza *et al.*, 2016; Barreto *et al.*, 2018; Brown *et al.*, 2007; Clement *et al.*, 2003; Crouzillat *et al.*, 2000; Flament *et al.*, 2001; Lanaud *et al.*, 2004; Risterucci *et al.*, 2003). DIfferent statistical approaches to identification has been used, ranging from the traditional two-point approach to a multipoint approach using machine-learning algorithms and Hidden Markov Model (HMM) statistical models (Barreto *et al.*, 2018; Tong *et al*., 2010; Wu *et al*., 2002).

Different molecular markers have been used in the QTLs detection associated with BP resistance varied over time ranging from Restriction Fragment Length Polymorphism (RFLP), Random Amplification of Polymorphic DNA (RAPD), Amplified fragment length polymorphism (AFLP) and most commonly at present Single Sequence Repeats or Micro-satellites (SSR). Recent studies of QTLs associated with

BP and using SSR markers, the number of QTLs, location and percentage of the phenotypic variation (%PV) explained by each one varies significantly, which demonstrates the quantitative character inherent to the BP resistance. Brown *et al.* (2007) identified three QTLs associated with BP resistance, using a population of 256 F_1 individuals from a crossing Pound7 x UF273, a first QTL located in the linkage group (LG) 4,%PV= 8.7; the second in LG 8, %PV= 7.3 and one last in LG 10, %PV= 23. The three QTLs were identified using invivo inoculation of fruits in the field as a phenotyping strategy. More recently, Akaza *et al.* (2016) identified eight QTLs using the PRR (Pod Rot Rate, cumulative % of rotten fruit in relation to the total number of fruits for 3 years) phenotype methodology, and three QTLs using the methodology FOLRES (Foliar disc inoculation test), methodology developed by Nyassé *et al.* (1995). The authors applied the two methodologies in three segregating populations: (SCA6 x H) x C1 (179 individuals), (P7 x ICS100) x C1 (173 individuals) and (P7 x ICS95) x C1(183 individuals). For PRR, three QTLs were identified in the population (SCA6 x H) x C1: LG 1 %PV= 17.3, LG 6 %PV= 27.6 and LG 8 %PV= 13.2. Three QTLs for the population (P7 x ICS100) x C1: LG 4 %PV= 13.8, LG 6 %PV= 14.8 and LG 6 %PV= 20; and two QTLs for the population (P7 x ICS95) x C1: LG 2 %PV= 19.3 and LG 4 %PV= 21.7. The results from the FOLRES methodology found three QTLs only in the population (P7 x ICS100) x C1: LG 1 %PV= 13.8, LG 3 %PV= 14.5 and LG 10 %PV= 17.3. Barreto *et al.* (2018) identified six QTLs (for resistance to three species *P. palmivora, P. capsici and P. citrophthora*) in a F_1 (Full-sib) population of 256 individuals derived from the crossing of heterozygous TSH1188 (Trinidad Selected Hybrids) and CCN51(*Colección Castro Naranjal*), these QTLs were: LG 6 %PV= 2.543 for *P. palmivora*; LG 1 %PV= 1.819, LG 2 %PV= 2.101, LG 3 %PV= 1.776 and LG 4 %PV= 3.027 for *P. capsici* and LG 6 %PV= 3.299 for *P. citrophthora*, all identified by the leaf disc inoculation methodology.

The former results published pointed that the range of %PV varies from 1.776 to 27.6%, which evidences the effect of QTLs/genes of major effect and minor effect on the control of the characteristic. On the other hand, LGs and their location varies significantly in all cases studied. Considering these results, it is very important that in order to be able to use the information obtained by these authors in marker-assisted selection, these QTLs and their flanking markers needs the validation in populations other than those in which they of have been identified (different genetic background)

and/or needs the saturation of genomic regions under study (Collard & Mackill, 2008). Here, we aim to test the validity of these markers, reported as associated with BP resistance, in samples of three cocoa local varieties, as well as to identify the genes located in the QTL regions with *in-silico* tools, in order to test their potential as candidate genes for resistance responses.

2 REVISÃO BIBLIOGRÁFICA

2.1 *Theobroma cacao* L. e podridão-parda

O cacau (*Theobroma cacao* L.) é uma árvore perene pertencente à família Malvaceae Juss. (APG, 2016), com uma distribuição natural nas florestas tropicais da América do Sul, área que constitui uma região central de sua diversidade genética (Schultes; Cuatrecasas, 1964). A espécie apresenta grande importância social e econômica, representando um meio de vida para mais de 120 milhões de pessoas, em 50 países tropicais, gerando 12 bilhões (US$) por ano em receitas, principalmente relacionado à produção de chocolate (Green America, 2014; ICCO, 2017).

O Brasil é o principal produtor de cacau da América (ICCO, 2017) e o estado da Bahia é o principal produtor do país (IBGE, 2018.). O *Quarterly Bulletin of Cocoa Statistics* (QBCE) (2014) projetou que o período 2014/2015 teria um déficit na oferta global de cacau de aproximadamente 150000 toneladas. No entanto, o relatório anual da *INTERNATIONAL COCOA ORGANIZATION* (ICCO) de 2017 relatou esse déficit em aproximadamente 196000 toneladas. A FAOSTAT registrou uma redução de 17% na área plantada com a cultura no Brasil para o período 2016/2017, e comparando as toneladas de cacau produzidas em 2017 (235809) a produção diminuiu 15% em relação a 2015.

A podridão-parda (PP) (*Black Pod*) é uma das doenças mais devastadoras na cultura do cacau (Lanaud *et al*., 2009) devido aos seus efeitos diretos na perda de produção (Akaza *et al.,* 2016). A PP em cacau tem uma etiologia representado por sete espécies do gênero *Phytophthora* (Luz *et al.,*2004), sendo as mais importantes no pantrópico *P. palmivora* Butler (Butler) e *P. capsici* Leonian (Guest, 2007). O termo phytophthora vem do grego e significa, *phytón* "planta" e *phthorá* "destruição", "o destruidor de plantas", As perdas variam de 10% até 100%, dependendo do país,

espécies de *Phytophthora* e zonas climáticas (Dakwa, 1987; Akrofi *et al*., 2003; Ploetz, 2007; Pokou *et al*., 2008).

2.2 *Phytophthora palmivora* Butler (BUTLER)

O gênero *Phytophthora* foi originalmente descrito por Anton De Bary em 1876. Na atualidade o gênero *Phytophthora* pertence aos Chromalveolates, Reino Straminipila, filo Oomicota, classe Oomycetes (grupo monofilético) ordem Pythiales e família Pythiaceae. Durante muito tempo, o gênero foi classificado dentro do reino dos Fungos (Fungi), mas análises genéticas, fisiológicas e bioquímicas mostraram que elas possuem características que o relacionam mais diretamente com diatomáceas e algas marrons (Alexopoulos *et al.*, 1996; Luz & Matsuoka, 1996; Sogin & Silberman, 1998; Baldauf*et al.*, 2000; Haldar *et al*., 2006; Adl *et al.*, 2005; Lamour *et al.,* 2007; Verret *et al.*, 2010).

No patossistema *T. cacao* L. - *P. palmivora* o patógeno comporta-se de maneira hemibiotrófica, inicialmente infectando os tecidos vivos do hospedeiro (fase biotrófica) e, após matar as células infectadas, alimenta-se deles (fase necrotrófica) (Hardham, 2007).

A penetração do patógeno nos tecidos da planta hospedeira acontece preferencialmente por aberturas naturais como as células epidérmicas e os estômatos. *Phytophthora palmivora* apresenta quatro tipos de propágulos infectantes, três assexuados, esporângio, zoósporo e clamidósporo e um sexuado, o oósporo (Luz et al., 1997; Santos et al., 2004). Os propágulos de infecção assexual de *P. palmivora* podem acontecer de forma indireta, através dos zoósporos. Os zoósporos são liberados de esporângios e em presença de vento e água livre, encistam na superfície da folha. Em seguida, através da formação de tubos germinativos e, posteriormente de um haustório, penetram e colonizam o tecido da planta, representando a fase biotrófica da doença. De forma indireta, ocorre por meio dos esporângios decíduos que penetram nos tecidos do hospedeiro através da formação de hifas, caracterizando a fase necrotrófica da doença, gerando a morte celular do hospedeiro (Fawke *et al*., 2015; Zheng & Mackrill, 2016).

O principal sintoma da doença acontece nos frutos, apresentando uma mancha amarronzada de formato arredondado. Essa mancha torna-se mais escura entre três a quatro dias após a penetração. Às duas semanas da infecção do patógeno, este se espalha por todo o fruto, causando descoloração e deformação das amêndoas (Bowers *et al.*, 2001). Ao fim do desenvolvimento da infecção, os frutos que ainda estão aderidos à planta tornam-se mumificados e são colonizados por fungos oportunistas. A intensidade do sintoma depende da variedade do cacaueiro infectado e da agressividade do inóculo (Nyassé *et al.*, 1995; Érsek & Ribeiro, 2010; Vanegtern *et al.,* 2015).

Se a infecção acontece em partes da planta que não seja o fruto, os sintomas variam de acordo com o tecido. Nos chupões e lançamentos foliares novos, observa-se o aparecimento de lesões necróticas escuras tanto no limbo foliar quanto na haste, ramos e pecíolos. Em folhas novas observa-se necrose das nervuras e seca das folhas. Tais sintomas são observados com frequência em condições de viveiros e climas de alta umidade, além da murcha e seca das plântulas (Luz *et al.*, 2008).

As principais maneiras registradas de dispersão do patógeno são a partir do solo e de frutos doentes, respingos de chuva, insetos, almofadas florais, raízes, frutos mumificados, folhas, chupões e cancros contaminados (Guest, 2007; Vanegtern *et al.,* 2015).

O controle do patógeno nos frutos torna-se difícil pela presença no solo de estruturas de resistência que podem persistir e sobreviver por vários anos. Os propágulos também podem estar presentes na parte aérea das plantas e nos frutos durante grande parte do ano, aguardando as condições ambientais favoráveis para esporulação e dispersão (Evans, 1981; Soberanis *et al.*, 1999).

O controle da doença é difícil de ser realizado sem o uso de agroquímicos (Akaza *et al.,* 2016), apesar de tentativas mais responsáveis ambientalmente terem sido feitas (Opuku *et al.*, 2003; Mpika *et al*., 2009a, 2009b; Adebola & Amadi, 2010; Gadji *et al*., 2015; Fister *et al*., 2015). O emprego de práticas culturais como a remoção de frutos infectados e poda do cacaueiro, não controlam totalmente a doença, mas diminuem a fonte de inóculo. O desenvolvimento de cultivares geneticamente resistentes a esta doença é a estratégia mais efetiva, duradoura e

viável. O uso da seleção assistida por marcadores moleculares tem sido um meio para o desenvolvimento de cultivares desde a década de 2000, provando ser uma das estratégias mais eficazes para o controle genético (Michelmore, 2003; Macagnan, 2005; Zhang & Motilal, 2016).

2.3 Interação planta-patógeno

Na interação Planta-Patógeno, a troca e o reconhecimento de sinais moleculares representam, desde o início da interação, a possibilidade de respostas do hospedeiro e do patógeno às estratégias de ataque ou defesa. Em relação ao hospedeiro, ressaltamos que a capacidade de reconhecer sinais da presença de um patógeno representa uma fase fundamental para o possível combate do mesmo. Para o patógeno, o sucesso no ataque dependerá em parte da habilidade do hospedeiro em reconhecer e ativar estratégias de defesa e combate (Stahl & Bishop, 2000; Tyler *et al.*, 2006).

Nas plantas existem diferentes estratégias de defesa e combate à presença de um organismo patogênico. Algumas dessas estratégias são consideradas mecanismos estruturais de defesa, que podem ser pré-formados como exemplo: número, morfologia, localização e período de abertura dos estômatos, presença de cutículas, vasos condutores, densidade de tricomas e se estes possuem glândulas secretoras. Ou mecanismos pós-formados, que podem atuar gerando alterações morfológicas como o aumento do fluxo citoplasmático na direção da penetração, rearranjo do cito esqueleto, migração do núcleo, justaposição dos componentes da parede celular, barreiras estruturais pós-formadas (lignificação), geração de camadas de cortiça, geração de camada de abscisão e tilose.

Dentro das respostas pós-formadas também podem ocorrer mudanças bioquímicas como a produção de fenóis, alcalóides, fito toxinas, glicosídeos cianogênicos e fenólicos, proteínas relacionadas à patogênese (PR-*Pathogenesis-related*), influxo de cálcio,aumento da produção de etileno e óxido nítrico, aumento da produção de espécies reativas de oxigênio (ROS –*Reactive Oxygen Species*), dentre outros (Dodds & Rathjen, 2010; Taiz & Zeiger, 2013; Teixeira *et al.*, 2014).

A identificação de possíveis patógenos pelo sistema imune das plantas pode ocorrer de duas formas: primeira, a imunidade desencadeada por PAMPs (*pathogen associated molecular patterns*) usualmente denominado PTI (*PAMP-triggered immunity*) ou MAMPs (*microbe associated molecular patterns*). A segunda, é a imunidade desencadeada por efetores ETI (*effector-triggered immunity*) (Jones & Dangl, 2006). O reconhecimento das PAMPs é feito por receptores PRRs (*PAMP-recognition receptor*) que podem ser localizados na membrana celular ou dentro da célula, ativando uma cascata de transdução de sinais na célula hospedeira (Denoux *et al.*, 2008; Hogenhout *et al.*, 2009; Roux *et al.*, 2014; Zipfel *et al.*, 2004).

Se o patógeno estiver adaptado para contornar a PTI, desenvolve-se a suscetibilidade desencadeada por efetores (ETS–*Effector triggered susceptibility*), em que o patógeno secreta moléculas efetoras com a função de suprimir as possíveis respostas de combate da planta (Birch *et al.*, 2008; Kamoun, 2006). Algumas plantas possuem um sistema de reconhecimento de efetores por meio de proteínas de resistência (R). Se a planta apresentar proteínas R que reconheçam as moléculas efetores, resulta em reação incompatível e o patógeno não consegue desenvolver-se nos tecidos do hospedeiro (resistência). Ao contrário, quando as plantas não possuem essas proteínas de reconhecimento, ocorre a reação compatível e a doença se desenvolve no hospedeiro (suscetibilidade) (Howden & Huitema, 2012). O reconhecimento de efetores por proteínas R pode ser direto (modelo gene-a-gene) (Flor, 1942) ou indireto (modelo guarda) (Obwald *et al.*, 2014).

O reconhecimento do patógeno de acordo com o modelo guarda acontece quando há uma perda de um gene "S" (suscetibilidade) na planta, o que traz como consequência a possível diminuição na capacidade do patógeno de ter sucesso, o que pode resultar em resistência patogênica específica se o gene estiver envolvido na produção de componentes necessários para a sua penetração no hospedeiro ou na resistência de largo espectro se o gene suprimir as defesas constitutivas (Fawke *et al.,* 2015).

A resistência nas plantas pode ser dividida em resistência sistêmica induzida (ISR-*Induction of systemic resistance*) e resistência sistêmica adquirida (SAR-*Systemic Acquired Resistance*) (Van Loon *et al.*, 1998). No sistema ISR, o agente indutor não é patogênico e não há acúmulo de proteínas PR. O indutor não provoca

sintomas, como necrose ou outros, no local da infecção, mas desencadeia uma defesa sistêmica na planta (Pieterse *et al.,* 1998). Para a resistência do tipo SAR, a resistência desenvolve-se de maneira sistêmica a um amplo espectro de patógenos, desencadeando um tipo de resposta rápida que vai gerar uma lesão necrótica no local da infecção, conhecida como reação de hipersensibilidade (HR-*Hyper sensitive Response*), caracterizada por um colapso do tecido lesionado, liberando compostos que podem ou não ser tóxicos para o patógeno (Nimchuk *et al.*, 2003; Van Loon, 2006).

Existem alguns genes já relatados em plantas como relacionados à defesa e combate à *Phytophthora*, codificando PRRs ou PRs ou enzimas.O elicitor INF 1 foi descoberto em *P. infestanse* pertence à família de proteínas com função de transferência de lipídeos entre lipossomas (Gaulin *et al.*, 2010; Mikes *et al.*, 1998). O INF1 tem como receptor o BAK1/SERK3-dependent ELR, identificados em tabaco e batata (Chaparro-Garcia *et al.*, 2011; Kamoun *et al.*, 1997; Kamoun *et al.*, 1998; Du *et al.*, 2015). O elicitor lecitina de ligação à celulose (CBEL), de *P. parasítica* tem função de adesão à célula hospedeira (Mateos *et al.*, 1997; Larroque *et al.*, 2012). Há sinalização *downstream* após percepção de CBEL em células de tabaco mas não em células do protoplasto por falta de parede, sugerindo que a ligação da parede celular é necessária para as reações de defesa induzidas pelo CBEL (Denecke *et al.*, 2012).

Os receptores para CBEL não são totalmente conhecidos, mas sabe-se que são proteínas dependentes de celulose (Gaulin *et al.*, 2006). Os β-glucanos são polissacarídeos de monômeros de D-glicose unidos por ligações glicosídicas β e representam PAMPs proveniente de oomicetos fitopatogênicos. β-glucano com ramificações β(1-6) foi identificado provocando resposta de defesa em plantas de soja, e β(1-3), em plantas de tabaco (Cheong *et al.*, 1991; Klarzynski *et al.*, 2000). Os receptores para β-glucano em plantas, são glucano-dependente CEBiP CERK1 (Kaku *et al.*, 2006; Miya *et al.*, 2007; Nars *et al.*, 2013).

A OPEL é um polipeptídio secretado a partir de filtrados de cultura de *P. parasítica* durante a infecção na planta *Nicotiana tabacum*. A infiltração de proteína OPEL recombinante (atividade predita de laminarinase) resultou numa resposta imunitária melhorada da planta e resistência a *P. parasítica*, levando a respostas de

defesa como a deposição de calose, a morte celular, a síntese de espécies reativas de oxigênio, entre outros (Chang *et al.*, 2015). Sabe-se que os receptores para o OPEL são proteínas componentes da membrana plasmática das células hospedeiras (Nürnberger *et al.*, 1994; Nennstiel *et al.*, 1998).O elicitor Pep-13 foi isolado da parede celular de *P. sojae* sendo a primeira molécula definida como um PAMP. O Pep-13 é um peptídeo curto de 13 aminoácidos com uma superfície exposta dentro de um fragmento da proteína transglutaminase cálcio dependente que provoca respostas de defesa em Solanaceae. A mutação de apenas um destes grupos de aminoácidos é suficiente para prejudicar o reconhecimento mediado pela transglutaminase de *P. sojae*, inibindo assim respostas de defesa da planta (Nürnberger *et al*., 1994).

A proteina GmPR10 foi isolada e caracterizada como relacionada a patogênese em soja (*Glycinemax*) (Xu*et al.*, 2014). Três PR (PR14a, PR14b e PR14c) foram isoladas a partir de folhas de tomate (*Lycopersico nesculentum* Mill. cv Baby) infectadas com *P. infestans,* havendo inibição da germinação dos zoósporos de *P. infestans*, tanto em estudos *in vitro* quanto *in vivo* em discos foliares do tomateiro (Niderman *et al.*, 1995). A expressão constitutiva da proteína PR-5, mostrou o aumento da tolerância à *P. citrohpthora* em plantas de laranja transgênicas (*Citrus sinensis* L. Obs. Cv. Abacaxi) (Fagoaga *et al.,* 2001).

2.4 *Quantitative Trait Loci*

O mapeamento de Locos de Caracteres Quantitativos (*Quantitative Trait Loci*, QTL) tem sido utilizado em diferentes caracteres, como resistência a doenças (Clair, 2010; Kover & Caicedo, 2001). O mapeamento de QTLs associado à resistência a *Phytophthora* spp. foi realizado por diversos autores (Akaza *et al.*, 2016; Barreto *et al*., 2018; Brown *et al*., 2007; Clement *et al.*, 2003; Crouzillat *et al.*, 2000; Flament *et al*., 2001; Lanaud *et al.*, 2004; Risterucci *et al.*, 2003). Distintas abordagens estatísticas de identificação têm sido utilizadas, variando da abordagem tradicional de dois pontos até uma abordagem multiponto usando algoritmos de *machine learning* e modelos estatísticos *Hidden Markov Model* (HMM) (Barreto *et al.*, 2018; Tong *et al*., 2010; Wu *et al*., 2002).

Os marcadores moleculares utilizados na detecção de QTLs associados à resistência a PP variaram ao longo do tempo e vão desde RFLPs (*Restriction Fragment Length Polymorphism*), RAPDs (*Random Amplification of Polymorphic DNA),* AFLPs *(Amplified fragment length polymorphism*) e mais comum na atualidade SSRs (*Single Sequence Repeats* ou Microsatelites). No presente estudo foram validados os QTLs encontrados pelos estudos mais recentes e que apenas usaram SSRs em sua identificação. O número desses QTLs, localização e porcentagem da variação fenotípica (%VF) explicado por cada varia significativamente, que demonstra a condição de caráter quantitativo inerente à resistência a PP.

Brown *et al.* (2007) identificaram três QTLs associados à resistência de PP, usando uma população de 256 indivíduos de uma F_1 do cruzamento Pound7 x UF273. Um primeiro QTL localizado no grupo de ligação (LG) 4, %VF=8,7; o segundo no LG 8, %VF=7,3 e um último no LG 10, %VF=23. Os três QTLs foram identificados usando inoculação invivo dos frutos no campo como estratégia de fenotipagem. Posteriormente, Akaza *et al* (2016) identificaram oito QTLs usando a metodologia PRR (*PodRot Rate*, % acumulado de frutos podres em relação ao número total de frutos por 3 anos) de fenotipagem, e três QTLs utilizando a metodologia FOLRES (Teste de inoculação em disco foliar), metodologia desenvolvida por Nyassé *et al*. (1995). As duas estratégias foram avaliadas em três populações segregantes: (SCA6 x H) x C1 (179 indivíduos), (P7 x ICS100) x C1 (173 indivíduos) y (P7 x ICS95) x C1 (183 indivíduos). Para PRR identificou-se três QTLs na população (SCA6 x H) x C1: LG 1 %VF=17,3, LG 6 %VF=27,6 e LG 8 %VF=13,2; três QTLs para a população (P7 x ICS100) x C1: LG 4 %VF=13,8, LG 6 %VF=14,8 e LG 6 %VF=20; dois QTLs para a população (P7 x ICS95) x C1: LG 2 %VF=19,3 e LG 4 %VF=21,7. Com o uso da metodologia FOLRES de fenotipagem identificou-se três QTLs apenas na população (P7 x ICS100) x C1: LG 1 %VF=13,8, LG 3 %VF=14,5 e LG 10 %VF=17,3. Em um trabalho mais recente, Barreto *et al*. (2018) identificou seis QTLs (para resistência a três espécies *P. palmivora, P. capsici* e *P. citrophthora*) em uma população F_1 (*Full-sib*) de 256 indivíduos derivados do cruzamento de clones heterozigotos TSH1188 (*Trinidad Selected Hybrids*) e CCN51 (*Colección Castro Naranjal*), esses QTLs foram: LG 6 %VF= 2,543 para *P. palmivora*; LG 1 %VF=1,819, LG 2 %VF=2,101, LG 3 %VF=1,776 e LG 4

%VF=3,027 para *P. capsici*e LG 6 %VF=3,299 para *P. citrophthora*, todos identificados através da metodologia de inoculação de disco foliar.

