



Guilherme de Toledo e Silva

Estudos genético-moleculares em *Panicum maximum*: mapeamento genético-molecular e
análise de transcriptoma via RNA-seq

*Genetic and molecular studies in Panicum maximum: genetic and molecular mapping and
transcriptome analysis via RNA-seq*

Campinas

2013



UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Biologia



Guilherme de Toledo e Silva

Estudos genético-moleculares em *Panicum maximum*: mapeamento genético-molecular e análise de transcriptoma via RNA-seq

Genetic and molecular studies in Panicum maximum: genetic and molecular mapping and transcriptome analysis via RNA-seq

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Genética e Biologia Molecular na área de Genética Vegetal e Melhoramento

Thesis presented to the Institute of Biology of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Genetics and Molecular Biology, in the area of Plant Genetics and Genetic Breeding

Orientadora/ Supervisor: Profa. Dra. Anete Pereira de Souza

Co-orientadora/ Co-supervisor: Dra. Liana Jank

Este exemplar corresponde à versão final da tese defendida pelo aluno Guilherme de Toledo e Silva e orientada pela Profa. Dra. Anete Pereira de Souza.

Profa. Dra. Anete Pereira de Souza

CAMPINAS

2013

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

T575e Toledo-Silva, Guilherme, 1983-
Estudos genético-moleculares em *Panicum maximum* : mapeamento genético-molecular e análise de transcriptoma via RNA-seq / Guilherme de Toledo e Silva.
– Campinas, SP : [s.n.], 2014.

Orientador: Anete Pereira de Souza.

Coorientador: Liana Jank.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Capim-guine. 2. Microsatélites (Genética). 3. Transcriptoma. 4. RNA-seq. 5. Mapeamento cromossômico. I. Souza, Anete Pereira de, 1962-. II. Jank, Liana. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Genetic and molecular studies in *Panicum maximum* : genetic and molecular mapping and transcriptome analysis via RNA-seq

Palavras-chave em inglês:

Guinea grass

Microsatellites (Genetics)

Transcriptome

RNA-seq

Chromosome Mapping

Área de concentração: Genética Vegetal e Melhoramento

Titulação: Doutor em Genética e Biologia Molecular

Banca examinadora:

Anete Pereira de Souza [Orientador]

Gabriel Rodrigues Alves Margarido

Sérgio Furtado dos Reis

Patricia Menezes Santos

Flávio Antônio Maës dos Santos

Data de defesa: 28-02-2014

Programa de Pós-Graduação: Genética e Biologia Molecular

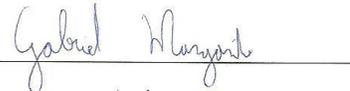
Campinas, 28 de Fevereiro de 2014

Banca Examinadora

Profa. Dra. Anete Pereira de Souza (orientadora)


Assinatura

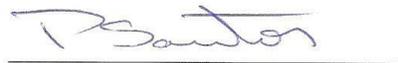
Profa. Dr. Gabriel Rodrigues Alves Margarido


Assinatura

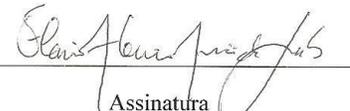
Profa. Dr. Sérgio Furtado dos Reis


Assinatura

Prof. Dra. Patricia Menezes Santos


Assinatura

Prof. Dr. Flávio Antônio Maës dos Santos


Assinatura

Prof. Dr. Michel Georges Albert Vincentz

Assinatura

Prof. Dr. Rodrigo Gazaffi

Assinatura

Dr. Rafael Silva Oliveira

Assinatura

RESUMO

No Brasil, a pecuária bovina é baseada principalmente na utilização de pastagens cultivadas para alimentação animal. A espécie *Panicum maximum* Jacq., popularmente conhecida no Brasil como capim colonião, encontra-se entre as espécies mais utilizadas na alimentação do gado de corte. Grande parte das áreas destinadas a pastagens encontra-se estabelecida com cultivares exóticas e de reprodução clonal. O monocultivo representa sério risco para todos os sistemas de produção baseados no uso de pastagens. Nesse sentido, o melhoramento genético de forrageiras e o lançamento de novas cultivares de *P. maximum* surgem como alternativa para a diversificação das pastagens. O presente trabalho teve como objetivo principal contribuir para o conhecimento básico sobre a genética e a biologia molecular de *P. maximum*. Primeiramente, foi realizada a montagem *de novo* e análise do transcriptoma da espécie utilizando a metodologia de sequenciamento massivo paralelo de cDNA (RNA-seq), resultando em 38.192 unigenes, os quais foram anotados em diferentes bancos de dados. A maioria dos genes relacionados as vias de fixação de carbono C₄ e de síntese de lignocelulose foram identificados. Ainda, foram identificados 5.035 marcadores microssatélites (SSR) e 346.456 marcadores do tipo *single nucleotide polymorphism* (SNP) em todo o transcriptoma. Na segunda parte deste trabalho, foram desenvolvidos 131 novos marcadores moleculares do tipo microssatélite para *P. maximum*. Estes marcadores foram utilizados juntamente a 43 marcadores SSR obtidos da literatura, para construção de mapas genético-moleculares, através da genotipagem dos híbridos F1 resultantes de cruzamento intraespecífico. Os mapas de ligação cobriram 672,2 cM e 802,2 cM dos genomas dos genitores S10 e Mombaça. Estes resultados contribuem para a pesquisa associada a *P. maximum* e forrageiras tropicais, assim como para os programas de melhoramento genético.

ABSTRACT

In Brazil, livestock feeding is mainly based on cultivated pastures. *Panicum maximum* Jacq., also known as capim-colonião, figures among the most cultivated tropical forage plants in Brazilian pastures, which are established with exotic cultivars of clonal reproduction, leading to monoculture, and consequently offering serious risks to associated production systems. One possible solution to monoculture relies on new cultivar releases through breeding programs, diversifying livestock pastures in Brazil. The work presented here had an objective of elucidating some basic aspects regarding the genetics and molecular biology of *P. maximum*. First, we conducted a *de novo* assembly and analysis of leaf transcriptome, through massive parallel cDNA sequencing (RNA-seq), resulting in 38.192 unigenes, which were annotated in different databases. Genes related to C₄ metabolism and lignocellulose synthesis are valuable resources to breeding programs and were identified among assembled transcripts. Thus, several putative molecular markers were located in unigenes: 5.035 microsatellites (SSR) motifs and 346.456 single nucleotide polymorphism (SNP) positions. Also, we developed 131 new genomic and expressed SSR markers for *P. maximum*. These markers were used with 43 published SSR markers to develop a genetic map, through the genotyping of F1 hybrids from an interspecific cross. Linkage maps covered 672,2 cM and 802,2 cM of parentals S10 and Mombaça genomes respectively. Results presented in this work contribute to *P. maximum* and tropical grasses related research and breeding programs.

SUMÁRIO

vii	Resumo
ix	Abstract
xiii	Agradecimentos
xvii	Organização da tese
1	Introdução
3	Revisão bibliográfica
21	Objetivos
23	Capítulo I
	<i>De novo transcriptome assembly for the tropical grass Panicum maximum Jacq.</i>
41	Capítulo II
	Mapeamento genético-molecular de <i>Panicum maximum</i> Jacq., gramínea forrageira tropical autotetraploide
92	Considerações finais
99	Resumo dos resultados
101	Conclusões
103	Perspectivas
105	Literatura citada
119	Anexo I

AGRADECIMENTOS

Aos meus pais, Joaquim e Marlene, e minha irmã Ticiane, por me apoiarem incondicionalmente por todos estes anos, aturando minhas constantes indecisões sobre o rumo de minha vida. Obrigado por todo carinho, incentivo e paciência.

À minha querida noiva Aline, por me mostrar a melhor definição de companheirismo que pude ter nesta vida. Foi a pessoa que, através da sua maneira delicada de ver a vida, me trouxe de volta ao caminho da pesquisa em uma época que estava totalmente desiludido, decidido a recomeçar minha vida. Seu incentivo a realizar esse doutorado na Unicamp foi crucial, sem ela certamente nada disso teria acontecido. Obrigado por ter tomado decisões tão importantes, logo cedo, em um momento em estávamos ainda nos conhecendo. Obrigado por me ensinar a ser uma pessoa melhor, e por iluminar cada dia de minha vida. E por ser a melhor pessoa que eu poderia ter ao meu lado, hoje e sempre.

À professora Dra. Anete Pereira de Souza. Por ter me ensinado tanto em tão pouco tempo, se tornando minha principal referência na pesquisa. Através da convivência pude entender como é possível crescer profissionalmente através de trabalho honesto, ético e competente. Sua paixão pela ciência, e a forma como luta pelos seus ideais serão sempre um norte em minha carreira. Obrigado por ter me aceitado como aluno, e pelas inúmeras oportunidades dadas ao longo deste doutorado. Espero ter correspondido as expectativas depositadas sobre mim, e que possamos realizar muitos projetos juntos ao longo dos anos que virão.

À minha co-orientadora Dra. Liana Jank, uma pessoa muito especial que tive a sorte de me aceitar como seu co-orientado. Obrigado por me acolher de maneira tão especial em minhas visitas a Embrapa Gado de Corte, por todos ensinamentos e conversas que tivemos em nossos momentos juntos. Espero ter feito um trabalho que faça jus a todo o esforço despendido para que eu tivesse amostras de qualidade e assim realizar meu trabalho.

Ao Dr. Juverlande, uma pessoa que acreditou em mim sem conhecer nada ao meu respeito, sendo crucial no processo de conseguir um lugar junto ao laboratório da professora Anete. Posso dizer sem sombra de dúvidas que foi um segundo pai pra mim durante esses anos, e desejo o melhor nessa vida.

Aos colegas do Laboratório de Análise Genética e Molecular (LAGM), por terem me acolhido, aguentando todos meus altos e baixos, minhas inúmeras dúvidas sobre genética, marcadores e tudo que era completamente novo para mim. Agradeço também por todos churrascos e

comemorações em nosso espaço gourmet. Não citarei nomes pelo medo de esquecer alguém. Um grande abraço a todos, de coração.

Dentre os colegas do LAGM, alguns se envolveram mais com meu trabalho, me ajudando além do esperado, visto que tinham inúmeras atribuições e mesmo assim despenderam parte de seu tempo. Obrigado Melissa, Livia, Prianda, Gustavo, Mauricio, Danilo, Melina e Benício.

Ao Dr. Antônio Augusto Franco Garcia e ao Dr. Marcelo Mollinari por todo apoio e direcionamento na construção dos mapas genéticos.