O intervalo de %VF nos três trabalhos variou de 1,776 a 27,6% evidenciando a ação de QTLs/genes de efeito maior e de efeito menor no controle da característica de estudo. Por outro lado, os LGs e a localização dos mesmos variam significativamente nos três casos, o que justifica que para poder usar a informação apresentada pelos autores na seleção assistida por marcadores. Estes QTLs e seus marcadores flanqueantes devem ser validados em populações diferentes das que foram identificadas (diferente *Background* Genético) e/ou as regiões genômicas sob estudo devem ser saturadas (Collard; Mackill, 2008).

2.5 Genômica

Nas últimas duas décadas, o surgimento de tecnologias de sequenciamento mais precisas, juntamente com a redução significativa do custo desses procedimentos, possibilitou a geração de diversas fontes de informações genômicas de alta utilidade em trabalhos de pesquisa, considerando tambem o aumento da capacidade computacional em geral e o aumento da capacidade e número de servidores (supercomputadores) foi possível analisar grandes quantidades de dados, o que até pouco tempo não era viável do ponto de vista computacional.

Para o organismo *T. cacao* L., existem dois genomas anotados e públicos, o genoma "CRIOLLO" (Argout*et al*., 2011) e "Matina1-6" (Motamayor *et al*., 2013). O genoma de CRIOLLO foi gerado a partir do clone B97-61/B2 multiplicado de um genótipo de um material antigo criollo coletado nas montanhas maias de Belize (Mooleedhar *et al*., 1995) e conservado no *International Cocoa GenBank* (ICG) em Trinidade. Esse genótipo criollo foi selecionado por sua capacidade de auto-cruzamento e seu alto nível de homozigose inicialmente relatado como 93% para 130 marcadores SSR e posteriormente como 99,9% para 795 SNP (Motilal *et al*., 2010), o que facilita a montagem do genoma devido à baixa frequência de locus heterozigotos, aqueles que geram dualidade no momento da montagem.

O genoma CRIOLLO foi sequenciado usando a estratégia de *Whole Genome Shotgun* (WGS) utilizando NGS (*Next-Generation Sequencing*) Roche/454 excetuando sequências de terminações BAC que foram produzidas através de sequenciamento *paired-end* de clones utilizando tecnologia a Sanger. A montagem de *reads* foi feita no NewBler V2.3 (*Contig* N50=19,8Kb e *Scaffolds* N50=473,8Kb). O acumulado de *scaffolds* alcançou 326,9 Mb, cerca de 24% menos que os 430 Mb do tamanho estimado do genoma. Dos 38737 unigenes de cacau, 97,8% foram confirmados na montagem utilizando BLAT (Kent, 2002), para *reads* pequenos foi utilizado SOAP (Li *et al*., 2008) (tamanho de *seed* = 12 pb e um tamanho máximo de *gap* de 3 pb por *read*).

A ancoragem no genoma foi realizada utilizando dois mapas de progênies, o primeiro desenvolvido e localizado na CNRA (*Centre National de Recherche Agronomique*, Divo, Costa do Marfim). Foi utilizada uma F1 de 256 indivíduos resultado do cruzamento de dois genótipos heterozigotos UPA402 (Forastero Peruano da alta Amazônia) com UF676 (híbrido entre forastero e criollo) selecionado na Costa Rica. O segundo mapa utilizado para ancoragem está localizado e foi desenvolvido pela CEPLAC (Comissão Executiva de Planejamento da Lavoura Cacaueira, Itabuna, Brasil) foi utilizada uma progênie F2 de 136 indivíduos produto do cruzamento de dois parentais heterozigotos, ICS21 (Trinitário seleto em Trinidad) e Scavina6 (Forastero da alta Amazônia). Dos 1259 marcadores (SSR e SNP) utilizados na ancoragem 1192 (94,3%) mostraram um único *hit* de 80% ou mais de identidade. Para a predição de genes codificantes, eles usaram EuGene (Foissac *et al*., 2008) usando diferentes bancos de dados de similaridade.

Um número variado de *tracks* foi varrido no genoma, sendo de grande interesse para o presente estudo a identificação de genes ortólogos a NBS, LRR-LRK e NPR1-like relacionados à resistência a doenças no cacau, que foi ancorado com uma meta-análise de 76 QTLs relacionados à resistência a doenças no cacau (Lanaud *et al*., 2009). O mapa comparativo forneceu um grande número de genes candidatos para resistência, e os autores recomendam uma abordagem funcional para a confirmação destes (Argout *et al*., 2011).

Este genoma foi recentemente melhorado (Argout *et al*., 2017), usando os mesmos dados dos *contigs* Roche/454, os dados Sanger (BAC *ends*) e as *Paired*

end reads, além destes dados foram incluídos quatro *inserts* de tamanho grande 3-5Kb, 5 -8Kb, 8-11Kb e 11-15Kb usando HiSeq2000 Illumina e, 78 SMRT *Cells Pacific Biociences* (correspondendo a 52x da cobertura de *reads* longas). Esta melhoria nas entradas de sequenciamento levou a melhorias em: *scaffolds* N50=6.5Mb (anteriormente 0.47Mb), ancoragem=96.7% (anteriormente 66.8%) e a diminuição em sítios desconhecidos de 5.7% (anteriormente 10.8%).

Quanto ao genoma de Matina (Motamayor *et al*., 2013), do qual foi publicada uma versão preliminar (V0.9) em 2010, foi sequenciado a partir de amostras do clone Matina1-6 pertencente ao grupo de germoplasma de cacau amelonado (Motamayor *et al*., 2008). Esse é o grupo mais comum em cultivos em todo o mundo, e tem uma baixa diversidade em comparação com os outros nove grupos já relatados (Motamayor *et al*., 2003; Motamayor *et al*., 2008). Apresenta um alto nível de homozigosidade (Bartley, 2005), uma condição favorável para montagem genômica. Esta condição de homozigosidade é explicada pela sua capacidade de auto cruzamento (autocompatibilidade).

Uma metodologia de sequenciamento mista foi usada incluindo Roche/454 e Sanger, usando uma estratégia WGS. A montagem das *reads* foi feita usando Arachne (versão 20071016) (Jaffe *et al*., 2003) (parâmetros maxcliq1 250, lap_ratio = 0.8, max_bad_look = 2000 e BINGE_AND_PURGE =True). A montagem final de *contigs* (N50 = 84.4Kb) e *scaffolds* (N50 = 34,4 Mb) gerou um acumulado de 445 Mb, que está dentro da faixa esperada para o cacau 411Mb-494Mb (Argout *et al*., 2011).

Possíveis contaminações foram identificadas usando Magablast (Altschul *et al*., 1990) com o Genbank NT e, blastx com bibliotecas de um grupo genes conhecidos de micróbios. A integração de mapas de ligação e a construção de pseudomoléculas em uma escala cromossômica foram realizadas usando uma combinação de mapas de marcadores, junções BAC/fosmid e informações adicionais de genes específicos em regiões funcionais, os marcadores dos mapas genéticos foram alinhados usando BLAT (Kent, 2002).

A anotação gênica foi feita através do isolamento do RNA das folhas, sementes e pistilos de cacau Matina1-6, em condições normais, sequenciando e usando essa informação como evidência de expressão na anotação. O

transcriptoma de cada coleção (folhas, sementes e pistilos) foram depositados no NCBITSA com BioProject 51633 e o transcriptoma foi refinado usando PASA (Haas, 2008).

A predição de proteínas (*ab initio*) foi realizada usando AUGUSTUS (Stanke & Waack, 2003), que foi treinado para parâmetros genéticos usando a coleção completa de cDNA extraída de folhas de Matina1-6. A anotação proteica foi realizada usando o banco de dados UniProt (UniProt, 2013) (genes de plantas) e o proteoma de oito espécies de plantas (*Arabidopsis thaliana* (TAGI, 2013), *V. vinifera* (Jaillon*et al*.), *Populus trichocarpa* (Tuskan *et al*.), *G. max* (Schmutz *et al*.), *Ricinus communis*(Chan *et al*.), *Fragaria vesca* (Shulaev *et al*.), *Solanum tuberosum* (Paterson *et al*.) e *Sorghum bicolor* (TSBG)) além de Matina1-6, a ortologia das nove plantas foi realizada usando OrthoMCL (Li *et al*., 2003).

Comparando os algoritmos de previsão de genes e proteínas utilizados pelos dois genomas acima mencionados, é apreciado que as metodologias de previsão eram diferentes. Para CRIOLLO, foi utilizado um software baseado na similaridade e para MATINA foi utilizada uma predição do tipo *ab initio*.

Os dois sistemas de previsão têm vantagens e desvantagens. O sistema por semelhança usa predição de proteína por homologia, isto significa que prediz em função de proteínas já conhecidas, o que pode ser muito útil para proteínas comuns ou bem descritas nos diferentes bancos de dados de proteínas. Em contrapartida, se as proteínas de *query* (*input*) são pouco conhecidas ou não foram estudadas, o poder de previsão deste tipo de *software* é quase nulo (Muniba, 2015).

O sistema de previsão *ab initio*, é ideal para a predição de proteínas de organismos não-modelos, dos quais pouca ou nenhuma informação está disponível sobre o perfil proteico e as proteínas individuais que compõem. Por outro lado, este sistema de previsão é mais complexo e com um custo computacional significativamente maior comparado ao sistema de previsão por homologia (Muniba, 2015; Lu *et al*., 2002; Floudas *et al*., 2006).

Considerando os prós e contras dos dois sistemas de previsão, escolhemos utilizar as proteínas previstas para os dois genomas e as duas metodologias como dados de entrada adequada para a anotação funcional de um organismo não

modelo, como *T. cacao* L.. A dupla fonte de dados pode garantir uma confiabilidade maior nas etapas subsequentes necessárias para a anotação e identificação de possíveis genes candidatos.

3 MATERIALS AND METHODS

3.1 Plant materials

The biological material used in this study was obtained from the leaves of 36 cocoa trees, located in the farm Novo Horizonte, Uruçuca, Bahia, Brazil (14° 36.907' S and 39° 15.891' W). Twelve individuals were selected in each of the three main diversity groups reported by Rhodes (2016), Within these 12 individuals per group of diversity, represented the three ancient local varieties (ALV) (Comum, Pará and Maranhão) reported by Santos *et al.* (2015)(Table 1) and for each subgroup of individuals from each ALV, phenotypically contrasting individuals were selected for the *P. palmivora*/*T. cacao* interaction, previously phenotyped by Rhodes, (2016) using the foliar disk methodology proposed by Nyassé *et al.* (1995, 2002), the method allowed to select individuals with the highest level of susceptibility to the pathogen and individuals with the minimum level of susceptibility (resistant) according to the gradient proposed by Nyassé *et al.* (2002). Additionally, four controls were used: (1) SCA6 resistant (Bahia *et al.*, 2015; Barreto *et al.*, 2015; Luz & Silva 2001), (2) SIC23 reported as moderately resistant (Luz *et al.*, 2006), (3) TSH1188 reported as moderately resistant or resistant (Bahia *et al.*, 2015) and (4) Catongo reported as susceptible (Barreto *et al.,* 2015). The plant material of these controls was collected in the Banco Ativo de Germoplasma of Centro de Pesquisa do Cacau (CEPEC/CEPLAC) located in the Km 22 of the highway Itabuna/Ilhéus, Ilhéus, Bahia, Brazil.

3.2 DNA extraction and PCR amplification

The Cetyl Trimethyl Ammonium Bromide (CTAB) method proposed by Doyle & Doyle (1990) was modified, for the extraction of total genomic DNA from leaves of 36

individuals of the local varieties (Comum, Pará and Maranhão) and the controls SCA6, SIC23, TSH1188 and Catongo. Proteinase-k was added to the extraction buffer at a final concentration of 400 µg/mL, all centrifugation times were duplicated and performed at a temperature of 4°C, a further purification of the nucleic acids was carried out with phenol (pH 8.0, equilibrated, ultrapure), chloroform and isoamyl alcohol in a proportion 25:24:1. The final precipitation of the nucleic acids was carried out with isopropanol at a volume of 1:1 with the sample. Finally, three washes were performed, two times with 70% ethanol and once with 95% ethanol and a final resuspension with 100 µL of TE (10 mM Tris-HCl, 1 mM EDTA [pH; 8.0]). The integrity and quantity of extracted DNA was visually quantified in 1.5% agarose gel comparing the samples with a standard lambda phage marker (λ 100 ng). The purified DNA solutions were subsequently diluted with water to a final concentration of 5 ng/µL before use for microsatellite analyses.

3.3 SSR analysis

Thirty-six SSR markers flanking the 20 QTLs associated with resistance to BP reported by Barreto *et al.* (2018), Akasa *et al.* (2016) and Brown *et al.* (2007) with a distance less than 10 cM, in the respective linkage maps, from each QTL were selected. The authors, sequences of the primers and other information of the markers are described in the Table 2. The mixture for polymerase chain reaction (PCR) amplification, up to a final volume of 13 µL with water, contained 15 ng template DNA, 0.131 mM forward primer 5' labeled with the M-13 tail sequence 5'-CACGACGTTGTAAAACGAC-3', 0.5 mM reverse primer, 0.5 mM M-13 fluorescent DYE (one for each SSR marker, 6-FAM™, NED™, PET ® , VIC ®), 0.254 mM dNTP mix, 1.5 mM $MgCl_2$, 15.7 nM BSA (bovine serum albumin, supplied in 10 nM Tris-HCl [pH 7.4 at 25 ºC], 100 mM KCl, 1 mM EDTA and 50% [v/v] glicerol), 10x PCR buffer (200 mM Tris-HCl [8.4], 500 mM KCl) and 1 U of Taq polymerase (Invitrogen, SP, Brazil). The PCR reactions were performed using a Veriti 96-Well Thermal Cycler (Applied Biosystems), the PCR program included an initial denaturation at 94 ºC for 5 min, followed by first 30 cycles of 30 s denaturation at 94 ºC, 45 s for annealing at 55 ºC and 45 s extension at 72 ºC, then eight cycles of 94 ºC denaturation for 30 s, 53 ºC annealing for 45 s and 72 ºC extension for 45 s and a

final extension at 72 ºC for 10 min. The PCR amplification quality and approximate size of the amplicons (using a 100 pb DNA Ladder) were evaluated on 1.5% agarose gel. Tetraplex systems of 10 µL of final volume were prepared, containing 7.8 µL HI-DI formamine, 0.2 µL GeneScam-500 LIZ ® size standard and 2 µL of mixture contained 0.5 µL of each four PCR products (each labeled with a different DYE, Table 2). The samples were denatured for 5 min at 94°C, immediately chilled on ice, then capillary electrophoresis was performed on an ABI 3500 Genetic Analyzer (Applied Biosystems) using POP-7™ polymer (Applied Biosystems), 50 cm capillary arrays and default instrument settings (all Applied Biosystems). GeneMarker® software (V2.6.3, SoftGenetics) was used to visualize, establish the peaks of filtering (RFU Relative Fluorescent Units) and interpretation, define the genotype of each individual, and generate the compilation of genotype data.

3.4 Genome alignment

In order to obtain the physical coordinates of the different markers, the sequences of the markers were downloaded in FASTA format available in the database GenBank of National Center for Biotechnology Information (NCBI) and aligned, using the BLAST (Basic local alignment search tool) (Altschul *et al.*, 1990) algorithm, specifically the BLASTn subprogram, aligning the sequence of the markers with the most recent version (V2.0) of the CRIOLLO genome (Argout *et al*., 2017) and with the last version (V1.0) of the Matina genome (Motamayor *et al*., 2013). An expected value (E-value) of 10 (e^{-10}) was used to obtain the alignment with a lower probability of detecting false positives and a maximum number of 20 aligned sequences to keep. Additionally, an alignment was made of the sequences of the primers (Table. 2) (in silico PCR) of each marker, published by the different authors and validated with the PROBE database of NCBI, with an expected amplicom size between 40 to 1000 pb, and a single mismatch was considered acceptable.

3.5 Analyses of population structure

The population structure of the individuals was estimated through the principal components analysis (PCA). It was necessary to adapt the information resulting from the capillary electrophoresis and the GeneMarker® software, in diploid HapMap (Haplotype Map) PHASE format using the IUPAC (International Union of Pure and Applied Chemistry) nomenclature (Favre *et al.*, 2013), Each allele per locus was encoded as a SNP (Single nucleotide polymorphisms) allele, the information of the physical coordinates of each locus resulting from the alignments made with the BLASTn algorithm was also inclued. This genotype data set was transformed to probabilistic numerical values (homozygous major is 1.0, homozygous minor is 0.0, and heterozygous is 0.5). PCA analysis was performed using a covariance matrix, the number of main components was selected based on the ratio of the number of components and the total variability covered with the addition of a new component.

The level of relatedness among the individuals of the population was obtained by using the CLADOGRAM pluging of TASSEL®, as input, the genotype data set in HapMap PHASE format was used. The distance model between each taxon used by TASSEL® is a modified Euclidean distance, the model treats a homozygote as 100% similar to itself and a heterozygote as only 50% similar to itself (due to the two different alleles are present). Two different approaches were used to analyze the distance matrix: Neighbor-Joining and UPGMA (Unweighted Pair Group Method with Arithmetic mean), the trees generated by the plugin were visualized using the Archaeopteryx tree (Zmasek *et al*., 2011) software.

3.6 Association mapping

The level of correlation between allelic variations and phenotypic variations and the effect of each allele per marker was evaluated on the phenotypic values of each individual using the GLM (General Linear Model) pluging (Casstevens & Wang, 2015) of TASSEL®, a unified matrix with phenotypic, genotypic (transformed to probabilistic values) and structure information was used as input. A test of 1000 permutations was carried out, based on the method proposed by Anderson & Ter Braak (2003), which calculates the predicted and residual values of the reduced model, then permutes the residuals and adds them to the predicted values.

3.7 Retrieving of the protein sequences

Once the physical coordinates of the flanking markers of the QTLs were located, the information contained between each pair of flanking markers and containing theoretically, as reported by the respective authors, each QTL were downloaded. This information was downloaded directly using the services of JBrowser (Skinner *et al.*, 2009) to the CRIOLLO genome, and GBrowser (Stein *et al.*, 2002) to the MATINA genome. The information was retrieved in genomic format GFF3 (General Feature Format) containing all the information in the form of Tracks of the targets regions available for genome. For the CRIOLLO genome, a direct download of the proteins predicted in the regions was also made in FASTA format. For the genome MATINA it was necessary to use a script in Perl language for the Retrieving of the predicted proteins for each genomic region contained in the GFF3 files, and later transform the protein sequences in FASTA format.

3.8 Alignment and functional annotation of complete protein sequences

The protein sequences contained in the FASTA files retrieved from both genomes for the subject genomic regions were aligned against the “nr” eukaryote protein database of the NCBI (NCBInrDB, 2018), downloaded in its most current version (18_08_2018) at the time the alignments were made. The alignment was made on the CACAU server, available at by the NBCGBI (*Núcleo de Biologia Computacional e Gestão de Informações Biotecnológicas*) of the *Universidade Estadual de Santa Cruz* through a command line interface using the BLAST software, specifically the submodule BLASTp, the alignment parameters were used: an expected e-value of 0.001 (-evalue 0.001), a maximum number of the twenty best alignments (-num_alignments 20), the results were requested in XML (Extensible Markup Language) format (-outfmt 5), the number of cores used for analysis (-num_threads) depended on the number of cores available in the server nodes in

which each analysis was allocated, the rest of the parameters were left in default for all analyzes.

3.9 Alignment and functional annotation of protein sequences based on domains

The retrieved FASTA files containing the protein sequences of the genomic regions of the two genomes were analyzed with the InterProScan (Jones *et al.*, 2014) package compendium, version (V5.33-72.0), in its command line interface, it was requested that the results be given in three formats (-f XML HTML TSV) each to be used in subsequent analyzes or visualizations, the number of cores to be used depended on the number available in the nodes of the server where the analyzes were allocated. The package was requested to compare the sequences with the following databases offered by the software:

1) CATH-Gene3D (Lees *et al.*, 2012) database which describes protein families and domain architectures in complete genomes, the protein families are formed using a Markov clustering algorithm, followed by multi-linkage clustering according to sequence identity. Mapping of predicted structure and sequence domains is undertaken using HMMs libraries representing CATH and Pfam domains.

2) CDD (NCBICDD, 2018) is a protein annotation resource that consists of a collection of annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST (Reverse PSI-BLAST). CDD content includes NCBI-curated domain models, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases.

3) MobiDB (Piovesan *et al.*, 2015) is a centralized resource for annotations of intrinsic protein disorder. The database has three levels of annotation: manually curated, indirect and predicted. The different sources present a clear tradeoff between quality and coverage. By combining them all into a consensus annotation,

MobiDB aims at giving the best possible picture of the "disorder landscape" of a given protein of interest.

4) HAMAP (Pedruzzi *et al.*, 2012) stands for High-quality Automated and Manual Annotation of Proteins. HAMAP profiles are manually created by expert curators. They identify proteins that are part of well-conserved proteins families or subfamilies.

5) Pfam (Punta *et al.*, 2012) is a large collection of multiple sequence alignments and HMMs covering many common protein domains.

6) PIRSF (Wu *et al.*, 2004) protein classification system is a network with multiple levels of sequence diversity from superfamilies to subfamilies that reflects the evolutionary relationship of full-length proteins and domains.

7) PRINTS (Attwood *et al.*, 2012) is a compendium of protein fingerprints.

8) ProDom (Bru *et al.*, 2005) protein domain database consists of an automatic compilation of homologous domains. ProDom are built using a novel procedure based on recursive PSI-BLAST searches.

9) PROSITE (Sigrist *et al.*, 2012) is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family a new sequence belongs.

10) SFLD (Structure-Function Linkage Database) (Akiva *et al.*, 2014)is a hierarchical classification of enzymes that relates specific sequence-structure features to specific chemical capabilities.

11) SMART (a Simple Modular Architecture Research Tool) (Letunic *et al.*, 2012) allows the identification and annotation of genetically mobile domains and the analysis of domain architectures.

12) SUPERFAMILY (De Lima Morais *et al*., 2011)is a library of profile HMMs that represent all proteins of known structure. The library is based on the SCOP (Murzin *et al.*, 1995) classification of proteins: each model corresponds to a SCOP domain and aims to represent the entire SCOP superfamily that the domain belongs to.

13) TIGRFAMs (Haft *et al.*, 2012) is a collection of protein families, featuring curated multiple sequence alignments, HMMs and annotation, which provides a tool for identifying functionally related proteins based on sequence homology. the rest of the parameters were left in default for all analyzes.

3.10 Functional annotation in Gene Ontology language

For assignment of terms in Gene Ontology (GO) (Ashburner *et al.*, 2000) language using his latest version (TGOC, 2018) the BLAST2GO software was used (Conesa *et al*. 2005) which classifies and assigns terms to the functions of the protein sequences in three large groups: biological process, molecular function and cellular component. The XML files generated by InterProScan and BLASTp were used as input, generating a mixed data set with the results of the two software, containing, among other information, the functions of the orthologous proteins found. A mapping was carried out in order to retrieve the GO terms associated to the functions found with the alignment softwares and finally an annotation of reliable functions.

3.11 Classification of genes associated with resistance to pathogens in plants

The files in tabulated TSV format generated by the InterProScan were used as input for a software based on Perl language and developed by Raner J. S. Silva (2018, personal information) , which uses the TSV annotated files to classify the results of genes associated with resistance to pathogens in thirteen gene classes: 1) Proteins with Coiled-coils (CC) and nucleotide binding sites (NBS) domains, 2) Proteins with CC, NBS and Leucine-rich repeat (LRR) domains, 3) Proteins with Mlo-like resistance proteins, 4) Proteins with just NBS domain, 5) proteins with NBS and LRR domains, 6) Proteins with Receptor-like Kinase (RLK), 7) Proteins with RLK and Ginkbilobin2 (GNK2) domains, 8) Proteins with Receptor-like without the Kinase domain, 9) Proteins with Resistance to Powdery Mildew 8 (RPW8) NBS and LRR domains, 10) proteins with Toll/interleukin-1 receptor (TIR) domain, 11) Proteins with TIR and NBS domains, 12) proteins with TIR, NBS and LRR domains and 13) (UNKNOWN) Proteins with Leucine-rich repeat (LRR) domains that do not fit any

other class. Files for each gene class containing the codes of the genes in their respective genomes of origin were generated.

Table 1. Characteristics of plants materials. (ID) Identifier of individuals, (V) Variety, (FN) Field number, (RL) Raistance level, (PC) Phenotypic class, (PC(S)) Phenotypic class (Simplified).

ID	V	FN	RL	PC	PC(S)
MA_R_4131	Maranhão	4131	1.325	a1	R
MA_R_4118	Maranhão	4118	1.475	a1	R
MA_R_4130	Maranhão	4130	1.775	a1	R
MA_R_4116	Maranhão	4116	1.8	a1	R
MA_R_4117	Maranhão	4117	1.8	a1	R
MA_R_4134	Maranhão	4134	1.8	a1	R
CO_R_4070	Comum	4070	1.575	a1	R
CO_R_4076	Comum	4076	1.4	a1	R
CO_R_4183	Comum	4183	1.55	a1	R
CO_R_4060	Comum	4060	1.625	a1	R
CO_R_4090	Comum	4090	1.825	a1	R
CO_R_4178	Comum	4178	1.925	a1	R
PA_R_4016	Pará	4016	1.725	a1	R
PA_R_4152	Pará	4152	1.85	a1	R
PA_R_4047	Pará	4047	1.95	a1	R
PA_R_4154	Pará	4154	1.95	a1	R
PA_R_4007	Pará	4007	2	a2	R
PA_R_4038	Pará	4038	2	a2	R
MA_S_4113	Maranhão	4113	3.55	a4	S
MA_S_4123	Maranhão	4123	3.55	a4	S
MA_S_4126	Maranhão	4126	3.2	a3	S
MA_S_4111	Maranhão	4111	3.075	a3	S
MA_S_4104	Maranhão	4104	3.05	a3	S
MA_S_4120	Maranhão	4120	3.025	a3	S
CO_S_4058	Comum	4058	4.45	a5	S
CO_S_4189	Comum	4189	4.3	a5	S
CO_S_4078	Comum	4078	3.725	a4	S
CO_S_4089	Comum	4089	3.525	a4	S
CO_S_4062	Comum	4062	3.45	a4	S
CO_S_4190	Comum	4190	3.4	a4	S
PA_S_4012	Pará	4012	3.625	a4	S
PA_S_4043	Pará	4043	3.4	a4	S
PA_S_4036	Pará	4036	3.225	a3	S
PA_S_4160	Pará	4160	3.1	a3	S
PA_S_4017	Pará	4017	3.025	a3	S
PA_S_4005	Pará	4005	2.975	a3	S
S_Cat	Catongo	-	2.969	a3	S
MS_TSH	TSH1188	-	2.824	a3	MS
R_SCA	SCA6	-	1.464	a1	R
MR_SIC	SIC23	-	2.6	a2	MR

Table 2. Characteristics of SSR loci. (LG) Linkage group, (QTL) quantitative trait loci, (RS) repeated sequence, (DYE) fluorescents used in genotyping, (*) Loci that did not amplify.

Marker Name	Marker Reference	LG	QTL	QTL Reference	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	RS	DYE
mTcCIR6	LANAUD, C. *et al*,1999	6	PRR/100/6.1 q1.BP-Pct	Akaza, M. *et al.* 2016 Barreto, M. et al 2018	TAAAGCAAAGCAATCTAACATA	TTCCCTCTAAACTACCCTAAAT	(TG)7(GA)13	PET
mTcCIR9	LANAUD, C. *et al*,1999	6	PRR/H/6 q1.BP-Pp	Akaza, M. *et al.* 2016 Barreto, M. et al 2018	ACATTTATACCCCAACCA	ACCATGCTTCCTCCTTCA	(CT)8N15(CT)5N9(TC)10	6-FAM
mTcCIR37	USDA	10	Phytoph3	Brown, J. *et al.* 2007	CTGGGTGCTGATAGATAAT	AATACCCTCCACACAAAT	(GT)15	VIC
mTcCIR61	CIRAD/USDA	10	Phytoph3	Brown, J. *et al.* 2007	GCAGTCTGAAACAAGATAA	TGACTATAATATAAGGCGAA	(CA)18	NED
mTcCIR77*	Pugh T. *et al* 2004	10	FOL/100/10	Akaza, M. *et al.* 2016	GTTCTCCCCCACTCTCT	AATAAATAAATAAACAATACG	(TC)9	-
mTcCIR81	Pugh T. *et al* 2004	3	q3.BP-Pc FOL/100/3	Barreto, M. et al 2018 Akaza, M. et al. 2016	ACAATCTGTCCATTATTCTG	TGAAACTCCCATACTACTGA	(CT)15	PET
mTcCIR95	Pugh T. *et al* 2004	4	q4.BP-Pc PRR/100/4	Barreto, M. et al 2018 Akaza, M. et al. 2016	CATCGTCTTCCTCTCATC	CTCCTTCCCTTTCTCTC	(TC)4 CC	6-FAM
mTcCIR118	Pugh T. *et al* 2004	1	PRR/H/1	Akaza, M. *et al.* 2016	TCTGCCTGAAAATGTCTC	TGGGGCACTAACTTTTG	(GA)10	VIC
mTcCIR131	Pugh T. *et al* 2004	3	q3.BP-Pc	Barreto, M. *et al 2018*	GATCATCGGTAAAGTAAAAT	TGAGTAAGAAAAAGTAGAAAA	(GA)9 C (GA)4	NED
mTcCIR136	Pugh T. *et al* 2004	6	PRR/100/6.1 q1.BP-Pct	Akaza, M. *et al.* 2016 Barreto, M. et al 2018	GAGGAGGTGAGAGCCA	GGTTTGTATTTTTGATTGAG	(GA)7 GC (GA)7	PET
mTcCIR152	CIRAD Data Base	2	q2.BP-Pc	Barreto, M. *et al 2018*	CAGTAGTCAAAACATCAAA	GTAATCCAAATAATAAGCAT	(TC)9 CC (TC)15	6-FAM
mTcCIR168	Pugh T. *et al* 2004	4	Phytoph1	Brown, J. *et al.* 2007	GGTAGTATTGAGGTGCGTAT	GTGAATGAATGGATGTGAAA	(TC)9	VIC
mTcCIR176	Pugh T. *et al* 2004	2	q2.BP-Pc	Barreto, M. *et al 2018*	TCACCAATTCTCTGCTC	AATGAAATTACCTCCTTAC	(TG)16	NED
mTcCIR183	Pugh T. *et al* 2004	4	PRR/95/4	Akaza, M. *et al.* 2016	GTTATCTTAGTTTCTAGCCAC	GTAGTCTTACACCTTGATTG	(AC)9	6-FAM
mTcCIR184	Pugh T. *et al* 2004	1	PRR/H/1	Akaza, M. *et al.* 2016	GGTTTTCTAGCTCCTCC	AGGAAAGAATGACTCATACTA	(CA)8 (CT)13	PET
mTcCIR200	Pugh T. *et al* 2004	8	Phytoph2	Brown, J. *et al.* 2007	CTTAAAATAAGCCCAAATAC	GCCAATTTCTGACCCA	(TG)8	VIC
mTcCIR202*	Pugh T. *et al* 2004	3	q3.BP-Pc	Barreto, M. *et al 2018*	CCTGAGTCAAAGTGTCCT	TCTCTCATAGCTCAAGCA	(TG)7 (GA)9	-
mTcCIR208*	Pugh T. *et al* 2004	6	PRR/100/6.2	Akaza, M. *et al.* 2016	GCAAGCCCCTAAAACT	AAAAAGCAAAAGAAGAAGA	(CT)10-25pb-(AC)11	-
mTcCIR213	Pugh T. *et al* 2004	4	PRR/95/4	Akaza, M. *et al.* 2016	GATCTCGCAAAACTAACA	TAAGTAAAATGAAGGTGTGA	(CT)26	NED
mTcCIR225*	Pugh T. *et al* 2004	8	Phytoph2	Brown, J. *et al.* 2007	AAGACAAAGGGAAGAAGA	AGGGGAAGAGCAAATC	(TC)10	-

mTcCIR237	Pugh T. *et al* 2004	4	q4.BP-Pc	Barreto, M. *et al 2018*	GAAGACAAGGATGGAGACT	GCAAAGAGAGCAGGAGA	(TC)10	PET
mTcCIR240	Pugh T. *et al* 2004	2	PRR/95/2	Akaza, M. *et al.* 2016	AGTGATTTATGGGACTTT	CATACCTACTACTGCTCTCT	(CT)22	6-FAM
mTcCIR241*	Pugh T. *et al* 2004	4	PRR/100/4	Akaza, M. *et al.* 2016	ACGAGTGAGAGAGTGAAGTT	CAGTTGGAGGGCATTT	(CT)23	-
mTcCIR252*	Pugh T. *et al* 2004	2	PRR/95/2	Akaza, M. *et al.* 2016	AATGTGTGCTTTGTTTCTA	TTCAAGGGCGTAACTC	(AC)10	-
mTcCIR255	Pugh T. *et al* 2004	6	q1.BP-Pp	Barreto, M. *et al 2018*	TTTACCTCCACCATCTT	TGGCACTTATCTATTACTGT	(AC)11	VIC
mTcCIR268	Pugh T. *et al* 2004	2	q2.BP-Pc PRR/95/2	Barreto, M. *et al 2018* *Akaza, M. et al. 2016*	CAGTGAAGAGGCAAGAGA	TGTAATCCAAATAATAAGCAT	(GA)17 GG (GA)9	NED
mTcCIR273	Pugh T. *et al* 2004	1	FOL/100/1 q1.BP-Pc	Akaza, M. et al. 2016 Barreto, M. et al 2018	ACGGCATTAGAGAGAGA	AGAATGATCGCAGAGAG	(CT)4 AC (CT)13 TT	PET
mTcCIR275	Pugh T. *et al* 2004	1	q1.BP-Pc	Barreto, M. *et al 2018*	GGTTTGGTTTTGGTAAGAC	TAAGAGAGAGTGATGCTGACA	(CT)11	6-FAM
mTcCIR276*	Pugh T. *et al* 2004	6	PRR/100/6.1	Akaza, M. *et al.* 2016	GTCCTATCTGCCTCACT	TCCTGCTTTTTAATACAT	(GA)14	-
mTcCIR282	Pugh T. *et al* 2004	8	PRR/H/8	Akaza, M. *et al.* 2016	AGCAAAGGCAATAATAATG	TGGTGAGGGGAGAGAA	(GA)2 GG (GA)6	VIC
mTcCIR291	Pugh T. *et al* 2004	6	PRR/H/6 PRR/100/6.2	Akaza, M. *et al.* 2016 Akaza, M. et al. 2016	AGTCCCATAGGTTCCAAT	CGAGGTTATCCCCAAA	(CT)12	NED
mTcCIR337	Fouet, O. *et al. 2011*	6	PRR/H/6	Akaza, M. *et al.* 2016	ACGAAGCCGTAACTTGG	TGCAGGACTCTCTGTCACT	(GGA)5	PET
mTcCIR343	Fouet, O. *et al. 2011*	4	PRR/100/4	Akaza, M. *et al.* 2016	GCTTTGCCCTTTCTTCTCT	AGCACTGAACCGAGCAA	(AT)10	6-FAM
mTcCIR410	Fouet, O. *et al. 2011*	3	FOL/100/3	Akaza, M. *et al.* 2016	TTTTGCCTCCCTTGTTCT	TGAATCGTTGAGCGAAAG	(TA)10	VIC
mTcCIR422	Fouet, O. *et al. 2011*	1	FOL/100/1	Akaza, M. *et al.* 2016	ACATCCTTTTCTCTGCCTTT	CCCTTCCCATCCCTCT	(GA)18	NED
mTcCIR444	Fouet, O. *et al. 2011*	8	PRR/H/8	Akaza, M. *et al.* 2016	TGAACCGGATTGTTGGA	GGGACTTAATCTGGACATGC	(TC)15	PET

4 RESULTS AND DISCUSSION

4.1 SSR analysis

A total of 39 SSR markers were amplified in order to evaluate the polymorphisms in three local cocoa varieties. The SSR markers used in the present study represent 80.5% of the total markers proposed by different authors in order to study the QTLs associated with BP resistance in cocoa (Akasa *et al.,* 2016; Barreto *et al.*, 2018 and Brown *et al.,* 2007).

The amplicons product of the PCRs were genotyped in an ABI 3500 Genetic Analyzer (Applied Biosystems). For the 1160 loci (29 markers x 40 individuals), only 44 did not amplify (data not shown), with an average of 1.1 loci per individual without amplification, which generated enough genotypic information for subsequent analyzes. The criterion for identifying the quality of the amplification and alelic profile was defined in GeneMarker® software (V2.6.3, SoftGenetics), in which a minimum RFU (Relative Fluorescent Units) cutoff criterion of 150 was used as proposed by Flores and Krohn (2013), with well defined peaks (Figure 1), avoiding flattened peaks of great amplitude, which could generate errors in the identification of the alleles. The standard deviation in the variation of the amplicons attributed to the same allele was evaluated, being 4.5 bp and a difference in length between the alleles of the same locus varying between 10-35 bp. From the 29 amplified loci, all were polymorphic, the number of alleles per locus range from two (mTcCIR184, mTcCIR273 and mTcCIR237) up to six alleles per locus (mTcCIR410 and mTcCIR37) (Table 3). The proportion of heterozygous individuals per locus varying from 0 to 0.9 (with an average of 0.43), which is also 0 (mTcCIR273) to 36 (mTcCIR337) (average of 16 individuals) of the 40 individuals

heterozygous for the same locus (Table 3). Relatively low values were obtained, for an unstructured free-cross population; a result that contrasts with what was expected due to the high diversity reported by some authors (Bartley, 2005; Bennett, 2003; Motamayor *et al.*, 2008), but being a result homologous to the low diversity reported by Santos *et al.* (2015), a study that included individuals from the population of Uruçuca as well, therefore the present study confirms these findings.

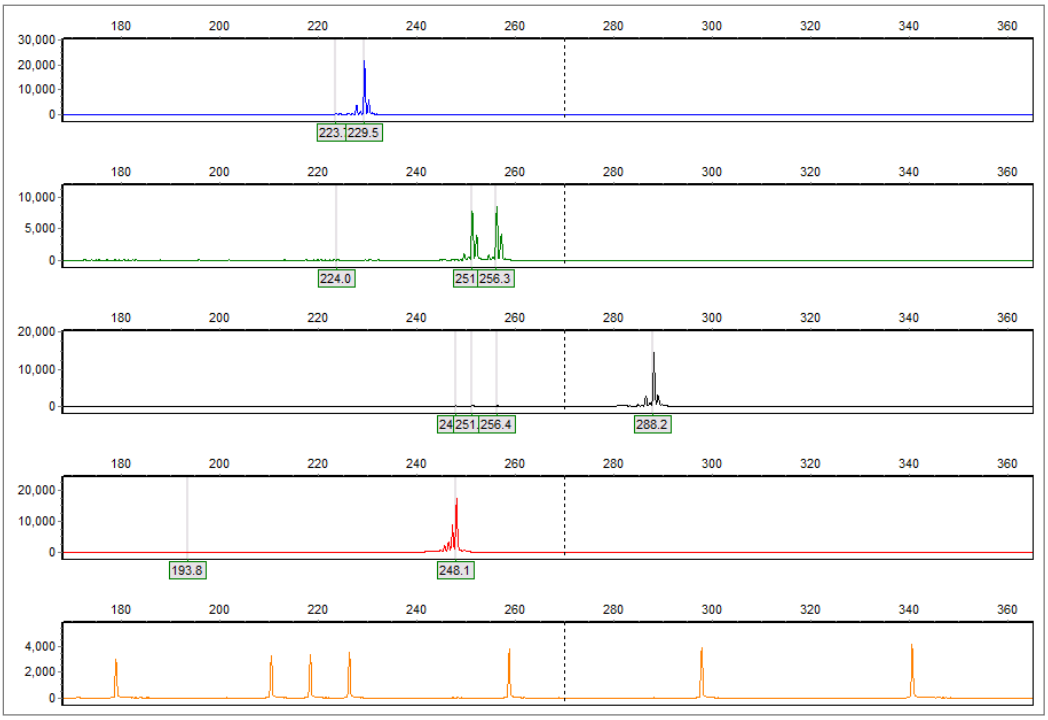

Figure 1. Visualization of amplicons in tetraplex. Sample 4123, markers: mTcCIR273 (PET-Red), mTcCIR275 (6-FAM-Blue), mTcCIR282 (VIC-Green) and mTcCIR291 (NED-Yellow)

The low diversity observed in the present study is probably associated to the origin of the varieties. All three varieties evaluated are direct descendants of varieties of Amelonado, which also has a low genetic diversity described by different authors (Lanaud *et al.*, 1987; Motamayor *et al.*, 2003). Santos et al.(2015) also reported a high positive index of fixation, explained by its mating system, which like other traditional cultivars is self-compatible (Aikpokpodion *et al.*, 2009; Motilal *et al.*, 2013;Santos *et al*. 2015, Sereno *et al.*, 2006).

Historically, the local cocoa varieties of the Bahia, come from a limited number of seeds (Santos *et al.*, 2015) and limited number of genotypes were introduced for more than 200 years. Both factors have increased the levels of inbreeding of these varieties.

Table 3. Alellic characteristics of SSR loci.

Marker	Major Allele Frequency	Minor Allele Frequency	Allele Number	Number Heterozygous	Proportion Heterozygous
mTcCIR184	0.9375	0.0625	2	5	0.125
mTcCIR118	0.6	0.325	4	30	0.75
mTcCIR273	0.975	0.025	2	0	0
mTcCIR422	0.5625	0.375	5	29	0.725
mTcCIR275	0.7625	0.1625	4	17	0.425
mTcCIR240	0.95	0.05	2	4	0.1
mTcCIR268	0.6625	0.2	5	11	0.275
mTcCIR152	0.55	0.35	4	32	0.8
mTcCIR176	0.7375	0.2375	3	19	0.475
mTcCIR410	0.775	0.1125	6	6	0.15
mTcCIR131	0.6625	0.225	4	21	0.525
mTcCIR81	0.6625	0.3125	3	9	0.225
mTcCIR168	0.4375	0.3375	3	31	0.775
mTcCIR213	0.625	0.175	4	13	0.325
mTcCIR183	0.5125	0.2375	4	24	0.6
mTcCIR95	0.475	0.4	5	31	0.775
mTcCIR237	0.9875	0.0125	2	1	0.025
mTcCIR343	0.5	0.45	4	8	0.2
mTcCIR6	0.5375	0.3	4	3	0.075
mTcCIR136	0.45	0.3625	4	26	0.65
mTcCIR255	0.85	0.1	3	9	0.225
mTcCIR337	0.5	0.425	3	36	0.9
mTcCIR9	0.725	0.125	5	6	0.15
mTcCIR291	0.675	0.1625	4	18	0.45
mTcCIR282	0.5125	0.2125	5	29	0.725
mTcCIR444	0.2375	0.225	5	17	0.425
mTcCIR200	0.5	0.325	4	34	0.85
mTcCIR61	0.7125	0.125	4	5	0.125
mTcCIR37	0.65	0.125	6	8	0.2

4.2 Structure and relatedness

The analysis of the relatedness among the individuals of the three varieties, including each group phenotypically contrasting individuals, was carried out by three approaches, Neighbor Joining (NJ) and UPGMA, both from distance matrices, and a Principal Components Analysis (PCA). For the NJ analysis (Figure 2) the formation of 9 main groups can be observed, not very different from the 8 main groups generated by the UPGMA analysis (Figure 3), and in general the groups contain almost the same individuals, which shows that genetic recombinations between ancestry was greater within the 8 or 9 main groups formed than it was between these general groups. There were certain similarities between the two phylogenetic analyzes, such as the formation of almost pure groups of resistant Maranhão in the two trees G2 NJ and G4 UPGMA. It can be seen that the three individuals that make up the G2 in NJ are present in the G4 of UPGMA, and can be considered as the same group in the methods, appreciating in the UPGMA tree, that the two remaining resistant Maranhão individuals are in an immediate group (G1 UPGMA), highlighting the grouping of all resistant Maranhão individuals into two contiguous groups for the method, showing a clear separation in the method for the contrasting phenotypes in the Maranhão variety, grouping relatively together all the resistant individuals and the susceptible individuals were distributed in the other groups. In the G1 of NJ a differentiated grouping according to the study phenotype can be observed, forming two groups, one with 6 only resistant individuals (4 Comum and 2 Pará) and the second subgroup only with susceptible individuals with representatives of the three varieties (3 Maranhão, 1 Comum and 1 Pará), again with a tendency for the Maranhão group to group differentially, if these subgroups are compared with the UPGMA tree, it can be seen that two of the resistant Comum individuals (4090 and 4178) and two of the resistant Pará individuals (4016 and 4047) make up an exactly equal subgroup of the G7 in the UPGMA tree, and the other subgroup of only susceptible individuals of the NJ G1 also has an almost exact homologous grouping in an internal branch of the G7 in UPGMA. In the NJ tree, the formation of a G3 basal group composed of only susceptible individuals from Comum and Pará can also be observed.