Aos professores membros de Exame de Qualificação, Banca Prévia e Banca de Defesa de Tese, por terem aceitados os convites, pelos ensinamentos, críticas e sugestões.

À Stephanie, que participou deste trabalho como aluna de iniciação científica, gerando resultados valiosos para a conclusão desta tese. Obrigado pelas sinceras conversas e toda dedicação e comprometimento. Que o caminho escolhido lhe traga a satisfação tão desejada.

Aos colegas Mauricio Durigan e Gustavo Mori, pela amizade construída nestes anos, por terem se tornado grandes amigos, pelo companheirismo nos inúmeros shows que fomos, pelas extensas noites de videogame, enfim, por todos momentos bons e ruins que passamos. Obrigado por tudo, desejo o melhor para suas vidas pessoais e profissionais. E um tomahawk em cada um no último segundo.

Aos colegas dos outros laboratórios que pude conhecer, em especial ao professor Dr. Michel Georges Albert Vincentz e seus alunos. O mesmo sentimento se estende ao professor Dr. Antônio Augusto Franco Garcia e seu grupo, que sempre me receberam da melhor maneira em minhas visitas a USP-ESALQ. Agradeço também aos doutores Renato Vincentini, Sérgio Furtado dos Reis e Maria Imaculada Zucchi.

A Tânia, Sandra, Gabriela, Eliseu, Patrícia, Aline, Alessandra, Lourdes e Beatriz. Obrigado por todo apoio prestado, pelos inúmeros auxílios de última hora e pelas ótimas conversas ao longo destes anos.

Aos colegas de Florianópolis e São José dos Campos, que a cada retorno, me receberam com muita alegria e satisfação. Obrigado pelo apoio e incentivo dado durante todo esse período.

À Universidade Estadual de Campinas e ao programa de pós-graduação em Genética e Biologia Molecular.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelas bolsas de doutorado que permitiram a realização deste trabalho.

ORGANIZAÇÃO DA TESE

Este trabalho busca elucidar algumas questões básicas relacionadas à genética e à biologia molecular da forrageira tropical *Panicum maximum* Jacq. Esta espécie, apesar de sua importância para a pecuária nacional e com potencial bioenergético, tem seu perfil molecular pouco estudado. Os resultados obtidos durante o período de doutoramento estão apresentados nesta tese no formato de dois artigos, dos quais um já foi publicado (Capítulo I).

O primeiro artigo, intitulado “*De Novo* Transcriptome Assembly for the Tropical Grass *Panicum maximum* Jacq.” e apresentado no Capítulo I, foi publicado na revista Plos One (DOI: 10.1371/journal.pone.0070781), refere-se a análise do transcriptoma de folhas de *P. maximum*, através de sequenciamento em larga escala de cDNA. Foram utilizados quatro genótipos que participam do programa de melhoramento genético da espécie na Embrapa Gado de Corte (Campo Grande – MS). Através de uma montagem *de novo*, foi possível reconstruir grande parte dos transcritos presentes nas folhas de *P. maximum*. O segundo artigo, “Mapeamento genético da forrageira autotetraploide *Panicum maximum* através de marcadores microssatélites”, ainda será submetido para publicação em periódico científico com seletiva política editorial. Esse artigo apresenta o desenvolvimento de 131 novos marcadores microssatélites e a construção de um mapa de ligação para a espécie, utilizando uma população de mapeamento localizada na Embrapa Gado de Corte (Campo Grande – MS), originada do cruzamento de genitores contrastantes quanto a importantes características de produção. A metodologia apropriada para o desenvolvimento de mapas genéticos em espécies autotetraploides foi utilizada pela primeira vez em *P. maximum*. Este trabalho é apresentado no Capítulo II.

Durante o primeiro ano do período de doutorado, realizou-se uma análise da diversidade genética presente no banco de germoplasma da leguminosa forrageira *Stylosanthes capitata*, localizado na Embrapa Cerrados (Planaltina - DF). Além da elucidação da estrutura

genética do banco de germoplasma, foi sugerida uma coleção nuclear de acessos que incluisse 100% dos alelos encontrados entre os genótipos. Esses resultados foram publicados no artigo “Using genetic diversity information to establish core collections of *Stylosanthes capitata* and *Stylosanthes macrocephala*” na revista *Genetics and Molecular Biology*” (DOI: 10.1590/S1415-47572012005000076). Este artigo encontra-se no Anexo I.

Os resultados apresentados nesta tese de doutorado trazem conhecimento inédito para importantes espécies de forrageiras tropicais, informações estas, que poderão ser aplicadas ao programa de melhoramento genético de *P. maximum*, *S. capitata* e *S. macrocephala*, bem como a estudos relacionados a biologia dessas espécies, assim como a outras gramíneas relacionadas.

INTRODUÇÃO

Panicum maximum (*Megathyrsus maximus* (Jacq.) Simon BK & Jacobs SWL), popularmente conhecido no Brasil como capim colonião, é de origem africana, mas devido sua alta adaptação e distribuição pelo território nacional, costuma ser considerada como espécie nativa. O plantio de *P. maximum* expandiu-se devido à sua grande adaptação, produtividade e qualidade, fácil propagação por sementes e por ser altamente palatável ao gado (JANK, 1995; JANK et al., 1997). Encontra-se entre as espécies mais utilizadas nas pastagens cultivadas. Devido a sua importância para a pecuária, existe um programa de melhoramento da espécie na Embrapa Gado de Corte (Campo Grande - MS). Além desta aplicação, gramíneas forrageiras como *P. maximum* também são opções à obtenção de energia a partir da biomassa. Neste sentido, o capim elefante (*Pennisetum purpureum* Schum.), é uma das forrageiras mais produtivas, porém sua propagação vegetativa dificulta a implantação. A espécie *P. maximum* pode ser considerada como uma alternativa viável, por também apresentar alta produtividade e ser propagada por sementes, sendo uma opção para o uso de biomassa e/ou seus resíduos para produção de etanol celulósico (COSTA e JANK, 2009; JANK et al., 2010).

A ciência aplicada ao melhoramento de forrageiras em países tropicais é relativamente jovem e somente foi intensificada nas últimas décadas. Estas gramíneas ainda não são consideradas domesticadas, e ciclos contínuos de cultivo e seleção são necessários para adequar às necessidades de mercado. Existe pouco conhecimento sobre a influência genética nas características agrônômicas de interesse, assim como o mecanismo envolvido nestas relações genótipo-fenótipo (JANK, VALLE e RESENDE, 2011). A maior parte da pesquisa em *P. maximum* envolve distintos trabalhos que abordam questões de interesse agrônômico e morfológico. Dessa forma, diversos aspectos no âmbito genético-molecular ainda são desconhecidos.

Trabalhos relacionados ao transcriptoma de *P. maximum* e os respectivos mecanismos de

regulação da expressão gênica são escassos, geralmente focando nos genes envolvidos com o processo reprodutivo (CHEN et al., 1999, 2005; VASIL et al., 1988; YAMADA-AKIYAMA et al., 2009). O banco de genes público Genbank (BENSON et al., 2013) apresenta apenas 445 sequências referentes a *P. maximum*, sendo 196 destas de transcritos não anotados. Através da metodologia de sequenciamento massivo paralelo de cDNA (RNA-seq), o transcriptoma de qualquer espécie pode ser estudado, incluindo as espécies não modelo (MARGUERAT e BÄHLER, 2010). Neste trabalho, foi analisado pela primeira vez o conjunto de transcritos presentes em folhas de *P. maximum* (TOLEDO-SILVA et al., 2013), o que possibilitará uma gama de estudos relacionados, como por exemplo, o descobrimento de mecanismos moleculares e a elucidação de vias metabólicas de interesse.

Outra carência em relação ao conhecimento biológico da espécie reside no fato de existir apenas um mapa de ligação para *P. maximum*, desenvolvido por pesquisadores japoneses (EBINA et al., 2005). Através do desenvolvimento e genotipagem de novos marcadores moleculares microssatélites (SSR – *simple sequence repeat*), este trabalho teve como objetivo a construção de um mapa genético para a espécie *P. maximum*, a partir de um cruzamento intraespecífico realizado na Embrapa Gado de Corte (Campo Grande – MS). O desenvolvimento de um mapa genético-molecular tem como objetivo identificar locos relacionados a características quantitativas de interesse (*quantitative trait loci* ou QTL), que estão sendo analisadas na Embrapa Gado de Corte. O uso efetivo da variabilidade genética para o melhoramento da espécie requer a compreensão da arquitetura genética das características desejadas pelos programas de melhoramento genético.

REVISÃO DE LITERATURA

Panicum maximum

A espécie *Panicum maximum*, popularmente conhecida como capim colonião, é a forrageira propagada por sementes mais produtiva no mercado brasileiro. Ela se adapta facilmente a diversos tipos de clima e solos, mas é exigente em fertilidade. Como espécie forrageira, sua qualidade é excelente e tem sido responsável por grande parte da engorda de bovinos no Brasil e em vários países latino-americanos (JANK et al., 2008). *P. maximum* é de origem africana (África do Leste), mais especificamente das regiões do Quênia e da Tanzânia (BOGDAN, 1977; SKERMAN e RIVEROS, 1990). Os primeiros exemplares dessa espécie introduzidos no Brasil vieram da África Ocidental nos navios negreiros, onde eram utilizados como cama para os escravos, e uma vez aqui, se alastraram rapidamente, dando origem à primeira cultivar, o Colonião (JANK, 1995). É uma espécie tão bem adaptada que é considerada nativa em muitas regiões do Brasil (SAVIDAN et al., 1989).

Panicum maximum apresenta hábito de crescimento cespitoso e estolonífero, podendo alcançar até 1,80 m de altura (JANK et al., 2008). Seu sistema radicular bem desenvolvido garante resistência a seca e sobrevivência em condições sub-ótimas. São encontradas folhas decumbentes e eretas, e inflorescências do tipo panícula (JANK, 1995) (Figura 1). A espécie apresenta ciclo vegetativo perene, com propagação por sementes e modo de reprodução apomítico. A apomixia é presente em mais de 35 famílias de plantas (GAUER e CAVALLI-MOLINA, 2000), sendo este o modo reprodutivo da maioria das gramíneas forrageiras tropicais de importância econômica. A apomixia caracteriza-se como um tipo de reprodução totalmente assexual, no qual o embrião desenvolve-se em sementes férteis a partir de divisões mitóticas de uma célula do nucelo, sem a união prévia de gametas. Espécies apomíticas como *P. maximum* apresentam todas as sementes com genótipo da planta mãe (DALL'AGNOL e SCHIFINO-WITTMANN, 2005; SAVIDAN, 2000).