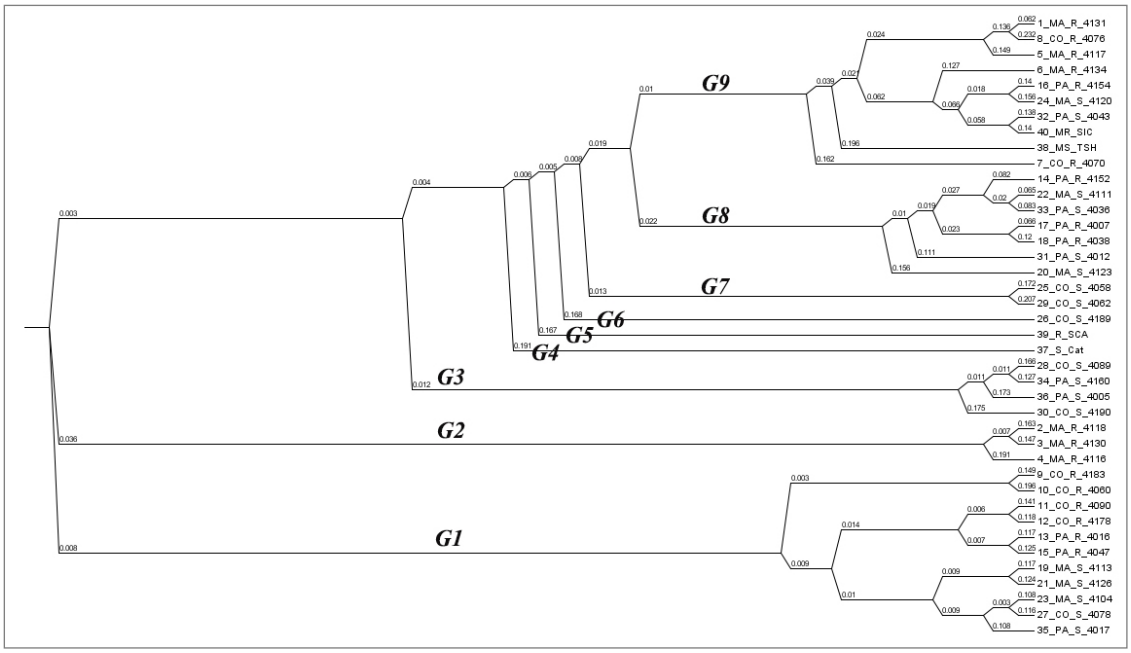

Figure 2. Neighbor Joining Phylogenetic Tree.

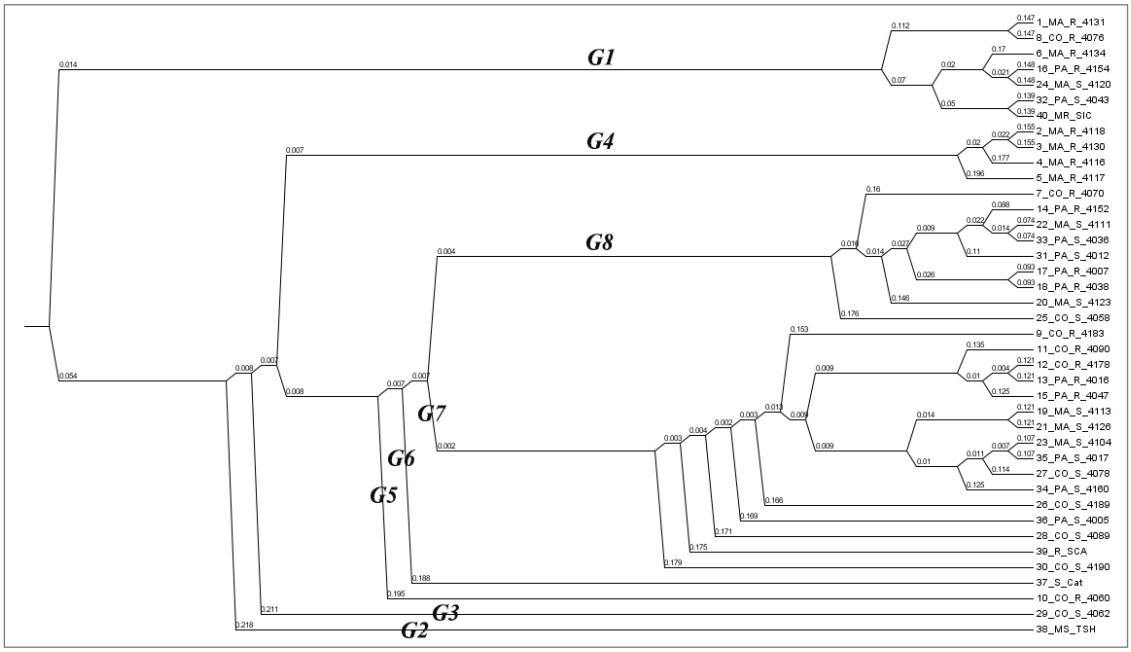

Figure 3. Unweighted Pair Group Method with Arithmetic Mean Phylogenetic Tree.

For the Principal Components Analysis, it was first evaluated what would be the minimum number of components to cover the greatest amount of variability, leaving only the components with the greatest influence on the distribution of the variables. They were reduced to two main components based on the comparison of the amount of total proportion of the variability covered with the inclusion of new components (Figure 4), with the decrease of the components to two, more than half (0.52) of the total proportion of the variability is included.

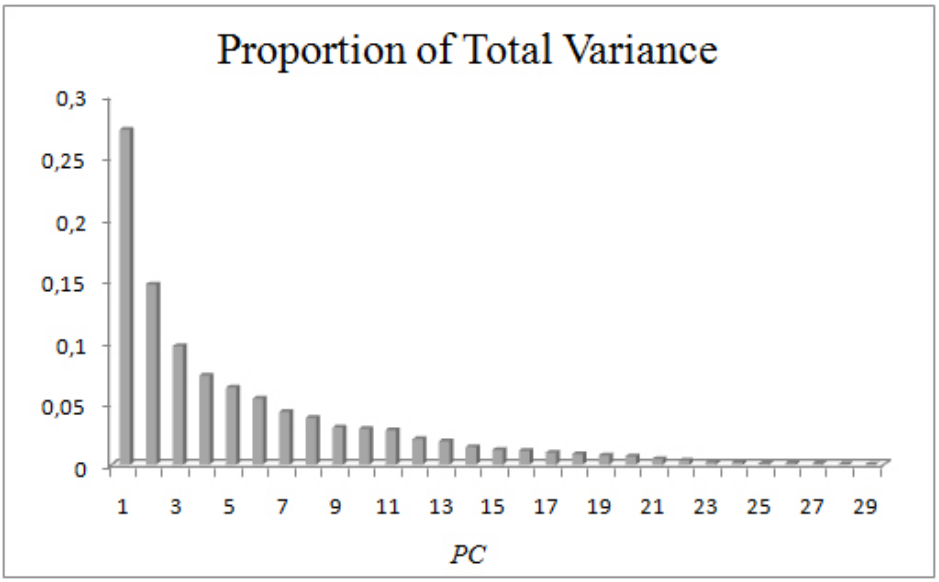

Figure 4. Proportion of Total Variance vs Number of Principal Components

Once the number of main components was selected, the two-dimensional distribution of the individuals was plotted (Figure 5). It can be seen that there is no clear formation of pure groups, but by comparing the formed subgroups, several homologies can be seen with the NJ and UPGMA phylogenetic trees. The G1 of the PCA analysis contains the same two individuals (4090 and 4178) of the Comum variety and resistant phenotype grouped together in both the NJ (G1) tree and a subgroup of the G7 of the UPGMA tree. The G2 group contains three Maranhão resistant individuals (4130, 4116 and 4118) that correspond exactly to the three individuals of the variety and phenotype grouped in the G2 of NJ and are 3 of the 4 grouped in the G4 of UPGMA. The G3 includes exactly the same two individuals grouped in the Pará resistant subgroup (including noise from a

susceptible Maranhão individual) in the G1 NJ and a subgroup of G7 that includes that variety and phenotype in UPGMA. G4 (containing G3 and G1) contains individuals grouped very closely in groups 7 of UPGMA and 1 of NJ. The G5 contains three susceptible individuals (4104, 4078 and 4017) grouped very closely in a sub-branch of the G1 of NJ.

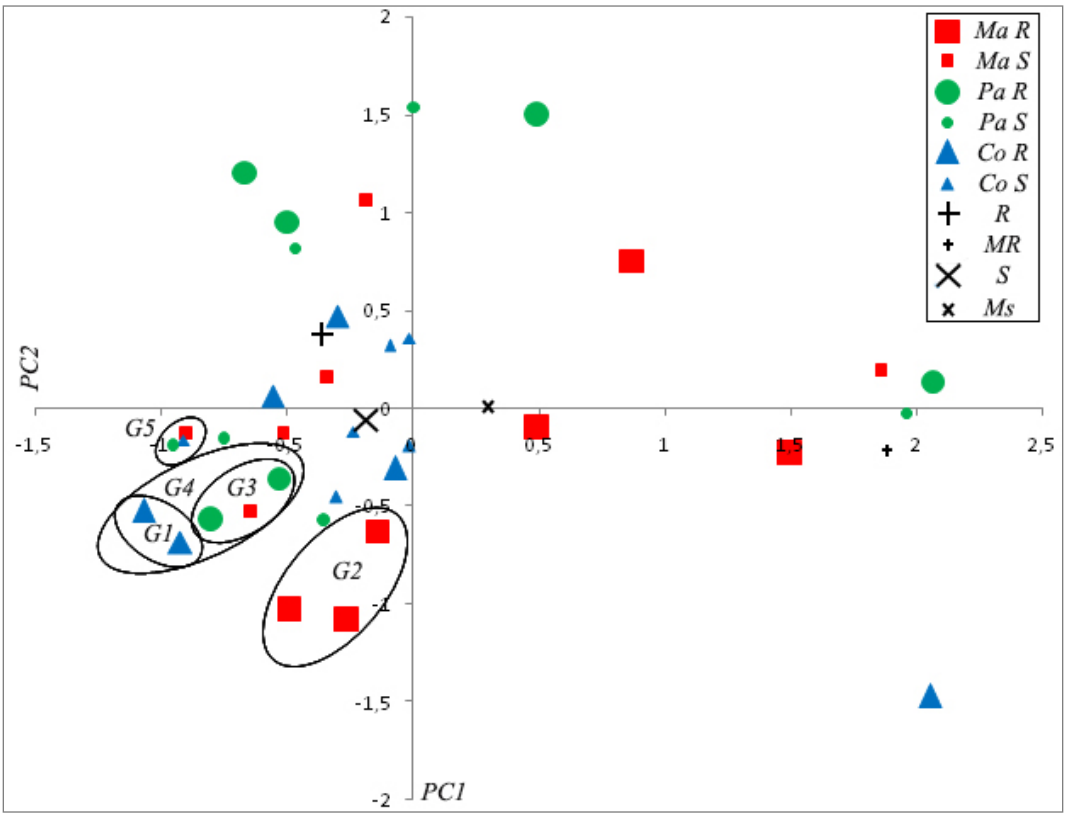

Figure 5. Principal Component 2D Plot. (PC1) Principal Component 1, (PC2) Principal Component 2, (Ma R) Maranhão Resistant, (Ma S) Maranhão Susceptible, (Pa R) Pará Resistant, (Pa S) Pará Susceptible, (Co R) Comum Resistant, (Co S) Comum Susceptible, (R) Resistant, (MR) Moderately Resistant, (S) Susceptible, (Ms) Moderately Susceptible

As the varieties have been plants for more than 200 years mixed, the crosslinks made that they are not currently varieties so genetically distant, despite still possessing very distinctive morphological characteristics (Rhodes, 2016), such as the shape of the fruit and the taste of the chocolate they produce (Santos *et al*., 2015). The differentiation of these varieties can be seen in the separation of individuals from foreign cultivars such as TSH118 (G2 UPGMA), this was recorded

by Santos *et al.* (2015) that showed that a more noticeable separation could be appreciated with non-local clones and cultivars than among local varieties, not forming differentiated groups among local varieties. When comparing the results of the three analyzes, a marked tendency is observed to the differentiated grouping of the Maranhão variety, specifically the resistant phenotypes, inferring that this group of plants is genetically more distant from the other two varieties. In general, several pure subgroups are denoted in the most basal branches of the trees, indicating the existence of varietal genetic subgroups, mainly for Maranhão and the varieties Comum and Pará being more difficult to differentiate genetically, which can be extrapolated to the difficulty to differentiate morphologically these two varieties in the field.

4.3 Association Mapping

The Association Mapping analyze was performed using the phenotypic information of Rhode (2016), the genotypic data of 29 SSR markers and the result of the analysis of PCA structure. The association of the SSR markers information with the phenotypic data and weighted by the structure, resulted in the identification of four significant loci for resistance to BP at a significant theshold (p) $\leq$0.05by using a GLM approach (Table 4). The theshold (p) was decided based on a QQ (Quantile-Quantile) test (Figure 6) where it can be seen that the correlation between the phenotypic values (PP_R_S) and the prediction of the values by the GLM analysis remain overlapped until approximately 0.15 of e-value, therefore a theshold < to 0.05 represents a high reliability.

Table 4. SSR Loci Associated to resistance to *Phytophthora palmivora.* (LG) Linkage Group, (Perm p) e-value permutations test, (R^2 (%)) Percentage of phenotypic variance explained by de locus, (add p) e-value Additivity test, (dom p) e-value Dominance test. (marker df) Marker degrees of freedom.

Marker Name	LG	Probability of marker (Marker p)	Perm p	R^2 (%)	add_p	dom_p	marker_df	minorObs
mTcCIR444	8	0.00015268	0.001	7.4312	2.6313E-4	3.5095E-6	7	7
mTcCIR200	8	0.005	0.069	3.7166	0.08077	0.00748	5	10
mTcCIR268	2	0.04508	0.521	2.7114	0.02145	0.91435	5	8
mTcCIR81	3	0.0489	0.556	1.9224	0.05314	0.15503	3	9

Two associated markers were detected in the LG8, mTcCIR444 with an e-value (p) of $1.5e^{-4}$ explaining a 7.43% (R^2%) of the phenotypic variation and mTcCIR200 with a p = 0.005 explaining a 3.72% of the phenotypic variation, being the markers with the highest level of association to the phenotypic variations. Additionally, one marker was identified in LG2, mTcCIR268 (p = 0.045508) explaining 2.71% of the phenotypic variation and one in LG3, mTcCIR81 with a value of p = 0.0489 explaining 1.92% of the phenotypic variation.

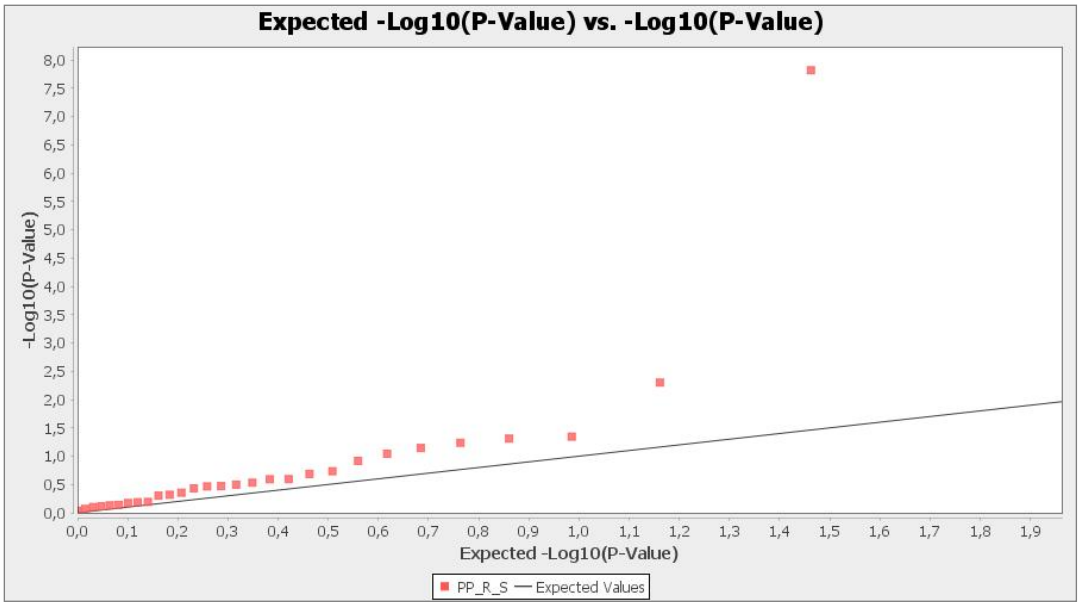

Figure 6. Quantile-Quantile Test of Phenotypic Data.

The mTcCIR444 marker showed, a significant value ($3.5095e^{-6}$) for the dominance test (f) compared to the additivity test. This QTL locus was also reported by Akasa*et al.* (2016) with a R^2% = 13.2 at a cross between (SCA6xH) xC1. The differences observed in the estimated % of the phenotypic variation explained by the locus can be due to two main reasons 1) the genetic difference of the cultivars where the QTL was identified with the genotypes of this study and 2) To the phenotyping methodology that was used, the author used the variable PRR, which was quantified as the accumulated percentage of rotten fruits in relation to the total number of fruits during three years of harvest, in contrast to the foliar disc

inoculation method used by Rhodes (2016). Table 5 shows the estimated contribution of locus allele pairs, where favorable alleles contribute significantly to the resistance, the favorable major allele (194 bp) in homozygosis contributes in the decrease of the susceptibility in -1.29 uni (based on the scale proposed by Nyassé *et al.* (2002) (0 maximum resistance and 5 maximum susceptibility). A second major allele (206 bp) in homozygosis contributing in -1.69 uni was detected and together both alleles -1.49 uni, information that validates the dominance/additivity test.

Table 5. Estimated contribution, length of alleles (pb) and number of observations of loci associated with resistance

Marker	LG	Pos	Obs	Allele	Estimate	Allele size (pb)
mTcCIR444	8	2251862	5	A	-1.2851E0	194
mTcCIR444	8	2251862	2	C	-6.6662E-1	213
mTcCIR444	8	2251862	7	G	0.63938	230
mTcCIR444	8	2251862	2	T	-1.6195E0	206
mTcCIR444	8	2251862	7	-	-3.9552E-1	-
mTcCIR444	8	2251862	3	S	0.15355	-
mTcCIR444	8	2251862	9	W	-1.4686E0	-
mTcCIR444	8	2251862	5	Y	0	-
mTcCIR200	8	4552522	3	A	-7.4174E-1	293
mTcCIR200	8	4552522	1	C	-2.5248E0	302
mTcCIR200	8	4552522	1	T	1.40417	282
mTcCIR200	8	4552522	1	+	-2.4E0	-
mTcCIR200	8	4552522	10	M	-1.2796E0	-
mTcCIR200	8	4552522	24	W	0	-
mTcCIR268	2	8244191	21	A	-7.2632E-1	367
mTcCIR268	2	8244191	2	G	0.10466	371
mTcCIR268	2	8244191	4	T	-1.3351E0	350
mTcCIR268	2	8244191	2	+	-1.3891E0	-
mTcCIR268	2	8244191	3	M	-5.9507E-2	-
mTcCIR268	2	8244191	8	W	0	-
mTcCIR81	3	36293182	22	A	-8.7655E-1	296
mTcCIR81	3	36293182	8	T	-7.0592E-2	317
mTcCIR81	3	36293182	1	+	0.66546	-
mTcCIR81	3	36293182	9	W	0	-

The second most significant locus, mTcCIR200 also located in LG8, was only reported by Brown *et al.* (2007) with a $R^2\%$ = 7.3, the phenotyping method (PRR) used and the populations also differ (Poun7xUF273) from those used in this study, this locus showed more significant values for the dominance test (0.00748) compared to the additivity test (0.08077), favorable alleles contribute significantly in resistance, the favorable major allele (293 bp) in homozygosis contributes in the decrease of the susceptibility in -0.7 uni and the second major allele (303 bp) in homozygosis contributing in -2.52 uni and when together the major and second major allele -1.28 uni, as for the mTcCIR444 marker, the dominance and additivity test showed consistency. The mtcCIR268 locus located in LG2 reported by Akasa *et al.* (2016) with a $R^2\%$ = 19.3 and by Barreto *et al.* (2018) with $R^2\%$ = 2.1, this last value being more consistent with that reported in this study, this could be due to the use of the same phenotyping methodology since this study was conducted in an F1 population of TSH1188xCCN51, both cultivars relatively genetically distant from the local varieties of Bahia. Flament *et al.*(2001) reported a QTL in this same LG explaining 11% of the phenotypic variation in the T60/887xIFC2 cross, IFC2 is a amelonado of the low Amazon that could be a little more related to the local varieties of Bahia, because of the type of molecular marker used by this author, AFLP cannot be sure with certainty if that QTL is located in the same locus as the one validated in this study, similarly using information from the adapters and selective nucleotides used by the author and realizing an alignment with BLAST software in the genome CRIOLLO, It was determined that the region where the QTL is found reported by Akasa *et al.* (2016), Barreto *et al.* (2018) and validated in the present study, has regions rich in the cleavage sequence associated with the enzymes used by said author, which could be an indication of the homology of the QTL with that of the other authors. For the analysis of the estimated contribution of the alleles of this locus, the favorable major allele (367 bp) in homozygosis contributes in the decrease of the susceptibility in -0.72 uni and the second major allele (350 bp) in homozygosis contributing in -1.34 uni, a greater reliability was identified in the additivity test (0.021) compared to the dominance test (0.91), with

an increase in the favorable contribution towards resistance when the alleles are together.

For the locus mTcCIR81 located in the LG3 was also reported by Akasa *et al.* (2016) with a R^2% = 14.5 and by Barreto *et al.* (2018) with R2% = 1.78, again being the reporter value by Barreto *et al.* (2018) more concordant with the finding of this study, Flament *et al.*(2001) also reported a QTL in this LG explaining 9% of the phenotypic variation, but the type of marker used, AFLP, by the authors there is a high level of uncertainty of whether it is located in the same locus reported by the different studies and validated in this work , and when trying to find rich regions with cleavage sequences, no rich regions were found and physically close to the location of the QTL. Risterucci *et al.* (2003) reported a QTL associated with resistance to *P. megakarya*, in the same LG explaining 11.5% of the phenotypic variation and an expected effect of -0.25, the study does not provide information on the primers or selective nucleotides used, but in the linkage map presented by the author, the QTL is located in the initial LG zone between 3.7 cM and 8.1 cM, comparing this information with the results of the physical location obtained in this study for that locus (mTcCIR81) in the CRIOLLO genome, the marker was located in the average location (single coordinate obtained from the equidistance of the location of the primers) in 36,293,182 bp, and the genome reports a total length of chromosome 3 of 36,364,294, which places it in a terminal area of the chromosome (the beginning or end of a LG can be relative depending on how the genomic data are presented, for each DNA strand, 5'-3' = Watson = Sense = + or 3'-5 '= Crick = Antisense = -) which increases the probability of referring to the same locus.

4.4 Genomic localization

The physical genomic coordinates of the 36 flanking markers to the 20 QTLs reported by Akasa *et al.* (2016), Barreto *et al.* (2018) and Brown *et al.* (2007) were obtained. It was observed a significant difference in the number of bp of the coordinates when compared between the two genomes. This fact can be attributed mainly to differences in genotypes sequenced for each genome B97-61/B2 for

CRIOLLO and Matina1-6 for MATINA. Another reason may be the telomeric and centromeric regions of the chromosomes, which are highly variable and therefore differences between genotypes, such as those sequenced for assembling the genomes, are to be expected.

Once the genomic coordinates of each marker were obtained, they were grouped according to their respective QTLs to obtain the flanking coordinates of the QTLs in the two genomes (Table 6), the size of the regions ranged from 198,254 bp to 4882174 bp in the CRIOLLO genome (GC) and from 201,785 bp to 7479140 bp for the MATINA genome (GM), which may be due to the fact that these regions were obtained based on linkage maps in cM and not on physical maps, adding to that the fact that the cM from each QTL in downstream and upstream direction was very variable in each map and QTL (Max 10 cM downstream and upstream for each QTL), and not all QTLs were reported with two flanking markers but only one, in these cases an average distance of 1.5 MM of base pairs was assumed from the reported marker towards the QTL direction or when the only marker was reported at 0 cM from the QTL , distances of 750,000 bp were assumed in each direction.

The total sum of the 20 genomic regions added 31,812,061 bp for the GC and 30,682,127 bp for the GM, assuming that the differences were due to the different genotypes sequenced as commented previously.

4.5 Functional annotation

A mixed genomic annotation methodology was used, including data of whole protein sequence alignment against the NCBI nr database and an approach based on the extraction of functional domains and compared with several data banks. Once the genomic regions were delimited, and the genomic files corresponding to the genomic regions were downloaded, the protein sequences were retrieved and transformed into FASTA files. A total of 5714 predicted proteins were recovered for all regions in the GC and a total of 3800 predicted protein sequences for all the regions under study in the GM.

Table 6. Number of candidate genes per region and type of domains present. (CN) Proteins with Coiled-coils (*CC*) and nucleotide binding sites (NBS) domains, (CNL) Proteins with CC, NBS and Leucine-rich repeat (LRR) domains, (MLO) Proteins with Mlo-like resistance proteins, (N) Proteins with just NBS domain, (NL) proteins with NBS and LRR domains, (RLK) Proteins with Receptor-like Kinase (RLK), (RLKGNK2) Proteins with RLK and Ginkbilobin2 (GNK2) domains, (RLP) Proteins with Receptor-like without the Kinase domain, (RPW8NL) Proteins with Resistance to Powdery Mildew 8 (*RPW8*) NBS and LRR domains, (T) proteins with Toll/interleukin-1 receptor (TIR) domain, (UNKNOWN) Proteins with Leucine-rich repeat (LRR) domains that do not fit any other class (C) CRIOLLO, (M) MATINA.

QTL	LG	G	Physical potition	CN	CNL	MLO	N	NL	RLK	RLKGNK2	RLP	RPW8NL	T	UNKNOWN	Total
AKAfolICS100CHR10	**10**	**C**	2551520:4052040	-	-	2	-	-	-	-	-	-	-	2	**4**
		M	16571900:18072415	-	-	-	-	-	-	-	-	-	-	1	**1**
AKAFfolICS100CHR1	**1**	**C**	30694434:33468108	-	-	-	-	-	-	-	1	-	-	-	**1**
		M	28502617:31502617	-	-	-	-	-	-	-	-	-	-	5	**5**
AKAfolICS100CHR3	**3**	**C**	35179055:36293379	-	2	-	2	4	-	-	1	-	-	4	**13**
		M	29572563:30703073	1	3	-	2	1	-	-	1	-	-	3	**11**
AKAprrHCHR1	**1**	**C**	871438:2417006	-	3	-	-	-	-	-	1	3	-	3	**10**
		M	867922:2421710	-	2	-	-	-	-	-	1	2	-	3	**8**
AKAprrHCHR6	**6**	**C**	25659781:26148133	-	-	-	-	-	-	-	-	-	-	-	**0**
		M	16380796:16859704	-	-	-	-	-	-	-	-	-	-	-	**0**
AKAprrHCHR8	**8**	**C**	1022182:2251917	-	-	-	-	-	-	-	4	-	-	1	**5**
		M	7171923:8400711	-	-	-	-	-	-	-	3	-	-	1	**4**
AKAprrICS100CHR4	**4**	**C**	30388244:30586498	-	-	-	-	-	-	-	-	-	-	-	**0**
		M	25770441:25972979	-	-	-	-	-	-	-	-	-	-	-	**0**
AKAprrICS100CHR61	**6**	**C**	20000000*:21474437	-	2	-	-	2	-	-	7	-	-	1	**12**
		M	7295362*:8795362	-	-	-	-	-	-	-	-	-	-	4	**4**
AKAprrICS100CHR62	**6**	**C**	3494620:5000000*	-	-	-	-	-	1	-	2	-	-	2	**5**
		M	14881286:16381286*	-	2	-	-	-	-	-	2	-	-	-	**4**
AKAprrICS95CHR2	**2**	**C**	703843:2174093	-	8	-	-	11	-	-	2	-	-	5	**26**
		M	731519:1316407	-	3	-	-	4	-	-	-	-	-	3	**10**
AKAprrICS95CHR4	**4**	**C**	19242045:19491718	-	-	-	-	-	-	-	-	-	-	-	**0**
		M	11207385:11409170	-	-	-	-	-	-	-	-	-	-	-	**0**
BARq1BPPcCHR1	**1**	**C**	30694434:35306336	-	-	-	-	-	1	-	1	-	-	-	**2**
		M	39928772:41428772	-	4	-	-	-	-	-	2	-	-	3	**9**
BARq1BPPctCHR6	**6**	**C**	219626:689235	1	1	-	-	-	-	-	2	-	-	1	**5**
		M	6827201:7295820	-	2	-	-	-	-	-	2	-	-	1	**5**
BARq1BPPpCHR6	**6**	**C**	23520483:25804709	-	-	-	-	-	16	21	2	-	1	11	**51**
		M	16717720:19056211	-	-	-	-	-	6	15	2	-	1	9	**33**
BARq2BPPcCHR2	**2**	**C**	8244114:9182779	-	-	-	-	-	-	-	-	-	-	4	**4**
		M	7336749:8276242	-	-	-	-	-	-	1	-	-	-	4	**5**
BARq3BPPcCHR3	**3**	**C**	35366498:36293379	-	2	-	2	4	-	-	1	-	-	4	**13**
		M	29572563:30545307	1	3	-	2	1	-	-	1	-	-	3	**11**
BARq4BPPcCHR4	**4**	**C**	27592136:28579412	-	-	-	-	-	1	-	1	-	-	3	**5**
		M	21130572:22096856	-	-	-	-	-	1	-	1	-	-	3	**5**
BROphy1CHR4	**4**	**C**	1:1000000*	-	-	-	-	-	-	2	1	-	-	-	**3**
		M	1:1000000*	-	-	-	-	-	-	1	1	-	-	-	**2**
BROphy2CHR8	**8**	**C**	2391280:4552442	-	4	3	-	-	-	-	3	-	-	2	**12**
		M	4898997:7034634	-	3	2	-	-	-	-	3	-	-	2	**10**
BROphy3CHR10	**10**	**C**	16000000*:20882174	1	33	-	1	19	-	1	4	-	-	20	**79**
		M	2511121*:9990261	-	9	-	-	5	-	1	4	-	-	24	**43**
Total		**C**		**2**	**55**	**5**	**5**	**40**	**19**	**24**	**33**	**3**	**1**	**63**	**250**
		M		**2**	**31**	**2**	**4**	**11**	**7**	**18**	**23**	**2**	**1**	**69**	**170**

The entire sequences and domains contained in the sequences were aligned for each genomic region in each of the genomes, the XML files were mapped and annotated in BLAST2GO®. A total of 4,163 (Figure 7) sequences for the GC and 2,849 (Figure 8) were annotated for the GM. The number of sequences scored lower than the number of sequences retrieved from the genomes can be due to cause such as sequencing errors generating proteins without identifiable orthologs, gaps in the genome that can generate incomplete sequences, among other reasons, which generates sequences of no hits against the databases.

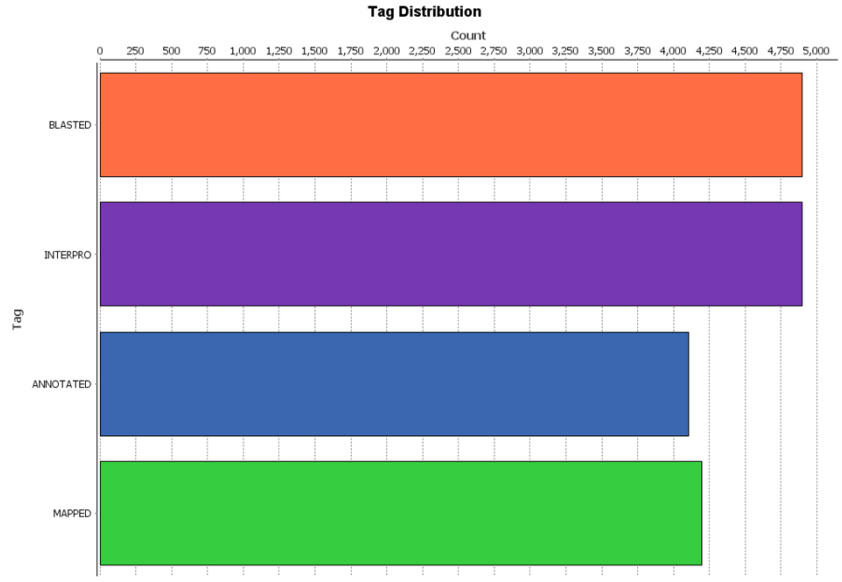

Figure 7. Summary annotation GO Gene Ontology, Genome CRIOLLO

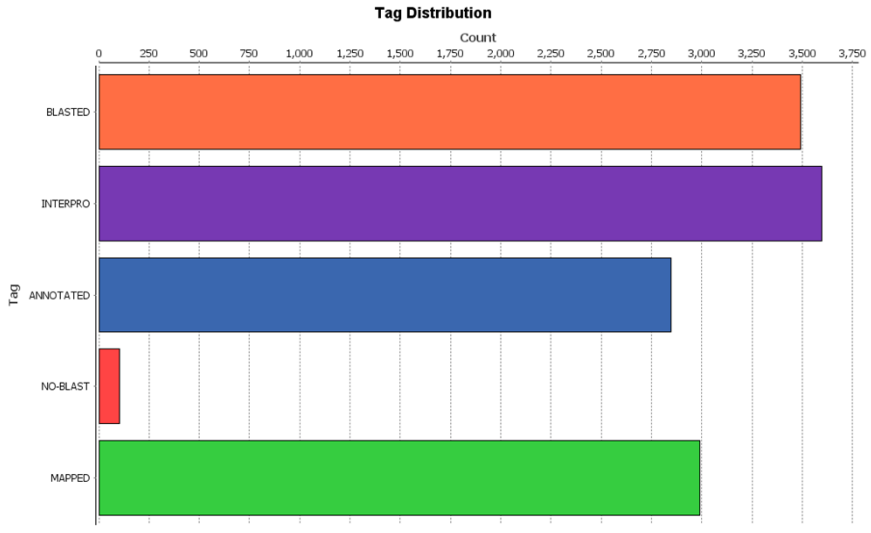

Figure 8. Summary annotation GO Gene Ontology, Genome MATINA

In general, results of the annotation in terms of GeneOntology (GO) for the GC (Figure 9) showed a differentiated grouping of genes for metabolic processes and cellular processes at the level of biological processes, for the level of molecular function, the greatest number of genes were grouped in the categories of catalytic activity and binding. At cellular component level most of the genes were grouped in four locations: membrane, membrane part, cell and cell part, these groupings occurred also for the GM (Figure 10). These results were expected, since these groups contain structural genes for the normal functioning of the cell. The present study focused on the genes grouped at the level of biological processes in the category response to stimulus, in which are normally grouped the defense and combat genes of pathogens in the cell, for this category were scored 353 genes for the GC and 258 for the GM.

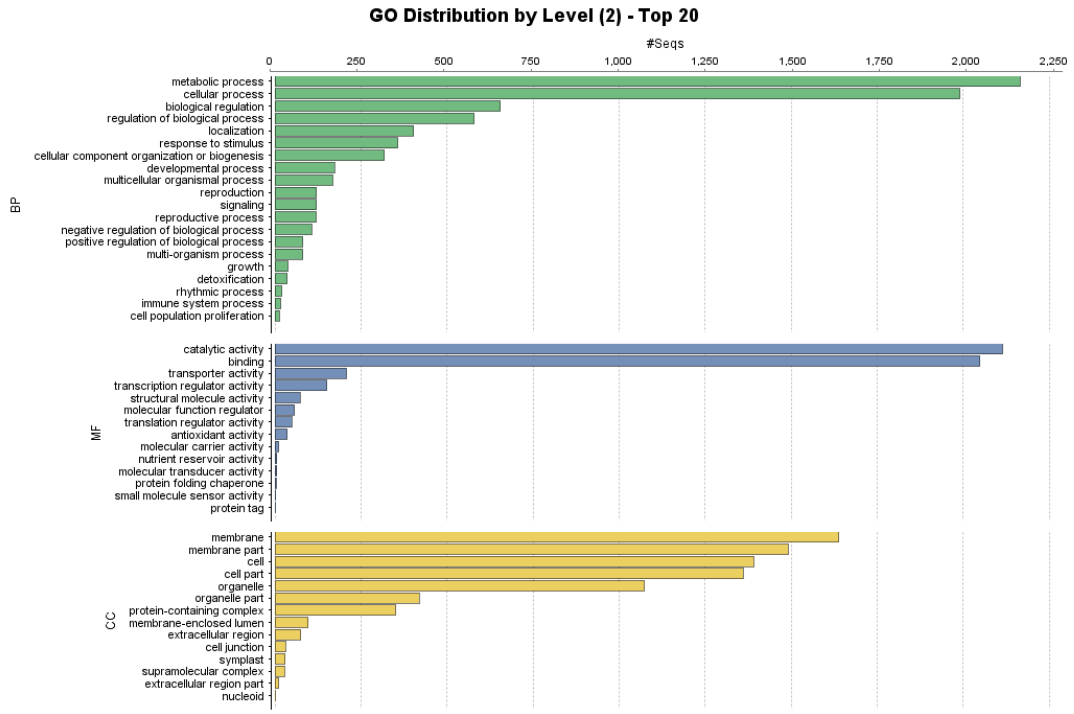

Figure 9. Distribution of proteins in GO levels,Genome CRIOLLO

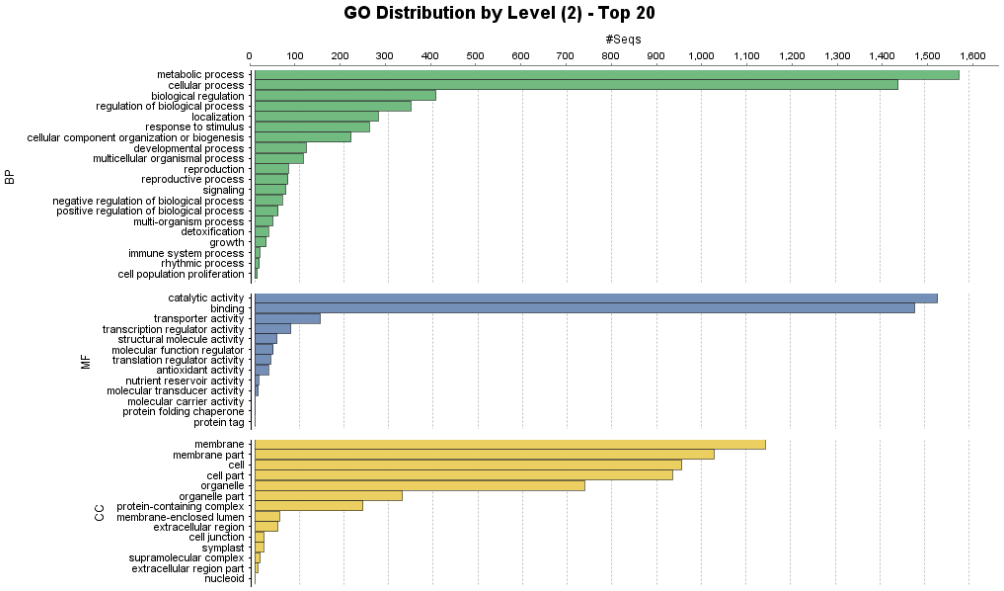

Figure 10. Distribution of proteins in GO levels,Genome MATINA

4.6 Selection of candidate genes

Once the protein sequences of the different QTLs and genomes were annotated, the results were classified into 12 gene classes associated with resistance of pathogens in plants, genes were identified only for 10 of these 12 classes (Table 6). A total of 250 candidate genes for the GC and 170 for the GM (Appendix 1), accumulating the majority in 5 classes RLK, NL, RLKGNK2, RLP and CNL being the last two those that more hits accumulated in the two genomes.

The distribution of the candidate genes in the different genomic regions (Table 6) showed that the three regions richest in candidate genes were Phyto3 (LG10) reported by Brown *et al.* (2007), q1-BPPp (LG6) reported by Barreto *et al.* (2018) and PRR / 95 (LG2) reported by Akasa *et al.* (2016) (Table 6). These results may be related to the total size of the regions, the wider the probability of finding genes and vice versa, so regions with fewer hits should not be ignored.

Similar results were obtained for some genic classes for the two reference genomes, such as the RLP class, which was also the most represented class, containing hit in 15 of the 20 regions for the GC and 12 regions of the 20 for the GM. However, not all the regions showed such a marked similarity, as for example the CN gene class, in which none of the few hits was located in the same genomic regions of the genomes. These results are probably associated to the different genotypes used

for sequencing and the protein prediction algorithms used, *ab initio* for GM and similarity for GC.

When reviewing the most frequent gene classes, it was observed a pattern in the hits of the CNL gene classes, of the 55 genes identified in the GC, 28 genes were annotated with disease resistance RPP13-Like, also for the GM, of the 31 genes of the Class 14 genes were annotated with the same hit, mainly concentrated in the LG10, in the Phyto3 region reported by Brown *et al.* (2007). RPP13 is a locus in *Arabidopsis* responsible for the resistance to *Peronospora parasitica* (Bittner *et al.,* 1999; Bittner *et al.*, 2000) organism that taxonomically belongs to the monophyletic group of the Oomycetes, as *Phytophthora palmivora*, inferring a high probability that this group of genes are involved in the fight against the pathogen in cocoa.

When joining the information of the loci identified in this study by mapping by association with the *in-silico* information (Table 7) it can be observed that almost all the regions (except q2-BPPc LG2) have candidate genes for the CG, being the richest groups PRR / 95 (LG2) in the GC reported by Akasa *et al.* (2016) and Phyto2 (LG8) in the GC reported by Brown *et al.* (2007). It is important to note that since the resistance to pathogens is a quantitative character, the fact that not all the markers were positive in the analysis of association does not mean that the genes in those regions are not involved in the fight against the pathogen. The no hit can be due to diverse causes like the density of the mapping.

Table 7. Candidate genes at loci associated with resistance BP.