Panicum maximum pertence à família Poaceae, subfamília Panicoideae, tribo Paniceae, e forma um complexo agâmico com outras duas espécies, *Panicum infestum* Anders e *Panicum trichocladum* K. Schum., o que é possível por compartilharem do mesmo número cromossômico ($x=8$) e o mesmo nível de ploidia (AKIYAMA et al., 2008; SAVIDAN et al., 1989). Desta forma, a forrageira *P. maximum* comporta-se como uma espécie autotetraploide ($2n=4x=32$) (NAKAJIMA et al., 1979; SAVIDAN e PERNÈS, 1982). Porém, em seu centro de origem, foram encontrados genótipos apresentando diferentes números cromossômicos, com exemplares diploides ($2n=16$), triploides ($2n=24$), pentaploides ($2n=40$), hexaplóides ($2n=48$), octaploides ($2n=64$), nonaploides ($2n=72$) e também plantas com números cromossômicos irregulares ($2n=30, 31, 34, 36, 37, 38$) (BOGDAN, 1977). Genótipos diploides sexuais foram encontrados na natureza, em contraste com as plantas apomíticas que são tetraploides (NAKAJIMA et al., 1979; SAVIDAN e PERNÈS, 1982). A hibridação de plantas apomíticas com plantas de reprodução sexuada cria uma oportunidade de produzir novas combinações genéticas e fixá-las em uma progênie heterozigota (BASHAW, HOVIN e HOLT, 1970). Para possibilitar a reprodução entre genótipos sexuais e apomíticos, as plantas sexuais encontradas na natureza são tratadas com colchicina, assim duplicando seu conjunto cromossômico (JANK et al., 2008). Esta estratégia de cruzamento entre plantas sexuais e apomíticas é de grande importância para os programas de melhoramento (HANNA, 1995; SAVIDAN, 1982). Neste contexto, a apomixia ganhou importância não só na história evolutiva das plantas como no processo de melhoramento genético de cultivares, uma vez que a manipulação deste processo reprodutivo permite fixar uma característica de interesse (DALL'AGNOL e SCHIFINO-WITTMANN, 2005).

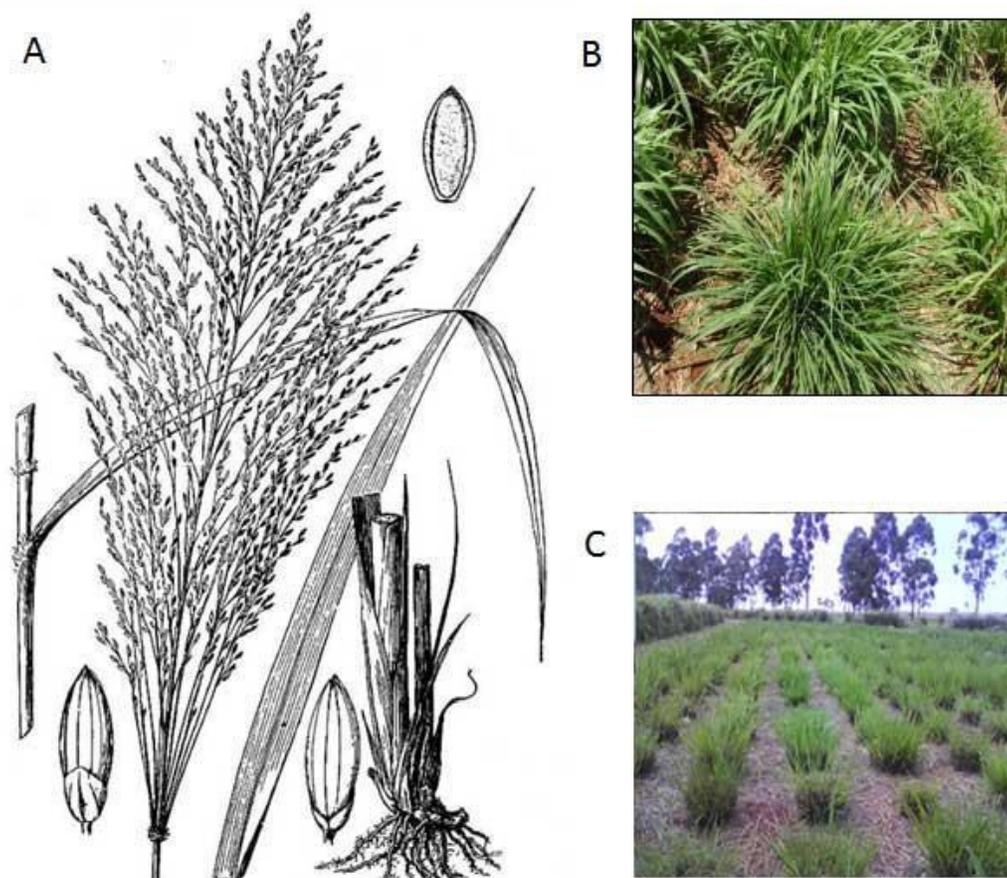


Figura 1: Ilustração e fotos de *Panicum maximum*. A) Ilustração de *P. maximum* (extraída de HITCHCOCK, 1971). B) Parte aérea da planta C) Híbridos de população de mapeamento, localizada na Embrapa Gado de Corte, Campo Grande – MS.

Atualmente, a Embrapa Gado de Corte dispõe de uma coleção de aproximadamente 426 genótipos no banco ativo de germoplasma (BAG) da espécie *P. maximum*. Através de um acordo de cooperação com o *Institut Français de Recherche Scientifique pour le Développement en Coopération* (ORSTOM), esta coleção foi introduzida no Brasil para o desenvolvimento do programa de melhoramento genético (JANK et al., 2008). Entre os genótipos do BAG, são encontrados acessos apomíticos e sexuais; e devido à eficiência e à abrangência das expedições de coleta realizadas no centro de origem, a coleção de genótipos presentes nos BAGs pode ser considerada como representativa da variabilidade natural existente para a espécie (SAVIDAN et al., 1989). Os genótipos pertencentes ao BAG da

Embrapa Gado de Corte foram caracterizados entre 1984 e 1988, demonstrando alta variabilidade agrônômica e morfológica (JANK, 1995; JANK et al., 1994, 1997). Posteriormente, os melhores genótipos foram avaliados sob pastejo, resultando no lançamento de três cultivares comerciais, *P. maximum* cv. Tanzânia-1 (1990), *P. maximum* cv. Mombaça (1993) e *P. maximum* cv. Massai (2001). Estas cultivares intensificaram os sistemas de produção, sendo que as cultivares Tanzânia-1 e Mombaça representam cerca de 10% das vendas totais de sementes de forrageiras (JANK, VALLE e RESENDE, 2011). O objetivo geral do programa de melhoramento genético de *P. maximum* é lançar genótipos produtivos, de boa qualidade e resistentes às cigarrinhas-das-pastagens, para suprir a demanda nacional por forrageiras visando à intensificação da produção. A geração de novas cultivares certamente contribuirá significativamente para a produção de carne, leite, couro e lã no país (JANK et al., 2008).

Contexto das forrageiras na pecuária nacional

No Brasil, a pecuária bovina de corte e de leite baseia-se, predominantemente, na utilização de pastagens para a alimentação animal, sendo que aproximadamente 90% dos nutrientes que os ruminantes precisam são obtidos diretamente do pasto (EUCLIDES et al., 2010). A criação de animais em pastos cultivados constitui-se em um dos fatores de competitividade e diferencial qualitativo para a carne brasileira, contribuindo de forma decisiva para a elevação do Brasil à condição de segundo maior produtor e primeiro exportador de carne do mundo (ALVES, 2008).

As maiores áreas de pastagens cultivadas no Brasil empregam as gramíneas dos gêneros *Brachiaria* e *Panicum*, devido à quantidade de sementes melhoradas comercializadas no país (SOUSA et al., 2011a). A utilização de pastagens com alto valor nutritivo é um dos fatores mais importantes para a redução de custos na produção agropecuária. As gramíneas fornecem

energia, proteína e fibras para promover a mastigação, ruminação e saúde do animal (LEOPOLDINO, 2000). As pastagens cultivadas cobrem extensas áreas, hoje estimadas em cerca de 101,4 milhões de hectares, e abrigam cerca de 176 milhões de cabeças de gado (IBGE, 2006), constituindo o maior rebanho comercial do mundo, que representa cerca de 14% do total mundial (FAOSTAT, 2011). As sementes de forrageiras também são economicamente importantes para o Brasil, pois movimentam um mercado de cerca de 240 milhões de dólares anuais, equivalente ao do mercado de sementes de milho (RESENDE et al., 2008).

Apesar da significativa importância, são poucas as cultivares disponíveis no Brasil resultantes de programas de melhoramento genético propriamente dito (JANK, 1995). A Embrapa Gado de Corte, através de seu programa de melhoramento genético, lançou as cultivares Mombaça, Tanzânia-1 e Massai, sendo que elas são as cultivares mais comercializadas da espécie no Brasil. Este dado mostra a importância do programa de melhoramento para a espécie (JANK, VALLE e RESENDE, 2011). Apesar da existência de cultivares muito utilizadas pelos criadores, o número de cultivares competitivas disponíveis para os produtores rurais ainda é pequeno (BRAZ, 2012).

Para que ocorra a intensificação da atividade pecuária, é necessário o desenvolvimento de cultivares de forrageiras com melhor desempenho e eficiência na utilização de insumos, adaptadas aos diversos ecossistemas pastoris do Brasil. Grande parte das áreas cultivadas por forrageiras encontra-se estabelecida com cultivares exóticas e de reprodução clonal, o que representa sério risco para todos os sistemas de produção baseados no uso de pastagens (VALLE et al., 2009). Este risco deve-se principalmente ao monocultivo de espécies forrageiras, que favorece a ocorrência de pragas e doenças. Nesse sentido, o melhoramento genético e o lançamento de novas cultivares surgem como possível solução para a diversificação das pastagens brasileiras (JANK, VALLE e RESENDE, 2011). A importância das forrageiras tropicais e de seu melhoramento continuado estão intimamente relacionados ao

potencial de aumentar a competitividade de toda a pecuária nacional (ALVES, 2008). Ainda, o Brasil apresenta cerca de metade das pastagens em variados estágios de degradação (DIAS-FILHO, 2007), o que poderia se beneficiar da substituição por novas cultivares provenientes de programas de melhoramento. Como resultado de anos de esforço em pesquisa e melhoramento das espécies forrageiras, assim como o manejo, foi obtido um aumento significativo na eficiência da utilização do pasto (EUCLIDES et al., 2010). No período entre 1995 e 2006, houve um aumento de 12% do rebanho brasileiro, enquanto a área coberta por pastagens diminuiu 8,3% (IBGE, 2006). Estes dados mostram que ainda existe espaço para expansão da indústria do gado, sem o aumento das áreas de pastagens (JANK, VALLE e RESENDE, 2011).