		CRIOLLO						MATINA							
	LG	RLP	CNL	MLO	NL	N	UN	RLP	CNL	MLO	NL	RLKGNK2	CN	N	UN
AKAprrHCHR8	8	4	-	-	-	-	1	3	-	-	-	-	-	-	1
BROphy2CHR8	8	3	4	3	-	-	2	3	3	2	-	-	-	-	2
AKAprrICS95CHR2	2	2	8	-	11	-	5	-	3	-	4	-	-	-	3
BARq2BPPcCHR2	2	-	-	-	-	-	4	-	-	-	-	1	-	-	4
AKAfolICS100CHR3	3	1	2	-	4	2	4	1	3	-	1	-	1	2	3
BARq3BPPcCHR3	3	1	2	-	4	2	4	1	3	-	1	-	1	2	3

An important region is q1-BPPp (LG6) reported by Barreto *et al.* (2018) as directly associated with resistance to *Phytophthora palmivora* and was the second region richest in candidate genes for the two genomes, accumulating the hits in two

main gene classes RLK and RLKGNK2, these class includes RLK (Receptor like Kinase) domains known by Detect specific pathogenic peptides that signal to Pelle-family kinases (Die´vart & Clark, 2004) and play central roles in signaling during pathogen recognition for the subsequent activation of defense mechanisms and developmental control (Afzal *et al*. 2008), and additionally the RLKGNK2 class contains a GnK2 domain (Ginkbilobin-2), Gnk2 is a protein secreted by *Ginkgo biloba* seeds that exhibits an antifungal activity (Sawano *et al.*, 2007; Wang & Ng, 2000). Gnk2 has a plant-specific cysteine-rich motif DUF26 (domain of unknown function 26, also known as stress-antifungal domain: PF01657) which belongs to cysteine-rich receptor-like kinases (CRKs) (Miyakawa *et al.*, 2014) not showing any similarity with other known antimicrobial proteins (Sawano *et al.*, 2007; Miyakawa *et al*., 2014). It was recently shown that Gnk2 can also activate actin-dependent cell death (Gao *et al*., 2016). Therefore, *Cast_Gnk2*-like may prevent pathogen growth either by its chemical properties or by inducing HR-related cell death. The hits reported in this genetic class (RLKGNK2), which was the third with the highest number of hits, represents a high potential as candidate genes for *Phytophthora palmivora* resistance in cocoa. In a recent study Santos *et al.* (2017), the authors reported a high differentiation in the expression of *cas_GnK2* in Castanea in trials with inoculation with *Phytophthora cinnamomi*, reporting the high potential that could have the isolation and purification of *Cast_Gnk2*-like protein to the development of an antimicrobial phytopharmaceutical against *P. cinnamomi*.

5 CONCLUSIONS

Local Ancient Varieties of Bahia have a differentiated alelic profile, low heterozygosity rate and medium to low allelic diversity in SSR molecular markers associated with phenotypic variations in *Theobroma cacao* L. / *Phytophthora palmivora* Butler (Butler) interaction.

The Ancient Local varieties of Bahia, Comum and Pará, have a very overlapping population structure, with slightly differentiable alelic profiles between the two varieties and, Maranhão, forms a differentiated group compared to Comum and Pará, forming a varietal genetic group

QTL genomic regions associated with resistance in the *Theobroma cacao* L. / *Phytophthora palmivora* Butler (Butler) interaction are regions rich in genes associated with resistance to pathogens in plants.

Genes with RLK and RLK + GNK2 domains, grouped in Chromosomes 4 and 6 of *Theobroma cacao* L., have a great potential to be partially responsible for the recognition and response in the interaction with the pathogen *Phytophthora palmivora* Butler (Butler).

Genes with CNL domains grouped in chromosomes 1, 3, 6, 8 and 10 of *Theobroma cacao* L., have a great potential to be partially responsible for the recognition and response in the interaction with the pathogen *Phytophthora palmivora* Butler (Butler).

6. REFERÊNCIAS

ADEBOLA, M. O. &AMADI, J. E.**Screening three Aspergillus species for antagonistic activities against the cocoa black pod organism (*Phytophthora palmivora*)**. Agriculture and Biology Journal of North America. ISSN: 2151-7525. 2010.

ADL, S.M.; SIMPSON, A.G.B.; FARMER, M.A. *et al.* **The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists.** J. Eukaryot. Microbiol. 52(5), 399–451, 2005.

AFZAL, A.J.; WOOD, A.J.; LIGHTFOOT, D.A. **Plant receptor-like serine threonine kinases: roles in signaling and plant defense**. Mol Plant Microbe Interact 21:507–517. 2008.

AIKPOKPODION, P.O.; MOTAMAYOR, J.C.; ADETIMIRIN, V.O.; ADU-AMPOMAH, Y.; INGELBRECHT, I.; ESKES, A.B. *et al.* **Genetic diversity assessment of sub-samples of cocoa, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker**. Tree Genet and Genomes. 5: 699–711. doi: 10.1007/ s11295-009-0221-1.2009.

AKAZA, M.J.; KOUASSI, A.B.; AKAFFOU, D.S. *et al.***Mapping QTLs for Black pod (*Phytophthora palmivora*)resistance in three hybrid progenies of cocoa (*Theobroma cocoa L.*) using SSR markers.** International Journal of Scientific and Research Publications 6(1), 298-311, 2016.

AKIVA, E.; BROWN, S.; ALMONACID, D.E. *et al.* **SFLD The Structure-Function Linkage Database.** Nucleic Acids Res. Jan;42 (Database issue): 521-30. doi: 10.1093/nar/gkt1130. 2014

AKROFI, A. Y.; APPIAH, A. A.; OPOKU, I. Y. **Management of Phytophthora pod rot disease on cocoa farms in Ghana.** Crop Protection 22: 469-477, 2003.

ALEXOPOULOS, C.J.; MIMS, C.W.; BLACKWELL, M. **Introductory mycology**. New York, John Wiley e Sons, Inc. 869. 1996.

ALTSCHUL, S.F.; Gish, W.; Miller, W.; Myers, E.W.;Lipman, D.J.**Basic local alignment search tool**. Journal of molecular biology. 215:403-410. 1990.

ANDERSON, M.J.; TER BRAAK, C.T.F. **Permutation tests for multifactorial analysis of variance.** J. Stat. Comput. Simul. 73, 85–113. 2003.

APG IV. **An update of Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APGIV**. Bot. J. Linn. Socc. 181(1): 1-20, 2016.

ARAÚJO, M.; ALGER, K.; ROCHA, R.; MESQUITA, C.A.B. **A Mata Atlântica do sul da Bahia: situação atual, ações e perspectivas.** Costa JPO, editor. Reserva da Biosfera da Mata Atlântica—MAB—UNESCO; 1998.

ARGOUT, X.; SALSE, J.; AURY, J.M. *et al.***The genome of *Theobroma cacao*.** Nature Genetics. 43:101-108. 2011

ARGOUT, X.; MARTIN, G.; DROC, G. *et al.***The cocoa Criollo genome v2.0: An improved version of the genome for genetic and functional genomic studies**. BMC Genomics. 18. 10.1186/s12864-017-4120-9. 2017.

ASHBURNER *et al*. **Gene ontology: tool for the unification of biology.** Nat Genet. 25(1):25-9. 2000.

ATTWOOD,T.K. *et al.* **The PRINTS database: a fine-grained protein sequence annotation and analysis resource–its status in 2012**. Database, 2012, bas019. 2012.

BAHIA, R. C.; VILDOSO, C. I. A.; LUZ, E. D. M. N.; LOPES, U. V.; MACHADO, R. C. R.; CORRÊA, R. X. **Resistance to black pod disease in a segregating cocoa tree population.** Tropical plant pathology, v.40, p. 13-18, 2015.

BALDAUFL, S.; ROGER, A. J.; WENK-SIEFERT, I.; DOOLITTLE, W. F. **A Kingdom-Level Phylogeny of Eukaryotes Based on Combined Protein Data.** Science 290, 972-977, 2000.

BARRETO, M.A.; SANTOS, J.C.S.; CORRÊA, R.X.; LUZ, E.D.M.N.; MARELLI, J.; SOUZA, A.P**. Detection of genetic resistance to cocoa black pod disease caused by three Phytophthora species**. Euphytica 206:677–687. 2015.

BARRETO, M.A.; ROSA, J.; HOLANDA, I. *et al.* **QTL mapping and identification of corresponding genomic regions for black pod disease resistance to three *Phytophthora* species in *Theobroma cacao* L.** Euphytica. 214:188. 2018.

BARTLEY, B.G.D. **The genetic diversity of cocoa and its utilization**. Wallingford: CABI Publishing; 2005.

BENNETT, A.B. **Out of the Amazon: *Theobroma cacao* enters the genomic era**. Trends Plant Sci. 8:561–563. 2003

BIRCH, P. R.J.; BOEVINK, P.C.; GILROY, *et al.* **Oomycete RXLR effectors: delivery, functional redundancy and durable disease resistance**. Current Opinion in Plant Biology. 11: 373–379. 2008.

BITTNER-EDDY, P.; CAN, C.; GUNN, N. PINEL, M. *et al.* **Genetic and Physical Mapping of the RPP13 Locus, in Arabidopsis, Responsible for Specific Recognition of Several *Peronospora parasitica* (Downy Mildew) Isolates.** Molecular Plant-Microbe Interactions. 12:9. 792-802. 1999.

BITTNER-EDDY, P.; CRUTE, I.; HOLUB, E.; BEYNON, J. **RPP13 *is a simple locus in* Arabidopsis thaliana *for alleles that specify downy mildew resistance to different avirulence determinants in Peronospora parasitica*.** The plant journal The Plant Journal. 21 (2), 177±188. doi.org/10.1046/j.1365-313x.2000.00664.x. 2000.

BOWERS, J. H.; BAILEY, B. A.; HEBBAR, P. K.; SANOGO, S.; LUMSDEN,R. D. **The impact of Plant Diseases on World Chocolate Production**. PlantHealth Progress. 2001.

BROWN, J.S.; PHILLIPS-MORA, W.; POWER, E.J. *et al.* **Mapping QTLs for resistance to frosty pod and black pod diseases and horticultural traits in *Theobroma cacao*.**CropScience47, 1851–8, 2007.

BRU,C. *et al.* **The ProDom database of protein domain families: more emphasis on 3D**. Nucleic Acids Res. 33, 212–215. 2005.

CHAN, A.P.;CRABTREE, J.;ZHAO, Q. *et al.***Draft genome sequence of the oilseed species *Ricinus communis*.**Disponível em: <http://castorbean.jcvi.org/downloads.php>.Acesso em: 11 setembro 2018.

CHANG, Y.H.; YAN, H.Z.; LIOU, R.F. **A novel elicitor protein from*Phytophthoraparasitica*induces plant basal immunity and systemic acquiredresistance.**Molecular Plant Pathology. 16:123 –136. 2015.

CHAPARRO-GARCIA, A.; WILKINSON, R.C.; GIMENEZ-IBANEZ, S *et al.***The receptor like-kinase SERK3/BAK1is requiredfor basalr esistance against thelate blight pathogen *Phytophtho rainfestans*in *Nicotiana benthamiana*.**PLoS One. 6:1. 608. 2011.

CHEONG, J.J.; BIRBERG, W.; FUGEDI, P. *et al.* **Structure-activity relationships of oligobetaglucoside elicitors of phytoalexin accumulation in soybean**. The PlantCell. 3:127–136. 1991.

CLAIR, D.A. **Quantitative disease resistance and quantitative resistance loci in breeding**. Annual Review of Phytopathology48, 247–68. 2010.

CLEMENT, D.; RISTERUCCI, A.M.; MOTAMAYOR, J.C.; N'GORAN, J. A.; LANAUD, C. **Mapping QTL for yield components, vigor, and resistance to *Phytophtho rapalmivora*in *Theobroma cacao* L.**Genome46, 204–12, 2003.

COLLARD, B.C.Y.; MACKILL, D.J. **Marker-assisted selection: an approach for precision plant breeding in the twenty-first century**. Phil. Trans. R. Soc. B 363, 557–572, 2008

CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J.M.; TEROL, J.; TALÓN, M.; ROBLES, M. **Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research.** Bioinformatics 21 (18):3674–3676. 2005.

CROUZILLAT, D.; MENARD, B.; MORA, A.; PHILLIPS, W.; PETIARD, V. **Quantitative trait analysis in *Theobroma cacao* using molecular markers.**Euphytica114, 13–23, 2000.

DAKWA, J.T. **A serious outbreak of black pod disease in a marginal area of Ghana.** In Proceedings of the Xth International Cocoa Research Conference, Santo Domingo, República Dominicana 447-452, 1987.

DENECKE, J.; ANIENTO, F.; FRIGERIO, L.*et al.***Secretory pathway research: the more experimental systems the better.**The Plant Cell. 24:1316–1326. 2012.

DENOUX, C.; GALLETTI, R.; MARMMARELLA, N. *et al.***Activation of defense response pathways by OGs and Flg22 elicitors in *Arabidopsis* seedlings**. Molecular Plant.1: 423–445. 2008.

DIE´VART, A.; CLARK, S.E. **LRR-containing receptors regulating plant development and defense**. Development 131(2):251–61. 2004.

DAVIE, J.H. **Chromosome studies in the malvaceae and certain related families**. II. Genetica 17:487–498. 1935.

DODDS, P.N.; RATHJEN, J.P. **Plant immunity: towards an integrated view of plant-pathogen interactions.** Nature Review Genetics. 11: 539–548. 2010.

DOYLE, J.J.; DOYLE, J.L. **Isolation of plant DNA from fresh tissue**. Focus, Rochester 12 (1), 13-15, 1990.

DU, J.; VERZAUX, E.; CHAPARRO-GARCIA, A. *et al.* **Elicitin recognition confers enhanced resistance to *Phytophtho rainfestans* in potato.** Nature Plants.1:15034. 2015.

ÉRSEK, T.; RIBEIRO, O. **Mini review article: an annotated list of new *Phytophthora* species described post**. Acta Phytopathologica et Entomologica Hungarica. 251-266. 1996.

EVANS, H.C. **Witches' broom disease: a case study**. Cocoa Growers Bulletin. 32: 5-19, 1981.

FAGOAGA, C.; ISMAEL, R.; CONEJERO, V. *et al.* **Increased tolerance of *Phytophthora citrophthora* in transgenic orange plants constitutively expressing a tomato pathogenesis related protein PR-5**. Molecular Breeding. 7 (2): 175–185. 2001.

FAVRE, H.A. & POWELL, W. H. **Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Name 2013.** Cambridge, UK: The Royal Society of Chemistry, ISBN 978-0-85404-182-4. 2013.

FAWKE, S.; DOUMANE, M.; SCHORNACK, S. **Oomycete interactions with plants: infection strategies and resistance principles.** Microbiology and Molecular Biology Reviews. 79 (3): 263-280. 2015.

FAOSTAT **Food and Agriculture Organization of the United Nations**. Database. Rome, Italy: FAO. Disponível em: <http://faostat3.fao.org/home/E>

FISTER, A. S.; O'NEIL, S. T.; SHI, Z.*et al.* **Two *Theobroma cacao* genotypes with contrasting pathogen tolerance show aberrant transcriptional and ROS responses after salicylic acid treatment. Journal of Experimental Botany**.doi:10.1093/jxb/erv334. 2015.

FLAMENT, M.H.; KEBE, I.; CLEMENT, D. *et al.* **Genetic mapping of resistance factors to *Phytophthora palmivora*in cocoa.**Genome44, 79–85, 2001.

FLOR, H.H. **Inheritance of pathogenicity in *Melampsoralini*.**Phytopathology.32: 653-669. 1942.

FLORES, L. & KROHN, A. **Scoring Microsatellite Loci.** Article *in* Methods in molecular biology (Clifton, N.J.) . DOI: 10.1007/978-1-62703-389-3_21. 2013

FLOUDAS, C. A.; FUNG, H. K.; MCALLISTER, S. R.; MÖNNIGMANN, M.; RAJGARIA, R. **Advances in protein structure prediction and de novo protein design: A review.** *Chemical Engineering Science*, *61*(3), 966-988. 2006.

FOISSAC, S. *et al.* **Genome annotation in plants and fungi: EuGène as a model platform**. Curr. Bioinform. 3, 87–97. 2008.

FOUET, O.; ALLEGRE, M.; ARGOUT, X. *et al.* **Structural characterization and mapping of functional EST-SSR markers in *Theobroma cacao*.** Tree Genetics & Genomes. 7:799–817. DOI 10.1007/s11295-011-0375-5. 2011.

GADJI, A. A. G.; YAPO, O. B.; ABO, K.*et al.* **In vitro assessment of biopesticide bacillus thuringiensis var. kurstakihd- 1 effectiveness on phytophthora palmivora, agent of cocoa black pod rot in côted'ivoire**. European Scientific Journal, 11 (21).ISSN 1857-7431. 2015.

GAO, N.; WADHWANI, P.; MÜHLHÄUSER, P.; LIU, Q.; RIEMANN, M.; ULRICH, A. S. *et al.* **An antifungal protein from *Ginkgo biloba* binds actin and can trigger cell death.** Protoplasma 253, 1159–1174. doi: 10.1007/s00709-015-0876-4. 2016.

GAULIN, E.; DRAMÉ, N.; LAFITTE, C. *et al.***Cellulose binding domains of a *Phytophthora* cell wall protein are novel pathogen-associated molecularpatterns.**The Plant Cell. 18:1766 –1777. 2006.

GAULIN, E.; BOTTIN, A.; DUMAS, B. **Sterol biosynthesis in oomycete pathogens.** Plant Signaling & Behavior. 5:258 –260. 2010.

GREEN AMERICA. **Big Chocolate" brings in $83 billion per year. Where does all this money end up?**. Green America. 2014. Disponívelem: <http://blog.greenamerica.org/2014 /10/15/big-chocolate-brings-in-83-million-per-year-where-does-all-this-money-end-up/>. Acessoem: 13 setembro 2018.

GUEST, D. **Black pod: diverse pathogens with a global impact on cocoa yield.** Phytopathology 97,1650–3, 2007.

HAAS, B.J. **Analysis of alternative splicing in plants with bioinformatics tools.**Curr Top MicrobiolImmunol. 326:17-37. 2008.

HAFT,D.H. *et al.* **TIGRFAMs and Genome Properties in 2013**. Nucleic Acids Res., 40, 387–395. 2012.

HALDAR, K.; KAMOUN, S.; HILLER, N.L.; BHATTACHARJE, S.; VAN OOIJ, C. **Common infection strategies of pathogenic eukaryotes.** Review. NATURE Reviews/Microbiology 4, 922-31, 2006.

HARDHAM, A.R. **Cell biology of plant-oomycete interactions**. Cell Microbiology. 9. 31–39. 2007.

HOGENHOUT, S.A.; HOORN, R.A.L.V.; TERAUCHI, R.; KAMOUN, S.**Emerging concepts in Effector Biology of Plant-Associated Organisms**. Molecular Plant-Microbe Interactions. 22(2). 115-122. 2009.

HOWDEN, A. J.; HUITEMA, E. **Effector-triggered post-translational modifications and their role in suppression of plant immunity**. Frontiers in Plant Science. 3. 160. 2012.

IBGE Instituto brasileiro de geografia e estadística. **Sistema IBGE de Recuperação Automática - SIDRA, Brasil**. Disponível em:<https://sidra.ibge.gov.br/tabela/1618#notas-tabela>. 2018.

ICCO. **The International Cocoa Organization. Annual report 2014/2015**. London United Kingdon, 2017.

JAFFE, D.B.; GNERRE, S.; MAUCELI, E.*et al.* **Whole-genome sequence assembly for mammalian genomes: Arachne 2**. Genome Research 13:91-96. 2003.

JAILLON, O.; AURY, J.; NOEL, B. *et al.***The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**.Disponível em:<ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Vvinifera/>.Acesso em: 13 outubros 2018.

JONES, J.D.; DANGL, J.L. **The plant immune system**. Nature. 444: 323– 329.2006.

JONES, P.; BINNS, D.; CHANG, H.Y.; *et al.* **InterProScan 5: genome-scale protein function classification**. Bioinformatics. 30:9.1236–1240. doi:10.1093 2014.

KAKU, H. *et al.* **Plant cells recognize chitin fragments for defense signaling through a plasma membrane receptor**. Proceedings of the National Academy of science of the USA. 103:11086 –11091. 2006.

KAMOUN, S.; VAN WEST, P.; DE JONG, A.J.*et al.* **A gene encoding a protein elicitor of *Phytophtho rainfestans*is down-regulated during infection of potato.** Molecular Plant Microbe Interact10:13–20. 1997.

KAMOUN, S.; VAN WEST, P.; VLEEHOUWERS, V.G.; DE GROOT, K.E.;GOVERS, F. **Resistance of *Nicotiana benthamiana*to *Phytophthora infestans*is mediated by the recognition of the elicitor protein INF1**. The Plant Cell10:1413–1426. 1998.

KAMOUN, S. **A catalogue of the effector secretome of plant pathogenic oomycetes.**Annual Review of Phytopathology. 44: 41–60. 2006.

KENT, W.J. **BLAT--the BLAST-like alignment tool.**Genome Research12, 656-664 2002.

KLARZYNSKI, O.; PLESSE, B.; JOUBERT, J.M.*et al.* **Linear beta-1,3 glucans are elicitors of defense responses in tobacco.** Plant Physiology. 124:1027–1038. 2000.

KOVER, P.X.; CAICEDO, A.L. **The genetic architecture of disease resistance in plants and the maintenance of recombination by parasites.** Molecular Ecology10, 1–16. 2001.

LAMOUR, K.; WIN, J.; KAMOUN, S. **Oomycete genomics: new insights and future directions.** FEMS MicrobiolLett 274, 1–8, 2007.

LANAUD, C.; SOUNIGO, O.; AMEFIA, Y.K.; PAULIN, D.; LACHENAU, P.; CLÉMENT, D. **Nouvelles données sur le fonctionnement du système d'incompatibilité du cocoayer et ses conséquences pour la élection**. Café, Cocoa, Thé. 31: 267–277. 1987.

LANAUD, C.; RISTERUCCI, A. M.; PIERETTI, I. *et al.* **Isolation and characterization of microsatellites in *Theobroma cacao* L.** Molecular Ecology. **8**, 2141–2152. 1999.

LANAUD, C.; RISTERUCCI, A.M.; PIERETTI, I.; N'GORAN, J.A.; FARGEAS, D. **Characterisation and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao* L.).**MolecularBreeding13, 211–27, 2004.

LANAUD, C.; FOUET, O.; CLEMENT, D. *et al.* **A meta–QTL analysis of disease resistance traits of *Theobroma cacao* L.**MolecularBreeding24, 361–74, 2009.

LARROQUE, M.; BARRIOT, R.; BOTTIN, A. *et al.* **The unique architecture and function of cellulose interacting proteins in oomycetes revealed by genomic and structuralanalyses.**BMC Genomics13:605. 2012.

LEITE, P.B.; MACIEL, L.F.; OPRETZKA, L.C.F.; SOARES, S.E.; BISPO, E.S. **Phenolic compounds, methylxanthines and antioxidant activity in cocoa mass and chocolates produced from "witch broom disease" resistant and non resistant cocoa cultivars.** Ciência e agrotecnologia. 37: 244–250. Available: doi: 10.1590/S1413-70542013000300007. 2013.

LEES, J. *et al.* **Gene3D: a domain-based resource for comparative genomics, functional annotation and protein network analysis**. Nucleic Acids Res., 40, D465–D471. 2012.

LETUNIC,I. *et al.* **SMART 7: recent updates to the protein domain annotation resource**. Nucleic Acids Res., 40, 302–305. 2012.

LI, L.; STOECKERT, C.J.; ROOS, D.S. **OrthoMCL: identification of ortholog groupsfor eukaryotic genomes.** Genome Res 13, 2178-2189. 2003.

LI, R.; LI, Y.; KRISTIANSEN, K.; WANG, J. SOAP: **short oligonucleotide alignment program**. Bioinformatics (Oxford, England)24, 713-714.2008.

LU, L.; LU, H.; SKOLNICK, J. **MULTIPROSPECTOR: an algorithm for the prediction of protein-protein interactions by multimeric threading.** Proteins: Structure, Function, and Bioinformatics, 49(3), 350-364. 2002.

LUZ, E.D.M.N.; MATSUOKA, K. **Taxionomia e sistemática do gênero*Phytophthora***. Revisão Anual de Patologia de Plantas. 4: 297-328. 1996.

LUZ, E.D.M.N. *et al*. **Cacau (*Theobroma cocoa* L.) controle de doenças.**In: Ribeiro do Vale, F. X. e Zambolim, L. (eds) Controle de doenças deplantas-grandes culturas.Viçosa-UFV. 2: 617–622. 1997.

LUZ, E. D. M. N.; SILVA, S. D. V. M. **Podridão-Parda dos frutos, cancro e outras doenças causadas por *Phytophthora* no cacaueiro.** In: LUZ, E. D. M.N.; SANTOS, A. F.; MATSUOKA, K.; BEZERRA, J. L. (Eds.). Doenças causadas por *Phytophthora* no Brasil. Campinas: Rural, p. 175-265, 2001.

LUZ, E.D.M.N.; SGRILLO, R.B.; SANTOSFILHO, L.P. **Estimativas de danos e perdas causadas por doencas no cacaueiro.** Proceedings of the Proceedings of the WorkShop de Epidemiologia de doenca. Vicosa, 67–79, 2004.

LUZ, E. D. M. N.; PAIM, M. C. A.; SILVA, S. D. V. M.; PIRES, J. L.; SANTOS, L. P.; MONTEIRO, W. R. **Correlations of cocoa clones reaction to diferent species of *Phytophthora***. In: ESKES, A. B.; EFROIN, Y. Global appoaches to cocoa germolasm utilization and conservation. Final report of the CFC/ICCO/IPGRI project on "Cocoa Germplasm Utilization and Conservation: a Global Approach" (1998-2004). CFC, Amsterdam, The Netherlands/ICCO, London, UK/IPGRI, Rome, Italy p.132-135, 2006.

LUZ, E.D.M.N.*et al*. Eds. **Glossário Ilustrado de *Phytophthora*: Técnicas Especiais para o Estudo de Oomicetos**. Itabuna.117. 2008.

MACAGNAN, D. **Isolamento e seleção de bactérias endosporâgenicas e do tipo actinomicetos visando o biocontrole da vassoura-de-bruxa (*Crinipellis perniciosa*) da podridão-parda (*Phytophthora* spp.) do cacaueiro (*Theobroma cocoa* L.) e estudos dos mecanismos de antagonismos ao fungo *Crinipellis perniciosa***. 122p. Tese(Doutorado em Fitopatologia) – Universidade Federal de Viçosa, Viçosa,Minas Gerais. 2005.

MATEOS, F.V.; RICKAUER, M.; ESQUERRÉ-TUGAYÉ, M.T. **Cloning and characterization of a cDNA encoding an elicitor of Phytophthora parasitica var. nicotianae that shows cellulose-binding and lectin-like activities**. Molecular Plant-Microbe Interactions. 10:1045–1053. 1997.

MICHELMORE, R.W. **The impact zone: genomics and breeding for durable disease resistance.** Current Opinion in Plant Biololy6, 397–404. 2003.

MIKES, V.; MILAT, M.L.; PONCHET, M.*et al*. **Elicitins, proteinaceous elicitors of plant defense, are a new class of sterol carrier proteins**. BiochemBiophys Res Commun. 245:133–139.1998.

MIYA, A.; ALBERT, P.; SHINYA, T.*et al*.**CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis.** Proceedings of the National Academy of Science.104:19613–19618. 2007.

MIYAKAWA, T.; HATANO, K.; MIYAUCHI, Y.; SUWA, Y.; SAWANO, Y. & TANOKURA, M. **A secreted protein with plant-specific cysteine-rich motif functions as a mannose-binding lectin that exhibits antifungal activity**. Plant Physiol. 166, 766–778. doi: 10.1104/pp.114.242636. 2014.

MONTEIRO, W.R.; AHNERT, D**. Melhoramento Genético do Cacaueiro.** In: Valle RR, editor. Ciência, Tecnologia e Manejo do Cacaueiro. Itabuna: Vital LTDA; pp. 1–16. 2007.

MOOLEEDHAR, V.; MAHARAJ, W.; O'BRIEN, H. **The collection of Criollo cocoa germplasm in Belize**. Cocoa Grower's Bull. 49, 26-40. 1995.

MOTAMAYOR, J.C.; RISTERUCCI, A.M.; HEATH, M.; LANAUD, C. **Cocoa domestication II: progenitor germplasm of the Trinitario cocoa cultivar**. Heredity 91:322-330. 2003.

MOTAMAYOR, J.C.; LACHENAUD. P.; DA SILVA E MOTA, J.W. *et al*. **Geographic and genetic population differentiation of the Amazonian chocolate tree (*Theobroma cacao* L)**. PLoS ONE 3:e3311. 2008.

MOTAMAYOR, J. C.; MOCKAITIS, K.; SCHMUTZ, J. *et al*. **The genome sequence of the most widely cultivated cocoa type and its use to identify candidate genes regulating pod color**. Genome biology. 14. R53. 10.1186/gb-2013-14-6-r53. 2013.

MOTILAL, L.A. *et al.* **The relic Criollo cocoa in Belize–genetic diversity and relationship with Trinitario and other cocoa clones held in the International Cocoa Genebank, Trinidad.** Plant Genetic Resources8, 106-115. 2010.

MOTILAL, L.A.; ZHANG, D.; MISCHKE, S.; MEINHARDT, L.W.; UMAHARAN, P. **Microsatellite-aided detection of genetic redundancy improves management of the International Cocoa Genebank, Trinidad**. Tree Genet Genomes. 9: 1395–1411. doi: 10.1007/s11295-013-0645-5. 2013;

MPIKA, J.; KEBE, I. B.; DRUZHININA, I. S. *et al.* **Inhibition de *Phytophthora palmivora*, agent de pourriturebrune des cabosses de cocoayer en Côte d'Ivoire, par *Trichoderma* sp.** Sciences & Nature 6 (1) : 49-62. 2009a.

MPIKA, J.; KÉBÉ, I. B.; ISSALI, A. E.*et al.* **Antagonist potential of *Trichoderm aindigenous* isolates for biological control of *Phytophthora palmivora* the causative agent of black pod disease on cocoa (*Theobroma cacao* L.) in Côte d'Ivoire**. African Journal of Biotechnology. 8 (20): 5280-5293. ISSN 1684–5315. 2009b

MUNIBA, F. **ab-initio prediction of protein structure: An introduction.** Bioinformatics Review 4(8). ISSN 2455-6645. 2015.

MURZIN A. G.; BRENNER S. E.; HUBBARD T.; CHOTHIA C. **SCOP: a structural classification of proteins database for the investigation of sequences and structures**. *J. Mol. Biol.* 247, 536-540. 1995.

NARS, A.; LAFITTE, C.; CHABAUD, M. *et al*. ***Aphanomy ceseuteiches* Cell Wall Fractions Containing Novel Glucan-Chitosaccharides Induce Defense Genes and Nuclear Calcium Oscillations in the Plant Host *Medicago truncatula***. PLoS ONE8(9): e75039.2013.

NCBICDD. **The conserved Domain Dtabase V3.16.** Disponível em: <HTTPS://www.ncbi.nlm.nih.gov/cdd/>. 2018.

NCBInrDB. Disponível em: < ftp://ftp.ncbi.nlm.nih.gov/blast/db/>. 2018.

NENNSTIEL, D.; SCHEEL, D.; NÜRNBERGER, T. **Characterization and partial purification of an oligopeptide elicitor receptor from parsley (*Petroselinum crispum*)**. FEBS Lettrs. 431:405–410. 1998.

NIMCHUK, Z.; EULGEM, T.; HOLT, B. III.; DANGL, L. **Recognition and Response in the Plant Immune System.** Annual Review of Genetics. 37: 579-609. 2003.

NÜRNBERGER, T.; NENNSTIEL, D.; JABS, T.*et al*. **High affinity binding of a fungal oligopeptide elicitor to parsleyplasma membranes triggers multiple defense responses.**Cell78:449 – 460.1994.

NYASSÉ, S.; DESPREAU, D.; CILAS, C. **Validity of a leaf inoculation test to assess the resistance to *Phytophthora megakarya* in cocoa (*Theobroma cacao* L.) diallel mating design.** Euphytica**,** v.123, p. 395-399, 2002.

NYASSÉ, S.; CILAS, C.; HERAIL, C.; BLAHA, G. **Leaf inoculation as an early screening test for cocoa (*Theobroma cacao* L.) resistance to *Phytophthora* black pod disease.**Crop Protection14, 657–63, 1995.

OPUKU, I. Y.; AKROFI, A. Y.; HOLDERNES, M.*et al*. **Phosphonic acid: an alternative approach to the control of black pod disease of cocoa caused by *Phytophtho ramegakarya*.** In Proceedings of INCOPED 4th International Seminar on Dealing with Pressing Crop Protection problems, Accra, Ghana, 19 - 21st October 2003, Akrofi, A. Y., Ackonor, J. B. &Ollenu, L. A. A.(Editors), Ghana Cocoa Board, Ghana, p 59-69. 2003.

OBWALD, W.; FLEISCHMANN, F.; RIGLING, D.*et al*. **Strategies of attack and defence in woody plant–*Phytophthora*interactions**. ForestPathologya. 44: 169-190. 2014.

PATERSON, A.H.; BOWERS, J.E.; BRUGGMANN, R.*et al*. **Genome sequence and analysis of the tuber crop potato**.Disponível em:<http://potatogenomics.plantbiology.msu.edu/index.html>. Acesso em: 18 outubros 2018.

PEDRUZZI, L. *et al.* **HAMAP in 2013, new developments in the protein family classification and annotation system**. Nucleic Acids Res., 41, 584–589. 2012.

PIETERSE, C.M.J.; VAN WEES, S.C.M.; VAN PELT, J.A.*et al.* **A Novel Signaling Pathway Controlling Induced Systemic Resistance in *Arabidopsis***.Plant Cell. 10:1571- 1580. 1998.

PIRES, J.L. **Avaliação Quantitativa e molecular de germoplasma para o melhoramento do cacaueiro com ênfase na produtividade, qualidade de frutos e resistência a doenças**. D.Sc. Thesis, University of Viçosa. Disponível em: <http://alexandria.cpd.ufv.br:8000/teses/genetica%20e%20melhoramento/2003/178016f.pdf>. 2003.

PIOVESAN, D.; TABARO, F.; PALADIN, L. *et al*. **MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins**. *Nucleic Acids Research*. 46, 471-476. doi: 10.1093/nar/gkx1071. 2018.

PLOETZ, R. C. **Cocoa diseases: Important threats to chocolate production worldwide.** Phytopathology 97:1634-1639, 2007.

POKOU, N. D.; N'GORAN, J. A. K.; KÉBÉ, I.; ESKES, A.; TAHI, M.; SANGARÉ, A. **Levels of resistance to *Phytophthora* pod rot in cocoa accessions selected on-farm in Côte d'Ivoire.**Crop Prot., 27: 302-309. Doi : 10.1016/j.cropro.2007.07.012, 2008.

PUGH, T.; FOUET, O.; RISTERUCCI, A.M. *et al.* **A new cocoa linkage map based on codominant markers: Development and integration of 201 new microsatellite markers**. Theor Appl Genet.108:1151–1161. DOI 10.1007/s00122-003-1533-4. 2004

PUNTA,M. *et al.* **The Pfam protein families database**. Nucleic Acids Res., 40, 290–301. 2012.

QBCE. **Quarterly Bulletin of Cocoa Statistics. CocoaYear 2013/14.**XL (1), 2014.

RHODES, M. **diversidade genética e variabilidade morfológica de frutos em cultivares locais de cacau resistentes à podridão parda provenientes de Uruçuca, Bahia.** 2016. 98f**.** Dissertação **(**Mestre em Produção Vegetal) - Universidade Estadual de Santa Cruz, Ilhéus, 2016.

ROUX, F.; VOISIN, D.; BADET, T.*et al*. **Resistance to phytopathogens etuttiquanti: placing plant quantitative disease resistance on the map**.Molecular Plant Pathology.15. (5): 427-432. 2014.

RISTERUCCI, A.M.; PAULIN, D.; DUCAMP, M,; N'GORAN, J.A.; LANAUD, C. **Identification of QTLs related to cocoa resistance to three species of *Phytophthora*.**Theoretical and AppliedGenetics108, 168–74, 2003.

SANTOS, A.F; LUZ, E.D.M.N; DOS SANTOS, J.T. ***Phytophtho rabohemeriae*c ausando a gomose da acácia-negra no Brasil.**Fitopatologia Brasileira.29:144. 2004.

SANTOS, E. S. L.; BERNARD, C.; MORI, G. M.; ANHERT, D.; MELLO, D. L. N.; PIRES, J. L.; CORREA, R. X.; SOUZA, A. P. **Genetic structure and molecular diversity of cocoa plants established as local varieties for more than two centuries: the genetics history of cocoa plantations in Bahia, Brazil**. Plos One**,** v.10, n.12, p. 1-18. Dicembro, 2015.

SANTOS, C.; DUARTE, S.; TEDESCO, S.; FEVEREIRO, P.; & COSTA, R. **Expression Profiling of *Castanea* Genes during Resistant and Susceptible Interactions with the Oomycete Pathogen *Phytophthora cinnamomi* Reveal Possible Mechanisms of Immunity**. Frontiers in Plant Science. 8:515. doi: 10.3389/fpls. 2017.

SAWANO, Y.; MIYAKAWA, T.; HIROSHI, Y.; TANOKURA, M.; & HATANO, K. **Purification, characterization, and molecular gene cloning of an antifungal protein from *Ginkgo biloba* seeds**. Biol. Chem. 388, 273–280. doi: 10.1515/BC.2007.030. 2007.

SCHMUTZ, J.; CANNON, S.B.; SCHLUETER, J. *et al.* **Genome sequence of the palaeopolyploid soybean**.Disponível em:<ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Gmax/>. Acesso em: 22 outubros 2018.

SCHULTES, R.E.; CUATRECASAS, J**. Cocoa and its allies. A taxonomic revision of the genus *Theobroma***. Review. Economic Botany Journal19, 416–7, 1964.

SERENO, M.L.; ALBUQUERQUE, P.S.B.; VENCOVSKY, R.A.F.; FIGUEIRA, A. **Genetic diversity and natural population structure of cocoa (*Theobroma cacao* L.) from the Brazilian Amazon evaluated by microsatellite markers.** Conserv Genet. 7: 13–24. doi: 10.1007/s10592-005-7568-0. 2006.

SHULAEV, V.; SARGENT, D.J.; CROWHURST, R.N. *et al.* The genome of woodland strawberry (Fragariavesca). Disponível em:<https://strawberry.plantandfood.co.nz/gbrowse/navbar/strawberry/download.html>. Acesso em: 22 outubros 2018.

SIGRIST,C.J.A. *et al.* **New and continuing developments at PROSITE**. Nucleic Acids Res., 41, 344–347. 2012.

SKINNER, M.E.; UZILOV, A.V.; STEIN, L.D. *et al.* **JBrowse: A next-generation genome browser.** Genome Research. 19:1630–1638. ISSN 1088-9051/09. 2009.

SOBERANIS, W.; RÍOS, R.; ARÉVALO, E.*et al.* **Increased frequency of phytosanitary pod removal in cocoa(*Theobroma cacao*) increases yield economically in eastern Peru.** CropProtection. 18: 677-685. 1999.

SOGIN, M.; SILBERMAN, J.D. **Evolution of the protists and protistan parasites from the perspective of molecular systematic.** International Journal for Parasitology 28, 11-20, 1998.

STAHL, E.A.; BISHOP, J.G. **Plant-pathogen arms races at the molecular level.** Current Opinion in Plant Biology. 3: 299–304. 2000.

STANKE, M.; WAACK, S. **Gene prediction with a hidden Markov model and anew intron submodel.**BioInformatics19:ii215-ii225. 2003.

STEIN, L.D.; MUNGALL,C.; SHU, S.Q. *et al.* **The Generic Genome Browser: A Building Block for a Model Organism System Database**. Genome Research. 12:1599–1610. ISSN 1088-9051/02 2002.

TAIZ, L.; ZEIGER, E. **Fisiologia vegetal.** Tradução. SANTAREM, et al., 3ª ed., Porto Alegre: Artmed. 719, 2013.

TEIXEIRA, P.J.P.L.; THOMAZELLA, D.P.T.; REIS, O.*et al.***High-ResolutionTranscript Profiling of the Atypical Biotrophic Interaction between *Theobroma cocoa* and the Fungal Pathogen *Moniliophthora perniciosa***. The Plant Cell.26: 4245– 4269. 2014.

TGOC The Gene Ontology Consortium. **The Gene Ontology Resource: 20 years and still GOing strong.** Nucleic Acids Res.47, 330-338. 2018

(TAGI) THE ARABIDOPSIS GENOME INITIATIVE. **Analysis of the genome sequence of the flowering plant Arabidopsis thaliana**Disponívelem:<ftp://ftp.arabidopsis.org/home/tair/Sequences/blast_datasets/TAIR10_blastsets/>. Acesso em: 22 outubros 2018.

(UniProt) **The universal protein resource.** Disponível em: <www.uniprot.org>

THE SORGHUM BICOLOR GENOME AND THE DIVERSIFICATION OF GRASSES. Disponível em:<ftp://ftp.jgipsf.org/pub/JGI_data/phytozome/v7.0/Sbicolor/>.

TONG, C.; ZHANG, B.; SHI, J. **A hidden Markov model approach to multilocus linkage analysis in a full-sib family.** Tree Genetics & Genomes6, 651–62, 2010.

TUSKAN, G.A.; DIFAZIO, S.; JANSSON, S. *et al.***T he genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. Disponível em:<ftp://ftp.jgi-psf.org/pub/JGI_data/phytozome/v7.0/Ptrichocarpa/>.

TYLER, R.; .TRIPATHY, S.; ZHANG, X.*et al.****Phytophthora*** **genome sequences uncover evolutionary origins and mechanisms of pathogenesis**. Science. 313. 2006.

VAN LOON, L.C; BAKKER, P.A.H.M; PIETERSE, C.M.J. **Systemic resistance induced by rhizosphere bacteria.** Annual Review Phytopathology. 36: 453-483. 1998.

VAN LOON L.C, R.E.P.M.; PIETERSE, C.M. **Significance of inducible defence-related proteins in infected plants**. Annual Review Phytopathology.44: 135–162. 2006.

VANEGTERN, B.; MICHAEL, R.M.; SCOT, N. **Black Pod Rot of Cocoa Caused by *Phytophthora palmivora*.** Plant Disease. 2015.

VELLO, F.; GARCIA, J.R. **Características das principais variedades de cacau cultivadas na Bahia.** Theobroma.1: 3–10. 1971.

VERRET, F. *et al*. **Calcium channels in photosynthetic eukaryotes: implications for evolution of calcium-based signalling**. New Phytology. 187:23–43. 2010.

WANG, H. & Ng, T. B. **Ginkbilobin, a novel antifungal protein from *Ginkgo biloba* seeds with sequence similarity to embryo-abundant protein**. Biochem. Biophys. Res. Commun. 279, 407–411. doi: 10.1006/bbrc.2000.3929. 2000.

WU, R.; MA, C.X.; PAINTER, I.; ZENG, Z.B. **Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species.** Theoretical Population Biology61, 349–63, 2002.

WU,C.H. *et al.* **PIRSF: family classification system at the Protein Information Resource**. Nucleic Acids Res., 32, 112–114. 2004.

XU, P.; JIANG, L.; WU, J.*et al*. **Isolation and characterization of a pathogenesis related protein 10 gene (GmPR10) with induced expression in soybean (Glycine max) during infection with *Phytophthora sojae*.** Molecular Biology Reports. 41 (8): 4899-909.2014.

ZHANG, D.; MOTILAL, L. **SPRINGER - Cocoa diseases: A history of old enemies and new encounters**. Springer International Publishing. 3-623.2016.

ZHENG, L.; MACKRILL, J.J. **Calcium Signaling in Oomycetes: An Evolutionary Perspective.** Frontiers in Physiology.7. 2016.

ZIPFEL C.; ROBATZEK, S.; NAVARRO, L.*et al*. **Bacterial disease resistance in Arabidopsis through flagell in perception.**Nature. 428: 764–767. 2004.

ZMASEK, C. M.; EDDY, S. R. **ATV: display and manipulation of annotated phylogenetic trees.** Bioinformatics. United Kingdom: Oxford Journals. **17** (4): 383–384. doi:10.1093/bioinformatics/17.4.383. PMID 11301314. 2001.