Gramíneas e produção de biomassa

Ao longo da história da civilização, a humanidade se aproveitou da energia das ligações químicas, queimando biomassa como combustível ou ingerindo plantas pelo conteúdo nutritivo dos açúcares e amido (McKENDRY, 2002). Biomassa é definida como a matéria orgânica formada por captura fotossintética de energia solar, armazenada sob forma de energia química (NALLATHAMBI GUNASEELAN, 1997). Quando as ligações entre moléculas de carbono, hidrogênio e oxigênio são quebradas por processos de digestão, combustão ou decomposição, esta energia química armazenada é então liberada. Se a biomassa é processada eficientemente, seja quimicamente ou biologicamente, a energia química liberada é combinada com oxigênio, oxidando o carbono para produzir CO₂ e água. Assim, o processo é cíclico, disponibilizando CO₂ para formação de mais biomassa (McKENDRY, 2002).

A biomassa fossilizada (carvão e petróleo) é altamente explorada, contudo, como o processo de conversão de biomassa em combustíveis fósseis leva milhões de anos, estes não podem ser considerados renováveis dentro de um período de tempo que a humanidade possa usar. Além disso, contribuem para emissões de carbono e conseqüentemente para as mudanças

climáticas (SOLOMON et al., 2009). Entretanto, na queima de biomassa vegetal o CO₂ residual lançado na atmosfera é reabsorvido no novo ciclo de crescimento das plantas, tornando o processo renovável (McKENDRY, 2002). Atualmente, os países buscam alternativas energéticas para se adequar ao panorama global, na tentativa de reduzir o consumo de combustíveis fósseis e seus problemas associados (SANTOS, 2011). Potenciais culturas energéticas incluem espécies lenhosas e gramíneas/herbáceas perenes, culturas de amido e açúcar, e sementes oleaginosas. As características de uma cultura energética ideal incluem alta produção, baixo requerimento energético para produção, baixo custo operacional, composição com o mínimo de contaminantes, e baixo requerimento de nutrientes, além da proporção entre celulose, hemicelulose e lignina (McKENDRY, 2002).

Nos Estados Unidos e Europa, muito investimento foi feito em pesquisas em cultivares energéticas modelo como *Panicum virgatum* e *Miscanthus giganteus*, resultando em benefícios diretos na geração de biomassa total (CLIFTON-BROWN, BREUER e JONES, 2007; McLAUGHLIN e ADAMS KSZOS, 2005). No hemisfério sul, gramíneas forrageiras tropicais são opções à obtenção de energia a partir da biomassa. O capim elefante (*Pennisetum purpureum* Schum.) é um dos mais produtivos, porém sua propagação vegetativa dificulta a implantação. Desta maneira, *P. maximum* é uma alternativa com alto potencial, por também ser de alta produtividade, pela propagação por sementes e bons valores associados a produção de biomassa (COSTA e JANK, 2009; JANK et al., 2010).

Análise de transcriptomas

O transcriptoma de uma espécie representa o conjunto completo e a quantidade de transcritos, que estão sendo expressos em determinado tecido, estágio de vida ou organismo como um todo (EKBLÖM e GALINDO, 2011). O transcriptoma oferece informação sobre expressão e regulação gênica, e o conteúdo de aminoácidos das proteínas traduzidas. Ainda, a

análise de transcriptoma é essencial para interpretação dos elementos funcionais do genoma e dos constituintes moleculares de células e tecidos (MARTIN e WANG, 2011; WANG, GERSTEIN e SNYDER, 2009). Entre os principais objetivos da transcriptômica podemos destacar a descrição de todas as espécies de transcritos, como RNAs mensageiros (mRNA), RNAs não codificantes (ncRNA) e pequenos RNAs (sRNA), a determinação estrutural destes transcritos, definindo suas extremidades 5' e 3', padrões de *splicing* e outras modificações pós-transcricionais, e a quantificação dos níveis de expressão de cada transcrito, nas diversas etapas de desenvolvimento ou sob diferentes condições (WANG, GERSTEIN e SNYDER, 2009).

Por muito tempo, o conhecimento sobre transcriptomas foi derivado quase que exclusivamente através de predições de genes e evidências limitadas a partir de ESTs (*expressed sequence tags*), que por muitas vezes representavam informações parciais e viciadas (MARTIN e WANG, 2011). Contudo, ao longo dos últimos anos, o tradicional método de sequenciamento Sanger (SANGER e NICKLEN, 1977) foi parcialmente substituído por tecnologias de sequenciamento de nova geração (*next-generation sequencing* ou NGS). Esta nova era tecnológica proporciona uma oportunidade para laboratórios e institutos de pesquisa se engajarem em projetos ambiciosos que antes só seriam concebíveis em grandes grupos ou centros de pesquisa (MARGUERAT e BÄHLER, 2010). Os recentes avanços nas metodologias de sequenciamento permitiram o desenvolvimento do sequenciamento massivo paralelo de cDNA ou RNA-seq, uma técnica com boa relação custo-benefício, direcionada a obter enormes quantidades de dados relativos ao transcriptoma de qualquer organismo ou tecido (TRAPNELL et al., 2010). A princípio, os dados provenientes de RNA-seq nos permitem identificar praticamente todos transcritos expressos (GUTTMAN et al., 2010). Desta maneira, a metodologia RNA-seq está se tornando altamente atrativa para estudos em espécies não modelo, onde geralmente existe pouco ou nenhum recurso genético disponível (HORNETT e WHEAT, 2012). Nestes casos, muitos trabalhos que caracterizam

transcriptomas tem sido descritivos, proporcionando recursos valiosos para estudos posteriores. Após o transcriptoma ter sido anotado, através de diferentes bancos de dados, o mesmo pode ser utilizado como ponto de partida para caracterizações funcionais detalhadas, como por exemplo, a categorização funcional através do banco de dados Gene Ontology (HARRIS et al., 2004), ou a identificação de vias metabólicas através do *Kyoto Encyclopedia of Genes and Genomes* ou KEGG (KANEHISA e GOTO, 2000). Ainda, um estudo baseado na técnica de RNA-seq permite a descoberta de novas informações sobre *splicing* alternativo entre os transcritos, regulação cis vs. trans, dinâmicas envolvendo micro RNAs (miRNAs) e identificação de marcadores moleculares como os microssatélites (também chamados de *simple sequence repeats* - SSR). E assim como as aplicações genômicas de NGS, a análise por RNA-seq oferece resolução de base simples, no qual é possível detectar diferenças de uma única base para subsequente desenvolvimento de marcadores *single nucleotide polymorphisms* (SNP) (EKBLÖM e GALINDO, 2011).

Se por um lado o surgimento das plataformas de NGS possibilitou aumentos expressivos na geração de sequências, proporcionando uma favorável relação custo-benefício quando comparados com o método Sanger, existe a perda em relação ao tamanho das sequências geradas (*reads*) em algumas das principais plataformas (MOROZOVA e MARRA, 2008), proporcionando novos desafios para as análises de bioinformática. O principal obstáculo é a reconstrução ou montagem dos *short reads* em sequências que representam os transcritos originais. Dependendo da existência ou não de um genoma que sirva de referência, as estratégias atuais de montagem de transcriptoma dividem-se em três categorias principais: montagem baseada em referência, montagem *de novo*, e uma estratégia híbrida combinando as duas últimas (MARTIN e WANG, 2011). A estratégia de montagem *de novo* não utiliza um genoma referência, ao invés disso, são identificadas sobreposições entre os *reads*, resultando no transcrito final (GRABHERR et al., 2011; MARTIN e WANG, 2011). Pelo menos 60

transcriptomas de plantas foram sequenciados *de novo* ao longo dos últimos anos (SCHLIESKY et al., 2012). Através de ferramentas de bioinformática apropriadas, as montagens *de novo* vêm tornando-se amplamente difundidas em estudos de organismos não modelos como as forrageiras tropicais.

A metodologia de RNA-seq também pode ser utilizada na caracterização de perfis de expressão gênica, sendo que o objetivo não é somente identificar quais genes são expressos, mas também o nível específico (absoluto ou relativo) da expressão dos genes encontrados (EKBLÖM e GALINDO, 2011). Até o surgimento das metodologias de NGS, a expressão de mRNA era mensurada principalmente por técnicas de *microarray* ou reação de polimerase em cadeia (*Polymerase Chain Reaction* ou PCR) em tempo-real. O primeiro método não possui grande sensibilidade, com limitações quanto às quantidades de RNA iniciais, quantificação dos níveis de expressão e a necessidade de informação prévia sobre as sequências, enquanto o segundo é de relativo alto custo e assim inviabiliza a análise global da expressão gênica (MARDIS, 2008). Em contraste, as metodologias rápidas e de relativo baixo custo de NGS oferecem a análise em larga escala de perfis de expressão gênica (MUTZ et al., 2013). Em síntese, o processo de análise de expressão gênica diferencial por RNA-seq inicia-se com os *reads* gerados pelo sequenciamento sendo alinhados a um genoma ou transcriptoma referência. Nos organismos que não possuem tais referências, uma montagem *de novo* pode resolver o problema oferecendo um conjunto inicial de transcritos (MARTIN e WANG, 2011). A quantidade de *reads* mapeados é contada e o nível de expressão gênica calculado por algoritmos especializados, e então testes estatísticos apropriados são aplicados, e a expressão gênica diferencial é determinada (Figura 2) (MUTZ et al., 2013).

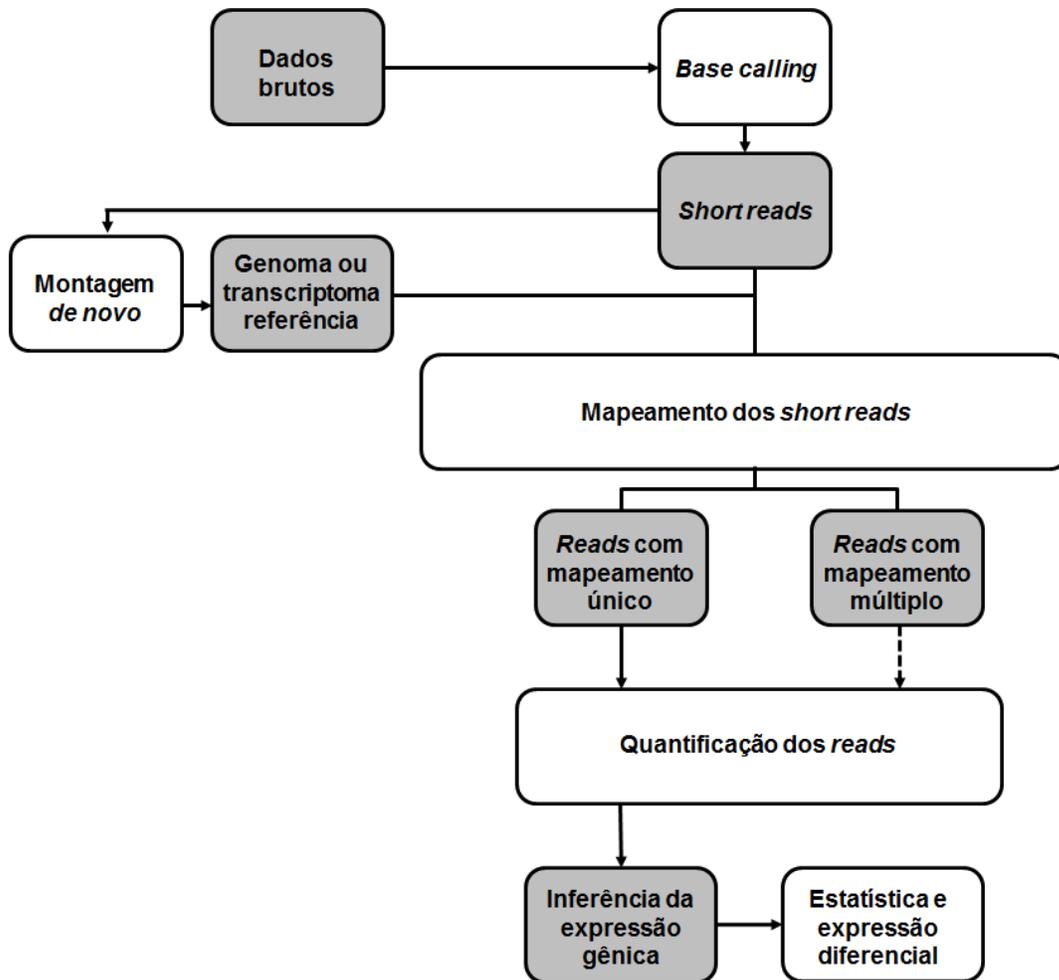


Figura 2: *Pipeline* usual de um experimento de RNA-seq, considerando as diversas etapas possíveis. Adaptado de MUTZ et al., 2013.

A disponibilidade de um banco de dados de seqüências para uma determinada espécie vegetal traz grandes avanços a pesquisa, como foi evidenciado pela história de sucesso de *Arabidopsis thaliana*. Contudo, a complexidade de muitos genomas de espécies vegetais e o alto custo freou o sequenciamento de seus genomas. Ao invés disso, foram sequenciados muitos transcriptomas de diferentes tecidos de diversas plantas cultivadas. Atualmente, o projeto 1KP (www.onekp.com) tem como objetivo o sequenciamento de transcriptomas de 1000 espécies vegetais. Com a informação obtida a partir de sequenciamentos de genomas e

transcriptomas, em união com estudos de mapeamento e associação, é possível identificar genes e *quantitative trait loci* (QTLs) chaves, redes metabólicas que podem levar a maiores níveis de produção, vias relacionadas com resistências a doenças, e auxiliar o cultivo em condições extremas e com menor impacto ambiental (ROUNSLEY et al., 2009).

Marcadores moleculares

Os marcadores moleculares são regiões existentes no genoma que podem atuar como sinalizadores na busca por regiões genômicas de interesse, atuando como variantes alélicas que permitem a identificação de indivíduos no nível molecular, representando assim, uma forma indireta de avaliar e comparar diferentes genótipos (LABORDA, 2011). Os diversos tipos de marcadores moleculares disponíveis diferenciam-se pela tecnologia utilizada, pela sua capacidade de detectar diferenças entre indivíduos, custo relativo, facilidade de uso, consistência e reprodutibilidade (SOUZA, 2012). A partir do surgimento da metodologia de amplificação de DNA via PCR, diversos marcadores moleculares foram desenvolvidos: AFLPs (*Amplified Fragment Length Polymorphisms*) (VOS et al., 1995), RAPDs (*Random Amplified Polymorphic DNA*) (WELSH e McCLELLAND, 1990), SSR (*Simple Sequence Repeats*) (LITT e LUTY, 1989), ISSR (*Inter Simple Sequence Repeat*) (ZIETKIEWICZ, RAFALSKI e LABUDA, 1994), SNP (*Single Nucleotide Polymorphism*) (SYVÄNEN, 2001), entre outros.

Marcadores do tipo microssatélites ou SSRs representam um importante avanço para estudos genéticos. São caracterizados como pequenas regiões compostas de 1 a 6 nucleotídeos repetidos em *tandem*, distribuídas aleatoriamente no genoma dos eucariotos (LITT e LUTY, 1989). Os marcadores SSR são codominantes e multialélicos, cuja variação advém da diversidade do número de motivos repetitivos, fornecendo um elevado nível de informações genéticas (FERREIRA e GRATTAPAGLIA, 1996). Além disso, a sua detecção é possível a partir de qualquer tipo de tecido em qualquer estágio de desenvolvimento. Apesar de apresentar

altas taxas mutacionais (10^{-3} a 10^{-6}) quando comparados com outros marcadores genéticos, os SSRs são conservados em suas regiões flangeadoras entre genótipos de uma mesma espécie e de espécies relacionadas. Consequentemente, eles podem ser transferidos dentro de espécies do mesmo gênero ou de genes similares filogeneticamente (GOLDSTEIN e SCHLOTTERER, 1999). A presença dessas características faz com que os microssatélites sejam marcadores indicados para diversos estudos, como mapeamento genético, identificação e discriminação de genótipos e estudos relacionados a genética de populações (CAIXETA et al., 2009).

Por outro lado, variações de uma única base em uma sequência de DNA, incluindo SNPs e pequenas inserções e deleções (InDels), são consideradas a mais abundante fonte de variação entre genomas, incluindo plantas (HENRY, 2008; KWOK, 2001). Diversos estudos sugerem que a frequência de SNPs em plantas varia de um SNP a cada 21 pb em batata até um SNP a cada 7.000 pb em tomate (LABATE e BALDO, 2005; NESBITT e TANKSLEY, 2002; RICKERT et al., 2003). Marcadores moleculares baseados em SNPs tornam possível o desenvolvimento de mapas genéticos de alta resolução, e também facilitam os estudos de clonagem gênica baseados em mapeamento, assim como análises de associação com base em haplótipos (BYERS et al., 2012). Quando a associação entre características fenotípicas e locos candidatos é estudada, as análises em SNPs considerando loco simples apresentam perda de informação devido a sua natureza bialélica, quando comparado com os marcadores SSRs multialélicos (VIGNAL et al., 2002). Baixa variabilidade significa menor número de meioses informativas, tornando difícil a detecção de ligação entre marcadores. Para contornar este problema, os SNPs são utilizados em grande número, sendo frequentemente espaçados de maneira uniforme e com uma alta cobertura do genoma. Ainda, é possível construir mapas combinando marcadores microssatélites, que apresentam baixa densidade, porém alta variabilidade, com marcadores SNPs, com alta densidade e baixa variabilidade. A ideia dessa estratégia é que os marcadores SSRs atuem como âncoras e façam com que os marcadores

SNPs não ligados sejam assimilados aos grupos de ligação. Essa combinação de diferentes marcadores produz mapas com maior acurácia e resolução, favorecendo a identificação de QTLs (BALL et al., 2010). Estudos em *Arabidopsis* utilizando SNPs obtiveram mapas de resolução 100 vezes superior àquela apresentada por mapas existentes. Isso deve-se ao fato da densidade de marcadores SNPs ser superior a de marcadores comumente utilizados (CAIXETA et al., 2009). É possível a identificação de SNPs através de sequenciamento direto do DNA ou cDNA, de informações disponíveis em bancos de dados, e através de métodos bioquímicos que investigam a presença de variantes. Com o advento de novas tecnologias de sequenciamento (NGS) e RNA-seq, a descoberta de grande número de SNPs teve seu custo reduzido quando comparado com abordagens tradicionais (BYERS et al., 2012). Após a identificação de SNPs é necessário validá-los, ou seja, determinar a frequência do SNP em um grupo de indivíduos, certificando-se de sua natureza polimórfica.

Através do uso de ferramentas como os marcadores moleculares, que apresentam um grande potencial na melhoria da eficiência dos programas de melhoramento, o tempo de lançamento de uma cultivar pode ser diminuído, graças à rápida identificação de características a serem selecionadas em uma geração. Além disso, permitem a identificação de híbridos e relações de parentesco, a caracterização molecular de cultivares, identificação e quantificação de fluxo gênico e também mensuração da diversidade genética. Informações obtidas a partir destes marcadores têm contribuído para um melhor entendimento da evolução e genética tanto de espécies diplóides quanto poliplóides (SOLTIS, SOLTIS e RIESEBERG, 1993).

A pesquisa relacionada a *P. maximum* vem aproveitando-se das metodologias associadas a marcadores moleculares. EBINA et al. (2007) desenvolveram 13 marcadores microsatélites a partir de clones genômicos e de cDNA, e os aplicaram para estimar as relações genéticas de 77 acessos de *P. maximum*. O estudo revelou seis grandes grupos e sugeriu que o centro de origem da espécie seja no Quênia, corroborando com estudos

anteriores baseados em análises morfológicas. CHANDRA e TIWARI (2010) isolaram e caracterizaram 15 marcadores SSR em 34 acessos *P. maximum* e a transferência dos mesmos para outras cinco espécies do gênero *Panicum*. Em nosso grupo, SOUSA et al. (2011a) desenvolveram 20 novos marcadores polimórficos do tipo microssatélites, sendo validados e utilizados para estimar as relações genéticas de 25 genótipos selecionados do banco de germoplasma. Ainda, esses marcadores tiveram sua transferibilidade testada para outras espécies forrageiras, obtendo sucesso na amplificação em quatro diferentes espécies, sendo estas *Brachiaria humidicola*, *B. brizantha*, *Paspalum regnelli* e *Paspalum notatum*. SOUSA et al. (2011b) avaliaram 396 genótipos do banco de germoplasma de *P. maximum* (apomíticos e sexuais) a partir de 30 marcadores SSR, e estes foram atribuídos a quatro grupos distintos, evidenciando uma ampla diversidade genética. Os genótipos sexuais foram agrupados entre os apomíticos autotetraplóides. Essa análise permitiu a caracterização e a discriminação das variedades comerciais brasileiras e estrangeiras, além da identificação de duplicatas. Ainda, conforme citado anteriormente, (EBINA et al., 2005) desenvolveram o único mapa de ligação disponível para a espécie, a partir de marcadores RAPD e AFLP, identificando marcadores co-segregando com a característica de apomixia.