6 Appendices

Appendix 1. Candidate genes for recognition, activation and combat. (Genome) Reference genome, (QTL) Quantitative Trait Loci, (GC) Gene Class, (Protein ID) Identifier of the protein in its reference genome

Genome	QTL	GC	Protein ID
CRIOLLO	AKAfolICS100chr10	MLO	Tc10v2_p005800.2
CRIOLLO	AKAfolICS100chr10	MLO	Tc10v2_p005800.1
CRIOLLO	AKAfolICS100chr10	UNKNOWN	Tc10v2_p004280.1
CRIOLLO	AKAfolICS100chr10	UNKNOWN	Tc10v2_p004170.1
CRIOLLO	AKAfolICS100chr1	RLP	Tc01v2_p027470.1
CRIOLLO	AKAfolICS100chr3	CNL	Tc03v2_p026110.1
CRIOLLO	AKAfolICS100chr3	CNL	Tc03v2_p026450.1
CRIOLLO	AKAfolICS100chr3	N	Tc03v2_p026140.1
CRIOLLO	AKAfolICS100chr3	N	Tc03v2_p026480.1
CRIOLLO	AKAfolICS100chr3	NL	Tc03v2_p026190.1
CRIOLLO	AKAfolICS100chr3	NL	Tc03v2_p026170.1
CRIOLLO	AKAfolICS100chr3	NL	Tc03v2_p026160.2
CRIOLLO	AKAfolICS100chr3	NL	Tc03v2_p026160.1
CRIOLLO	AKAfolICS100chr3	RLP	Tc03v2_p027150.1
CRIOLLO	AKAfolICS100chr3	UNKNOWN	Tc03v2_p026470.2
CRIOLLO	AKAfolICS100chr3	UNKNOWN	Tc03v2_p026130.1
CRIOLLO	AKAfolICS100chr3	UNKNOWN	Tc03v2_p026890.1
CRIOLLO	AKAfolICS100chr3	UNKNOWN	Tc03v2_p026470.1
CRIOLLO	AKAprrHchr1	CNL	Tc01v2_p002490.1
CRIOLLO	AKAprrHchr1	CNL	Tc01v2_p002490.2
CRIOLLO	AKAprrHchr1	CNL	Tc01v2_p002300.1
CRIOLLO	AKAprrHchr1	RLP	Tc01v2_p001880.1
CRIOLLO	AKAprrHchr1	RPW8NL	Tc01v2_p004390.1
CRIOLLO	AKAprrHchr1	RPW8NL	Tc01v2_p004360.1
CRIOLLO	AKAprrHchr1	RPW8NL	Tc01v2_p004380.1
CRIOLLO	AKAprrHchr1	UNKNOWN	Tc01v2_p002080.1
CRIOLLO	AKAprrHchr1	UNKNOWN	Tc01v2_p002100.1
CRIOLLO	AKAprrHchr1	UNKNOWN	Tc01v2_p003280.1
CRIOLLO	AKAprrHchr8	RLP	Tc08v2_p003150.1
CRIOLLO	AKAprrHchr8	RLP	Tc08v2_p003920.1
CRIOLLO	AKAprrHchr8	RLP	Tc08v2_p002170.2
CRIOLLO	AKAprrHchr8	RLP	Tc08v2_p002170.1
CRIOLLO	AKAprrHchr8	UNKNOWN	Tc08v2_p003050.1
CRIOLLO	AKAprrICS100chr61	CNL	Tc06v2_p010180.2
CRIOLLO	AKAprrICS100chr61	CNL	Tc06v2_p010180.1
CRIOLLO	AKAprrICS100chr61	NL	Tc06v2_p010200.2
CRIOLLO	AKAprrICS100chr61	NL	Tc06v2_p010200.1

CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p010130.1
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p011070.1
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p011050.1
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p011030.1
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p011020.2
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p011020.1
CRIOLLO	AKAprrICS100chr61	RLP	Tc06v2_p010780.1
CRIOLLO	AKAprrICS100chr61	UNKNOWN	Tc06v2_p011450.1
CRIOLLO	AKAprrICS100chr62	RLK	Tc06v2_p003400.1
CRIOLLO	AKAprrICS100chr62	RLP	Tc06v2_p003670.2
CRIOLLO	AKAprrICS100chr62	RLP	Tc06v2_p003670.1
CRIOLLO	AKAprrICS100chr62	UNKNOWN	Tc06v2_p003530.1
CRIOLLO	AKAprrICS100chr62	UNKNOWN	Tc06v2_p003690.1
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001110.2
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001130.4
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001130.2
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001130.3
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001110.3
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001110.4
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001200.1
CRIOLLO	AKAprrICS95chr2	CNL	Tc02v2_p001110.1
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.2
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001190.3
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.3
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.6
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001190.1
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001100.1
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.4
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.1
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001190.2
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.7
CRIOLLO	AKAprrICS95chr2	NL	Tc02v2_p001120.5
CRIOLLO	AKAprrICS95chr2	RLP	Tc02v2_p001540.1
CRIOLLO	AKAprrICS95chr2	RLP	Tc02v2_p002630.1
CRIOLLO	AKAprrICS95chr2	UNKNOWN	Tc02v2_p001130.1
CRIOLLO	AKAprrICS95chr2	UNKNOWN	Tc02v2_p001020.2
CRIOLLO	AKAprrICS95chr2	UNKNOWN	Tc02v2_p002950.1
CRIOLLO	AKAprrICS95chr2	UNKNOWN	Tc02v2_p001170.1
CRIOLLO	AKAprrICS95chr2	UNKNOWN	Tc02v2_p001020.1
CRIOLLO	BARq1BPPcchr1	RLK	Tc01v2_p029920.1
CRIOLLO	BARq1BPPcchr1	RLP	Tc01v2_p027470.1
CRIOLLO	BARq1BPPctchr6	CN	Tc06v2_p000460.1
CRIOLLO	BARq1BPPctchr6	CNL	Tc06v2_p000490.1
CRIOLLO	BARq1BPPctchr6	RLP	Tc06v2_p000350.1
CRIOLLO	BARq1BPPctchr6	RLP	Tc06v2_p000550.1
CRIOLLO	BARq1BPPctchr6	UNKNOWN	Tc06v2_p000470.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016870.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016910.1

CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016880.3
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016900.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016860.2
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016880.2
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016840.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016930.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016900.2
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016860.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016800.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016880.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016920.1
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016870.2
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016920.2
CRIOLLO	BARq1BpPpchr6	RLK	Tc06v2_p016810.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017420.4
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017340.2
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017440.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017360.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017420.3
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017370.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017470.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017300.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017250.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017340.3
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017430.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017350.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017320.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017290.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017330.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017400.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017420.2
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017290.2
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017380.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017340.1
CRIOLLO	BARq1BpPpchr6	RLKGNK2	Tc06v2_p017390.1
CRIOLLO	BARq1BpPpchr6	RLP	Tc06v2_p018760.1
CRIOLLO	BARq1BpPpchr6	RLP	Tc06v2_p016150.1
CRIOLLO	BARq1BpPpchr6	T	Tc06v2_p015980.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p018570.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p018600.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016480.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p018590.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p017860.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p015100.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016540.2
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016540.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016450.1
CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016530.2

CRIOLLO	BARq1BpPpchr6	UNKNOWN	Tc06v2_p016530.1
CRIOLLO	BARq2BPPcchr2	UNKNOWN	Tc02v2_p014020.1
CRIOLLO	BARq2BPPcchr2	UNKNOWN	Tc02v2_p013390.1
CRIOLLO	BARq2BPPcchr2	UNKNOWN	Tc02v2_p013940.1
CRIOLLO	BARq2BPPcchr2	UNKNOWN	Tc02v2_p013380.1
CRIOLLO	BARq3BPPcchr3	CNL	Tc03v2_p026110.1
CRIOLLO	BARq3BPPcchr3	CNL	Tc03v2_p026450.1
CRIOLLO	BARq3BPPcchr3	N	Tc03v2_p026140.1
CRIOLLO	BARq3BPPcchr3	N	Tc03v2_p026480.1
CRIOLLO	BARq3BPPcchr3	NL	Tc03v2_p026160.1
CRIOLLO	BARq3BPPcchr3	NL	Tc03v2_p026170.1
CRIOLLO	BARq3BPPcchr3	NL	Tc03v2_p026190.1
CRIOLLO	BARq3BPPcchr3	NL	Tc03v2_p026160.2
CRIOLLO	BARq3BPPcchr3	RLP	Tc03v2_p027150.1
CRIOLLO	BARq3BPPcchr3	UNKNOWN	Tc03v2_p026470.2
CRIOLLO	BARq3BPPcchr3	UNKNOWN	Tc03v2_p026470.1
CRIOLLO	BARq3BPPcchr3	UNKNOWN	Tc03v2_p026890.1
CRIOLLO	BARq3BPPcchr3	UNKNOWN	Tc03v2_p026130.1
CRIOLLO	BARq4BPPcchr4	RLK	Tc04v2_p018750.1
CRIOLLO	BARq4BPPcchr4	RLP	Tc04v2_p019600.1
CRIOLLO	BARq4BPPcchr4	UNKNOWN	Tc04v2_p018480.1
CRIOLLO	BARq4BPPcchr4	UNKNOWN	Tc04v2_p019390.1
CRIOLLO	BARq4BPPcchr4	UNKNOWN	Tc04v2_p018530.1
CRIOLLO	BROPHY1chr4	RLKGNK2	Tc04v2_p001170.1
CRIOLLO	BROPHY1chr4	RLKGNK2	Tc04v2_p001160.1
CRIOLLO	BROPHY1chr4	RLP	Tc04v2_p000470.1
CRIOLLO	BROPHY2chr8	CNL	Tc08v2_p006710.1
CRIOLLO	BROPHY2chr8	CNL	Tc08v2_p006690.1
CRIOLLO	BROPHY2chr8	CNL	Tc08v2_p006710.2
CRIOLLO	BROPHY2chr8	CNL	Tc08v2_p006730.1
CRIOLLO	BROPHY2chr8	MLO	Tc08v2_p005430.1
CRIOLLO	BROPHY2chr8	MLO	Tc08v2_p005430.2
CRIOLLO	BROPHY2chr8	MLO	Tc08v2_p005450.1
CRIOLLO	BROPHY2chr8	RLP	Tc08v2_p007290.1
CRIOLLO	BROPHY2chr8	RLP	Tc08v2_p006800.1
CRIOLLO	BROPHY2chr8	RLP	Tc08v2_p006810.1
CRIOLLO	BROPHY2chr8	UNKNOWN	Tc08v2_p006220.1
CRIOLLO	BROPHY2chr8	UNKNOWN	Tc08v2_p005750.1
CRIOLLO	BROPHY3chr10	CN	Tc10v2_p011880.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.4
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.6
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p013430.2
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011850.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.7
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p010260.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p013440.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.12
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011930.1

CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012140.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.15
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.3
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.14
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.17
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.5
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012150.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011910.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.11
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p010220.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.9
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011860.4
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.10
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011860.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p010220.3
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.16
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012160.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.2
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.8
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p012180.13
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011860.2
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011860.3
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p013430.1
CRIOLLO	BROPHY3chr10	CNL	Tc10v2_p011920.14
CRIOLLO	BROPHY3chr10	N	Tc10v2_p011890.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.6
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013340.3
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013350.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.9
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.8
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013400.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013340.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012080.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.7
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013410.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013360.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.5
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013340.2
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012100.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.3
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.1
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.2
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p012170.4
CRIOLLO	BROPHY3chr10	NL	Tc10v2_p013390.1
CRIOLLO	BROPHY3chr10	RLKGNK2	Tc10v2_p012610.1
CRIOLLO	BROPHY3chr10	RLP	Tc10v2_p011640.1
CRIOLLO	BROPHY3chr10	RLP	Tc10v2_p011580.1
CRIOLLO	BROPHY3chr10	RLP	Tc10v2_p010160.1

CRIOLLO	BROPHY3chr10	RLP	Tc10v2_p011560.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012980.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p013370.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p010240.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p013420.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012850.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012830.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p011900.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012760.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012770.2
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p013380.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p011600.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p013380.2
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012700.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p011590.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012700.2
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012720.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012930.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p012750.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p011710.1
CRIOLLO	BROPHY3chr10	UNKNOWN	Tc10v2_p010310.1
MATINA	AKAfolICS100chr10	UNKNOWN	3244038
MATINA	AKAfolICS100chr1	UNKNOWN	3212940
MATINA	AKAfolICS100chr1	UNKNOWN	3211130
MATINA	AKAfolICS100chr1	UNKNOWN	3211656
MATINA	AKAfolICS100chr1	UNKNOWN	3214102
MATINA	AKAfolICS100chr1	UNKNOWN	3212639
MATINA	AKAfolICS100chr3	CN	3344436
MATINA	AKAfolICS100chr3	CNL	3344449
MATINA	AKAfolICS100chr3	CNL	3344063
MATINA	AKAfolICS100chr3	CNL	3344500
MATINA	AKAfolICS100chr3	N	3344016
MATINA	AKAfolICS100chr3	N	3344488
MATINA	AKAfolICS100chr3	NL	3344454
MATINA	AKAfolICS100chr3	RLP	3343189
MATINA	AKAfolICS100chr3	UNKNOWN	3344057
MATINA	AKAfolICS100chr3	UNKNOWN	3343513
MATINA	AKAfolICS100chr3	UNKNOWN	3344431
MATINA	AKAprrHchr1	CNL	3184683
MATINA	AKAprrHchr1	CNL	3184894
MATINA	AKAprrHchr1	RLP	3184106
MATINA	AKAprrHchr1	RPW8NL	3196185
MATINA	AKAprrHchr1	RPW8NL	3196210
MATINA	AKAprrHchr1	UNKNOWN	3185860
MATINA	AKAprrHchr1	UNKNOWN	3184356
MATINA	AKAprrHchr1	UNKNOWN	3184361
MATINA	AKAprrHchr8	RLP	3507182
MATINA	AKAprrHchr8	RLP	3508342

MATINA	AKAprrHchr8	RLP	3506200
MATINA	AKAprrHchr8	UNKNOWN	3507278
MATINA	AKAprrICS100chr6_1	UNKNOWN	3464167
MATINA	AKAprrICS100chr6_1	UNKNOWN	3464843
MATINA	AKAprrICS100chr6_1	UNKNOWN	3465023
MATINA	AKAprrICS100chr6_1	UNKNOWN	3464909
MATINA	AKAprrICS100chr6_2	CNL	3446028
MATINA	AKAprrICS100chr6_2	CNL	3446023
MATINA	AKAprrICS100chr6_2	RLP	3446117
MATINA	AKAprrICS100chr6_2	RLP	3445147
MATINA	AKAprrICS95chr2	CNL	3267623
MATINA	AKAprrICS95chr2	CNL	3267589
MATINA	AKAprrICS95chr2	CNL	3267585
MATINA	AKAprrICS95chr2	NL	3267643
MATINA	AKAprrICS95chr2	NL	3267613
MATINA	AKAprrICS95chr2	NL	3267653
MATINA	AKAprrICS95chr2	NL	3267579
MATINA	AKAprrICS95chr2	UNKNOWN	3267416
MATINA	AKAprrICS95chr2	UNKNOWN	3268296
MATINA	AKAprrICS95chr2	UNKNOWN	3268309
MATINA	BARq1BPPcchr1	CNL	3222330
MATINA	BARq1BPPcchr1	CNL	3222462
MATINA	BARq1BPPcchr1	CNL	3222320
MATINA	BARq1BPPcchr1	CNL	3222500
MATINA	BARq1BPPcchr1	RLP	3227564
MATINA	BARq1BPPcchr1	RLP	3223656
MATINA	BARq1BPPcchr1	UNKNOWN	3223367
MATINA	BARq1BPPcchr1	UNKNOWN	3224250
MATINA	BARq1BPPcchr1	UNKNOWN	3222923
MATINA	BARq1BPPctchr6	CNL	3462782
MATINA	BARq1BPPctchr6	CNL	3462815
MATINA	BARq1BPPctchr6	RLP	3462662
MATINA	BARq1BPPctchr6	RLP	3462976
MATINA	BARq1BPPctchr6	UNKNOWN	3462709
MATINA	BARq1BpPpchr6	RLK	3451662
MATINA	BARq1BpPpchr6	RLK	3451861
MATINA	BARq1BpPpchr6	RLK	3451845
MATINA	BARq1BpPpchr6	RLK	3451646
MATINA	BARq1BpPpchr6	RLK	3451708
MATINA	BARq1BpPpchr6	RLK	3451759
MATINA	BARq1BpPpchr6	RLKGNK2	3450801
MATINA	BARq1BpPpchr6	RLKGNK2	3451094
MATINA	BARq1BpPpchr6	RLKGNK2	3451045
MATINA	BARq1BpPpchr6	RLKGNK2	3451072
MATINA	BARq1BpPpchr6	RLKGNK2	3450891
MATINA	BARq1BpPpchr6	RLKGNK2	3451166
MATINA	BARq1BpPpchr6	RLKGNK2	3451244
MATINA	BARq1BpPpchr6	RLKGNK2	3451197

MATINA	BARq1BpPpchr6	RLKGNK2	3451228
MATINA	BARq1BpPpchr6	RLKGNK2	3451139
MATINA	BARq1BpPpchr6	RLKGNK2	3451271
MATINA	BARq1BpPpchr6	RLKGNK2	3450996
MATINA	BARq1BpPpchr6	RLKGNK2	3450973
MATINA	BARq1BpPpchr6	RLKGNK2	3451011
MATINA	BARq1BpPpchr6	RLKGNK2	3450875
MATINA	BARq1BpPpchr6	RLP	3449263
MATINA	BARq1BpPpchr6	RLP	3452732
MATINA	BARq1BpPpchr6	T	3452900
MATINA	BARq1BpPpchr6	UNKNOWN	3452124
MATINA	BARq1BpPpchr6	UNKNOWN	3449575
MATINA	BARq1BpPpchr6	UNKNOWN	3452246
MATINA	BARq1BpPpchr6	UNKNOWN	3449561
MATINA	BARq1BpPpchr6	UNKNOWN	3449566
MATINA	BARq1BpPpchr6	UNKNOWN	3452184
MATINA	BARq1BpPpchr6	UNKNOWN	3453912
MATINA	BARq1BpPpchr6	UNKNOWN	3452116
MATINA	BARq1BpPpchr6	UNKNOWN	3450344
MATINA	BARq2BPPcchr2	RLKGNK2	3318311
MATINA	BARq2BPPcchr2	UNKNOWN	3318761
MATINA	BARq2BPPcchr2	UNKNOWN	3318922
MATINA	BARq2BPPcchr2	UNKNOWN	3317758
MATINA	BARq2BPPcchr2	UNKNOWN	3317751
MATINA	BARq3BPPcchr3	CN	3344436
MATINA	BARq3BPPcchr3	CNL	3344500
MATINA	BARq3BPPcchr3	CNL	3344063
MATINA	BARq3BPPcchr3	CNL	3344449
MATINA	BARq3BPPcchr3	N	3344016
MATINA	BARq3BPPcchr3	N	3344488
MATINA	BARq3BPPcchr3	NL	3344454
MATINA	BARq3BPPcchr3	RLP	3343189
MATINA	BARq3BPPcchr3	UNKNOWN	3344057
MATINA	BARq3BPPcchr3	UNKNOWN	3343513
MATINA	BARq3BPPcchr3	UNKNOWN	3344431
MATINA	BARq4BPPcchr4	RLK	3379026
MATINA	BARq4BPPcchr4	RLP	3375376
MATINA	BARq4BPPcchr4	UNKNOWN	3379407
MATINA	BARq4BPPcchr4	UNKNOWN	3379338
MATINA	BARq4BPPcchr4	UNKNOWN	3378443
MATINA	BROPHY1chr4	RLKGNK2	3365094
MATINA	BROPHY1chr4	RLP	3364140
MATINA	BROPHY2chr8	CNL	3502117
MATINA	BROPHY2chr8	CNL	3502107
MATINA	BROPHY2chr8	CNL	3502049
MATINA	BROPHY2chr8	MLO	3504014
MATINA	BROPHY2chr8	MLO	3504085
MATINA	BROPHY2chr8	RLP	3501305

MATINA	BROPHY2chr8	RLP	3501983
MATINA	BROPHY2chr8	RLP	3501992
MATINA	BROPHY2chr8	UNKNOWN	3503001
MATINA	BROPHY2chr8	UNKNOWN	3503681
MATINA	BROPHY3chr10	CNL	3254341
MATINA	BROPHY3chr10	CNL	3257344
MATINA	BROPHY3chr10	CNL	3258828
MATINA	BROPHY3chr10	CNL	3256755
MATINA	BROPHY3chr10	CNL	3257002
MATINA	BROPHY3chr10	CNL	3257102
MATINA	BROPHY3chr10	CNL	3257324
MATINA	BROPHY3chr10	CNL	3257371
MATINA	BROPHY3chr10	CNL	3258892
MATINA	BROPHY3chr10	NL	3253595
MATINA	BROPHY3chr10	NL	3258850
MATINA	BROPHY3chr10	NL	3257309
MATINA	BROPHY3chr10	NL	3257298
MATINA	BROPHY3chr10	NL	3258821
MATINA	BROPHY3chr10	RLKGNK2	3258102
MATINA	BROPHY3chr10	RLP	3256681
MATINA	BROPHY3chr10	RLP	3256715
MATINA	BROPHY3chr10	RLP	3254087
MATINA	BROPHY3chr10	RLP	3256785
MATINA	BROPHY3chr10	UNKNOWN	3257050
MATINA	BROPHY3chr10	UNKNOWN	3254276
MATINA	BROPHY3chr10	UNKNOWN	3259317
MATINA	BROPHY3chr10	UNKNOWN	3258498
MATINA	BROPHY3chr10	UNKNOWN	3257945
MATINA	BROPHY3chr10	UNKNOWN	3255072
MATINA	BROPHY3chr10	UNKNOWN	3259346
MATINA	BROPHY3chr10	UNKNOWN	3256700
MATINA	BROPHY3chr10	UNKNOWN	3254281
MATINA	BROPHY3chr10	UNKNOWN	3256837
MATINA	BROPHY3chr10	UNKNOWN	3258221
MATINA	BROPHY3chr10	UNKNOWN	3254467
MATINA	BROPHY3chr10	UNKNOWN	3258264
MATINA	BROPHY3chr10	UNKNOWN	3258374
MATINA	BROPHY3chr10	UNKNOWN	3256961
MATINA	BROPHY3chr10	UNKNOWN	3256939
MATINA	BROPHY3chr10	UNKNOWN	3259166
MATINA	BROPHY3chr10	UNKNOWN	3259285
MATINA	BROPHY3chr10	UNKNOWN	3253920
MATINA	BROPHY3chr10	UNKNOWN	3255812
MATINA	BROPHY3chr10	UNKNOWN	3256909
MATINA	BROPHY3chr10	UNKNOWN	3258360
MATINA	BROPHY3chr10	UNKNOWN	3256739
MATINA	BROPHY3chr10	UNKNOWN	3256704