Mapas de ligação

Mapas genéticos ou mapas de ligação são ferramentas fundamentais para os estudos genéticos, e têm como objetivo determinar a posição linear de genes ou marcadores em um cromossomo, fornecer informação sobre taxas de recombinação em todo genoma, assim como rearranjos gênicos intra e interespecie. Contudo, uma das aplicações mais importantes dos mapas genéticos é a identificação de locos associados à características quantitativas (BALL et al., 2010). Estudos genéticos em plantas sempre estiveram na vanguarda em se tratando de avaliações sobre características complexas. Muitos dos genes responsáveis por efeitos de QTL

em plantas foram clonados com sucesso, resultado da aplicação persistente de técnicas da genética clássica. Atualmente, mapas genéticos de diversos organismos estão com uma resolução cada vez maior, enquanto os custos de genotipagem diminuem. Por estas razões, aumentam os casos de sucesso em estudos que objetivam identificar os genes (ou QTLs) responsáveis por efeitos fenotípicos em particular, através de um processo que se inicia na construção de mapas de ligação em populações derivadas de cruzamentos entre genitores divergentes (NORDBORG e WEIGEL, 2008). Desde sua concepção, os mapas genéticos foram desenvolvidos a partir de diversos tipos de marcadores moleculares. Nas últimas quatro décadas, os avanços na tecnologia relacionada à biologia molecular disponibilizaram diferentes tipos de marcadores genéticos, permitindo que mapas de ligação fossem construídos para um grande número de espécies, incluindo aquelas consideradas não-modelo (SLATE, 2005).

De forma geral, a construção de mapas de ligação é feita em duas etapas: (a) agrupamento dos marcadores moleculares nos chamados grupos de ligação; (b) estimativa da ordem e da distância entre os marcadores dentro destes grupos. É necessário portanto, que indivíduos de uma população segregante sejam genotipados (MOLLINARI, 2012).

A teoria e metodologias necessárias para construção de mapas genéticos em espécies diploides são bem estabelecidas, o que não pode ser considerado em espécies autopoliploides (LEACH et al., 2010; RIPOL et al., 1999). Grande parte do desafio inerente a construção de mapas de ligação em espécies poliploides reside nos complexos padrões de segregação da progênie resultante de cruzamento entre poliploides e a alta diversidade de genótipos que precisam ser resolvidos (GAR et al., 2011). Aproximadamente 75% das espécies vegetais são poliploides, com evidência de poliploidia atual ou ancestral em espécies de interesse para geneticistas de plantas (HENRY, 2008). Espécies economicamente importantes como batata, café, cana-de-açúcar entre outras são poliploides. Apesar de sua importância, as pesquisas com marcadores moleculares nestas espécies ainda não utilizaram todo o seu potencial. A

genotipagem precisa de espécies poliploides é necessária para o entendimento de mecanismos genéticos, e também de loci específicos que determinam fenótipos através de mapeamento de QTLs e estudos de associação.

Para contornar os problemas de ploidia em genomas vegetais complexos, a maioria dos trabalhos utilizam marcadores moleculares em dose única (MDU) (WU et al., 1992). Tais marcadores detectam a presença do polimorfismo em apenas um dos homólogos dos grupos de homologia no qual ele se encontra, e podem apresentar padrões de segregação 1:1 (dose única em apenas um genitor) e 3:1 (dose única em ambos os genitores). Nesse caso, marcadores RFLP ou SSRs funcionam como dominantes (SILVA, DA e SORRELLS, 1996) e outros padrões de dosagem são de difícil detecção. Apenas com marcadores em dose única é impossível estudar os efeitos de dosagem alélica, como demonstrado em milho (GUO, DAVIS e BIRCHLER, 1996) e *Arabidopsis* (WANG et al., 2006). Em poliploides menos complexos como o caso de tetraploides, alguns métodos foram desenvolvidos para mapeamento de QTLs baseados em mapas genéticos construídos com marcadores codominantes, como os SSRs (HACKETT, BRADSHAW e McNICOL, 2001; LEACH et al., 2010). Essas metodologias permitem que os modelos genético-estatísticos contemplem efeitos gênicos de alelos e interações gênicas de altas ordens (HACKETT, BRADSHAW e McNICOL, 2001).

OBJETIVOS

Principal

Contribuir para o conhecimento genético-molecular de *Panicum maximum*, visando auxiliar os programas de melhoramento genético.

Específicos

- Realizar a montagem *de novo* de *reads* provenientes do transcriptoma de folhas de quatro genótipos de *P. maximum* através da técnica de RNA-seq;
- Relacionar transcritos obtidos no transcriptoma de *P. maximum* à vias metabólicas de interesse como metabolismo C₄ e síntese de lignocelulose;
- Identificar marcadores moleculares putativos SSRs e SNPs no transcriptoma de *P. maximum*;
- Desenvolver novos marcadores microssatélites para *P. maximum*;
- Genotipar uma população F1-segregante, resultante de cruzamento intraespecífico de dois acessos contrastantes quanto a características de interesse, utilizando marcadores SSRs, incluindo marcadores publicados e desenvolvidos neste trabalho;
- Construir mapas de ligação para cada genitor da população de mapeamento, utilizando metodologia apropriada para espécies autotetraploides.

***De novo transcriptome assembly for the tropical
grass *Panicum maximum* Jacq.***

De Novo Transcriptome Assembly for the Tropical Grass *Panicum maximum* Jacq

Guilherme Toledo-Silva¹, Claudio Benicio Cardoso-Silva¹, Liana Jank², Anete Pereira Souza^{1,3*}

1 Molecular Biology Center and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, São Paulo, Brazil, **2** Embrapa Beef Cattle, Campo Grande, Mato Grosso do Sul, Brazil, **3** Department of Plant Biology, Biology Institute, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil

Abstract

Guinea grass (*Panicum maximum* Jacq.) is a tropical African grass often used to feed beef cattle, which is an important economic activity in Brazil. Brazil is the leader in global meat exportation because of its exclusively pasture-raised bovine herds. Guinea grass also has potential uses in bioenergy production due to its elevated biomass generation through the C₄ photosynthesis pathway. We generated approximately 13 Gb of data from Illumina sequencing of *P. maximum* leaves. Four different genotypes were sequenced, and the combined reads were assembled *de novo* into 38,192 unigenes and annotated; approximately 63% of the unigenes had homology to other proteins in the NCBI non-redundant protein database. Functional classification through COG (Clusters of Orthologous Groups), GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) analyses showed that the unigenes from Guinea grass leaves are involved in a wide range of biological processes and metabolic pathways, including C₄ photosynthesis and lignocellulose generation, which are important for cattle grazing and bioenergy production. The most abundant transcripts were involved in carbon fixation, photosynthesis, RNA translation and heavy metal cellular homeostasis. Finally, we identified a number of potential molecular markers, including 5,035 microsatellites (SSRs) and 346,456 single nucleotide polymorphisms (SNPs). To the best of our knowledge, this is the first study to characterize the complete leaf transcriptome of *P. maximum* using high-throughput sequencing. The biological information provided here will aid in gene expression studies and marker-assisted selection-based breeding research in tropical grasses.

Citation: Toledo-Silva G, Cardoso-Silva CB, Jank L, Souza AP (2013) De Novo Transcriptome Assembly for the Tropical Grass *Panicum maximum* Jacq. PLoS ONE 8(7): e70781. doi:10.1371/journal.pone.0070781

Editor: Zhanjiang Liu, Auburn University, United States of America

Received: April 7, 2013; **Accepted:** June 21, 2013; **Published:** July 29, 2013

Copyright: © 2013 Toledo-Silva et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors gratefully acknowledge the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) for financial support grant 2005/51010-0 and scholarships to GTS and CBCS, and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for research fellowships to APS and LJ, and Embrapa Beef Cattle and Association for Promotion of Breeding Research in Tropical Forages (Unipasto) for financial support for field activities. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: anete@unicamp.br

Background

Approximately half of the world's bovine meat is produced in tropical or subtropical areas [1]. Pastures serve as the basis for beef production in Brazil and occupy an area of 101.4 million hectares [2], which, when considered together with natural pastures, is similar to the area occupied by crops and forests [3]. Additionally, Brazilian commercial beef cattle herds are the largest in the world, and Brazil is the leader in global bovine meat exports [1]. Bovine meat production in Brazil is performed exclusively on pasture, according to international market concerns regarding food security [3]. The primary grasses used as forage for cattle belong to species of *Brachiaria* or *Panicum maximum* [4]. Guinea grass is a forage grass that is native to Eastern and Southern Africa and is found in tropical and subtropical regions [5]. African forage grasses evolved in the presence of large mammals, so they are able to tolerate intense grazing and are very productive, vigorous and robust [3]. *Panicum maximum* is important for pasture and for green silage and hay production in the tropical Americas because of its high yield and nutritional content. Guinea grass belongs to the family Poaceae, subfamily Panicoideae and tribe Paniceae and constitutes an agamic complex with *P. infestum* Anders and *P. trichocladium* K. Schum [5]. *Panicum maximum* is adaptable to diverse ecosystems

and is grown in several countries [6]. It is the most productive seed-propagated tropical forage grass and the second most cultivated forage grass in Brazil [7]. However, most of the land is cultivated by just a few clonal genotypes, which presents a considerable risk for livestock pasture-based systems. The development of new forage grass cultivars with adaptability to the various edaphic and climatic conditions found in Brazil to enhance the diversity of forage grasses is a top priority of breeding programs [3]. *Panicum maximum* is a tetraploid species ($2n = 4 \times = 32$) of autopolyploid origin [8] that undergoes gametophytic aposporous apomictic reproduction [9]. Plants produced by apomixis are genetically identical to the mother plant as a result of clonal propagation by seeds [3]. However, a few genotypes found in natural populations exhibit diploidy ($2n = 2 \times = 16$) and a sexual reproductive mode [6]. Sexual *P. maximum* are important in breeding programs because they represent an effective tool for breeding and promoting diversification. Tetraploidy can be artificially induced in diploid plants to permit hybridization with apomictic tetraploid genotypes. Apomixis in *P. maximum* is determined by a gene or group of genes through simple inheritance, and the progeny of a cross between sexual and apomictic plants will exhibit a 1:1 ratio of sexual and apomictic plants [6]. The inheritance of apomixis allows the fixation of

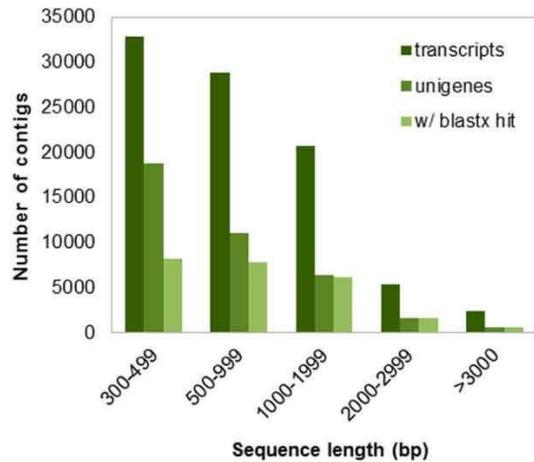


Figure 1. De novo assembly length distribution. Histogram of the sequence-length distribution of transcripts, unigenes and unigenes with significant BLASTx hits in the NCBI nr database. doi:10.1371/journal.pone.0070781.g001

desirable F₁ hybrids through cloning to generate heterosis and advantageous heterogeneous gene combinations, with subsequent seed production and commercialization [9].

Tropical forage grasses exhibit high growth rates and biomass yields as a consequence of the C₄ photosynthetic pathway [4]. C₄ photosynthesis involves several biochemical and anatomical adjustments to accumulate additional CO₂ compared with C₃ photosynthesis, using the key enzyme Rubisco. Grasses comprise the majority of C₄ plants (4,500 species), followed by sedges (1,500 species) and dicots (1,200 species). The importance of C₄ plants arises mainly from their global primary productivity; they account for approximately a quarter of total production [10]. In addition to their benefits for livestock, grasses such as *P. maximum* have potential as alternative energy sources through energy production via biomass generation, and this use is also considered in the species breeding objectives [3]. Species such as *Panicum virgatum* are being intensely studied for energy production [11–13]. Currently, the breeding of tropical forage grasses such as *P. maximum* and several *Brachiaria* species is based primarily on the assessment and selection of natural genotypes represented in germplasm banks, using the variability obtained from grasses of African origin [14,15]. The development of more productive and efficient cultivars can be improved through the use of genomics, transcriptomics and proteomics. The characterization of molecular markers is important for marker-assisted selection, germplasm assessment, the identification of hybrids and genome mapping [3]. Studies focusing on the molecular biology of *P. maximum*, including the characterization of molecular markers, genetic profiling, the search for apomixis-related genes and genetic evaluation of germplasm collections, have been performed [7,15–20]. However, knowledge about the Guinea grass transcriptome remains limited, and very few *P. maximum* protein and nucleotide sequences are available in current databases. The identification of transcribed regions of the genome using high-throughput sequencing (RNA-seq) provides a viable alternative for the analysis of non-model organisms with large genome sizes. RNA-seq methodology allows researchers to study the transcriptomes of non-model species and assess gene expression and patterns of regulation [21]. RNA-seq also allows the discovery of putative molecular markers, such as microsatellites (SSRs) and single nucleotide polymorphisms

(SNPs), because of the large quantity of data produced. Here, we present an overview of the transcriptome of *P. maximum* leaves. We constructed libraries from each of four genotypes currently used in the Guinea grass breeding program. We matched these sequences to known proteins in several databases using BLAST searches. Protein matches included a number of genes relevant to C₄ photosynthesis and lignocellulose biosynthesis. Furthermore, we screened the transcriptome for putative SSRs and SNPs, which will allow genome-wide screening of variation among different genotypes. The resulting assembled and annotated transcriptome sequences constitute a comprehensive genomic resource available for further studies and may enable the rapid identification of genes that are involved in pathways important for beef cattle grazing and potential biomass energy production.

Materials and Methods

Plant Material and RNA Extraction

Four different *P. maximum* genotypes were sampled from the Embrapa Beef Cattle collection (Campo Grande, MS, Brazil). The institution maintains a collection of Guinea grass germplasm and performs breeding research. The plants chosen for this study were the two sexual accessions that appeared the most promising after germplasm evaluation and two apomictic accessions that have been released as commercial cultivars [3]. The S10 and S12 genotypes were originally diploid with sexual reproduction and were tetraploidized using colchicine for crossing with tetraploid apomictic accessions. We used the tetraploid genotypes S10 and S12 for this study. The Tanzania and Mombaça genotypes are commercial cultivars of *P. maximum* that are autotetraploid with apomictic reproduction and desirable agronomic characteristics. Total RNA was isolated from the leaves using a modified lithium chloride protocol [22]. RNA quality was measured on a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA).

RNA-Seq Library Preparation and Sequencing

A transcript library was constructed using a TruSeq RNA sample preparation kit (Illumina Inc., San Diego, CA) according to the manufacturer's instructions. The quality of the library was

Table 1. Summary of assembled transcripts and unigenes of *P. maximum* leaves.

Total raw reads	168,053,718
Total data	13.44 Gb
Total clean reads	120,838,336
Q20 bases	96.32%
GC percentage	47.32%
Total transcripts	88.292
Transcripts mean length	942 bp
Transcripts N50	1272 bp
PE mapped reads (Bowtie)	85.96%
Unigenes	38.192
Unigenes mean length	758 bp
Unigenes N50	981 bp
Unigenes GC percentage	48.42%
Predicted ORFs (transcripts)	31.04%
Predicted ORFs (unigenes)	29.32%

doi:10.1371/journal.pone.0070781.t001

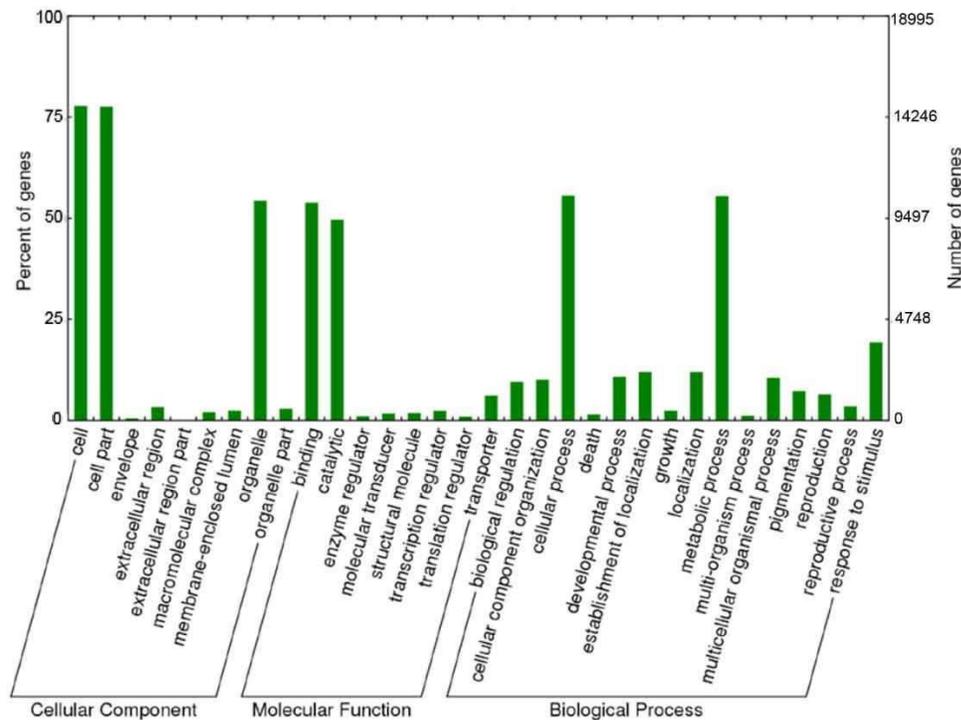


Figure 2. Gene ontology classification of *Panicum maximum* unigenes. Distribution of the GO categories assigned to the *P. maximum* transcriptome. Transcripts were classified into three categories: cellular components, molecular functions and biological processes. doi:10.1371/journal.pone.0070781.g002

assessed on a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA) and clustered using a TruSeq PE Cluster Kit on cBot (Illumina Inc., San Diego, CA). The samples were sequenced on an Illumina GAIIx using TruSeq SBS 36-Cycle kits (Illumina, San Diego, CA).

Raw Data Analysis and De Novo Transcriptome Assembly

High-quality reads (minimum 30 phred score) were filtered using a NGS QC Toolkit 2.3 [23]. Only paired-end reads (PE) from the four sequenced samples were assembled using Trinity software [24], which used three independent modules (Inchworm, Butterfly and Chrysalis) to assemble the transcriptome sequencing data *de novo*. Data from the four genotypes were assembled

together to obtain a reference transcriptome of *P. maximum* leaves. Short transcripts (<300 bp) were discarded from the resulting assembly. The Bowtie aligner [25] with default parameters was used to map the reads back onto transcripts. Our criteria for determining non-redundant contigs (unigenes) for transcriptome annotation considered the first Butterfly transcript generated per Chrysalis component as representative. Partial and complete open reading frames (ORFs) were predicted using the transdecoder script present in the Trinity package, with a minimum length of 200 amino acids. All reads were deposited in the NCBI Short Read Archive (SRA) under accession number SRR821833.

Annotation

Unigenes were used as queries to search protein databases using the BLAST+ program [26]. The queried databases included the NCBI non-redundant protein database (nr), UniProtKB-SwissProt, Clusters of Orthologous Groups (COG) and Phytozome grass data (www.phytozome.net); the Phytozome database includes protein sequences from *Sorghum bicolor*, *Zea mays*, *Setaria italica*, *Panicum virgatum*, *Oryza sativa* and *Brachypodium distachyon*. Homology searches applied BLASTx with an e-value cut-off of 1E-06 (in the case of the COG database, 1E-20 was used for increased stringency). Gene ontology (GO) terms were obtained from nr hits using Blast2GO software [27] with default parameters for the mapping and annotation steps, except that an e-value cutoff of 1E-10 was used for BLAST hits in the annotation step. Go-slim with plant slim (*Arabidopsis thaliana*) as an alias was used to summarize the GO term annotations of the transcriptome. WEGO [28] was used to functionally classify GO terms and graphically represent

Table 2. Annotation summary of 38,192 *P. maximum* unigenes.

Database	Hits	Hits percentage
NCBI non-redundant proteins (nr)	24,122	63.15%
SwissProt	16,396	42.93%
COG	22,473	58.84%
KEGG	4,110	10.76%
Grass	26,319	68.92%
Gene ontology	18,995	49.73%

doi:10.1371/journal.pone.0070781.t002

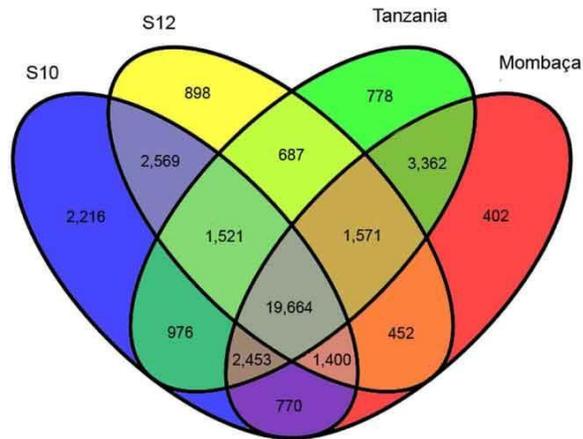


Figure 3. Shared and unique unigenes of *Panicum maximum* determined based on RSEM analysis.
doi:10.1371/journal.pone.0070781.g003

the distribution of unigene functions. The metabolic pathways were mapped using the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) [29] with a bi-directional best-hit strategy to assign KEGG orthology terms (KO) to unigenes. The identified pathways were settled using their respective KO assignments.

Abundance Estimation

RSEM (RNA-Seq by Expectation Maximization) software [30] was used to estimate the unigene FPKM values (fragments per kilobase of transcript per million mapped reads) based on read abundance using the Bowtie [25] aligner. Our discrimination of unique and shared transcripts among the sequenced plants was based on their FPKM values; only transcripts with FPKM values higher than 0.5 were considered.

Variant Detection

Microsatellites were searched using the MISA script [31], with the motif rules set to a six-copy minimum for dinucleotides and a four-copy minimum for trinucleotides to hexanucleotides, guaranteeing a minimum length of 12 bp. As shown in yeast [32], SSRs with lengths of less than 12 bp exhibit non-deterministic

variation and have a mutation potential similar to that of non-microsatellite regions. Burrows-Wheeler Aligner (BWA) [33] was used to align the reads back to the transcripts. Sequences from individual plants were mapped against unigenes. The default parameters were used to map the PE reads. The resulting alignment was analyzed using the Base Alignment Quality (BAQ) function of SAMtools [34], which provides an efficient and effective method for ruling out false SNPs caused by nearby insertions and/or deletions (INDELs). Subsequently, we used FreeBayes [35] for variant calling using the following parameters: ploidy 4, minimum read counting for variant calling 2, minimum base quality 30, minimum mapping quality 20, minimum coverage 20 and no INDELs or multiple nucleotide polymorphisms (MNP) called. Variants from sequenced genotypes were compared using the vcf-isec program from VCFtools [36], which generated a list of unique and shared SNPs.

Ethics Statement

We certify that no specific permits were required for the described field studies. This work was a collaborative research project developed by researchers from UNICAMP (Brazil) and EMBRAPA Beef Cattle (Brazil). Additionally, we confirm that the field studies did not involve endangered or protected species.

Results and Discussion

Sequencing and Assembling

We produced a total of 168,053,718 PE reads using Illumina technology, which generated 13.44 Gb of data (Table 1). No significant differences in sequencing data characteristics were observed among individuals (Table S1). After quality assessment and data filtering, 120,838,336 reads (96.32% Q20 bases and 47.32% GC content) were selected for *de novo* assembly. Using Trinity software [24], 88,292 transcripts were assembled, with a mean length of 942 bp and an N50 length of 1,272 bp. The Bowtie aligner [25] mapped 85.96% of the reads onto assembled sequences, considering only properly mapped paired-ends. We selected 38,192 sequences (43.25% of total transcripts) as unigenes, with a mean length of 758 bp and an N50 of 981 bp. The length distribution of the transcripts and unigenes is shown in Figure 1. We found that the average GC content of the *P. maximum* unigenes was 48.42%, slightly higher than the mean GC content of the raw reads. The average length of the unigenes was similar to that of chili pepper (712 bp) [37] and Ma bamboo (735 bp) unigenes [38] and higher than those of switchgrass (535 bp) [39] and safflower

Table 3. The 10 most abundant transcripts found in the Guinea grass leaf transcriptome.

Putative gene	E-value	FPKM	UniProtKB
Transcript antisense to ribosomal RNA protein 2	2.00E-016	10702.87	gij 74630365 sp Q8TGM7.1 ART2_YEAST
Carbonic anhydrase	1.00E-121	10021.35	gij 729003 sp P40880.1 CAHC_HORVU
Transcript antisense to ribosomal RNA protein 1	1.00E-011	7523.47	gij 74644329 sp Q8TGM6.1 TAR1_YEAST
Uncharacterized protein ycf76	2.00E-050	7410.46	gij 75121187 sp Q6ENQ6.1 YCF76_SACOF
Stem-specific protein TSJT1	1.00E-028	5228.23	gij 136452 sp P24805.1 TSJT1_TOBAC
Phosphoenolpyruvate carboxylase 1	0.0	5043.19	gij 115608 sp P04711.2 CAPP1_MAIZE
Chlorophyll a-b binding protein	4.00E-139	4389.51	gij 122246902 sp Q10HD0.1 CB23_ORYSJ
Pyruvate, phosphate dikinase 1	0.0	4164.02	gij 193806357 sp P11155.2 PPDK1_MAIZE
Metallothionein-like protein 3A	2.00E-008	3959.71	gij 158512839 sp A2_WLS0.1 MT3A_ORYSI
Metallothionein-like protein 1A	8.00E-009	3775.59	gij 158513336 sp A2ZH20.1 MT1A_ORYSI

doi:10.1371/journal.pone.0070781.t003

Table 4. List of genes comprising the C₄ photosynthetic pathway found among *P. maximum* unigenes.

Putative gene	KEGG orthology	Enzyme code	Unigenes
Carbonic anhydrase	K01673/K01674	4.2.1.1	8
Phosphoenolpyruvate carboxylase	K01595	4.1.1.31	12
Aspartate aminotransferase, cytosolic	K14454	2.6.1.1	1
Phosphoenolpyruvate carboxykinase	K01610	4.1.1.49	3
Pyruvate kinase	K00873	2.7.1.40	10
Alanine transaminase	K00814	2.6.1.2	1
Malate dehydrogenase (NADP+)	K00051	1.1.1.82	3
Malate dehydrogenase (oxaloacetate-decarboxylating)	K00029	1.1.1.40	4
Pyruvate, orthophosphate dikinase	K01006	2.7.9.1	1
Malate dehydrogenase	K00025	1.1.1.37	1
Malate dehydrogenase (decarboxylating)	K00028	1.1.1.39	2
Ribulose-bisphosphate carboxylase	K01602	4.1.1.39	1
Aspartate aminotransferase, chloroplastic	K00811	2.6.1.1	2

doi:10.1371/journal.pone.0070781.t004

(446 bp) unigenes [40]. The N50 value of the *P. maximum* unigenes was shorter than those of chili pepper (1,076 bp) and Ma bamboo (1,132 bp) and higher than that of safflower (555 bp). The *P. maximum* GC content was similar to that of rice (47.52%) and slightly higher than that of *A. thaliana* (41.10%) [41]. Direct comparisons of assembly metrics were challenging because the methods used for unigene definition and/or minimum contig settings have not been standardized. Nevertheless, these metrics showed that the current assembly was successful in obtaining useful leaf transcripts of *P. maximum*. The transdecoder script from the Trinity package was used to predict ORFs in the transcripts and unigenes. The total ORFs represented 31.04% of the assembled transcripts, whereas this value was slightly lower for unigenes (29.32%). A summary of the sequencing, assembly and ORF prediction process is presented in Table 1.

Annotation

The 38,192 assembled unigenes were queried against different protein databases, as shown in Table 2. We found 24,122 sequences (63.15%) similar to proteins in the nr database. The top-hit species distribution is presented in Figure S1. Among the nr BLASTx top hits, 10,851 were *Sorghum bicolor* proteins, followed by *Zea mays* (7,581), *Oryza sativa* (3,372), *Brachypodium distachyon* (1,020) and *Hordeum vulgare* (509). These five species accounted for ~96% of the total nr top hits, which was expected because these species are closely related to *P. maximum*. Additionally, as shown in Figure 1, there was a strong correlation between transcript length and annotation success; 53.59% of the sequences between 300–999 bp in length were successfully annotated, whereas 96.79% of the longer transcripts (>1,000 bp) retrieved hits above the e-value cutoff. A search for homology against the manually curated UniProtKB-SwissProt database produced 16,396 hits (42.93%)

Table 5. List of genes comprising the cellulose and lignin pathways found among *P. maximum* unigenes.

Putative gene	KEGG orthology	Enzyme code	Unigenes
Cellulose biosynthesis			
UTP–glucose-1-phosphate uridylyltransferase	K00963	2.7.7.9	2
Sucrose synthase	K00695	2.4.1.13	3
Cellulose synthase A	K10999	2.4.1.12	21
Sterol 3beta-glucosyltransferase	K05841	2.4.1.173	4
Lignin biosynthesis			
Phenylalanine ammonia-lyase	K10775	4.3.1.24	3
Trans-cinnamate 4-monooxygenase	K00487	1.14.13.11	4
4-Coumarate-CoA ligase	K01904	6.2.1.12	10
Cinnamoyl-CoA reductase	K09753	1.2.1.44	5
Cinnamyl alcohol dehydrogenase	K00083	1.1.1.195	6
Peroxidase	K00430	1.11.1.7	21
Shikimate O-hydroxycinnamoyltransferase	K13065	2.3.1.133	8
Caffeoyl-CoA O-methyltransferase	K00588	2.1.1.104	2
Ferulate-5-hydroxylase	K09755	1.14.-.-	2

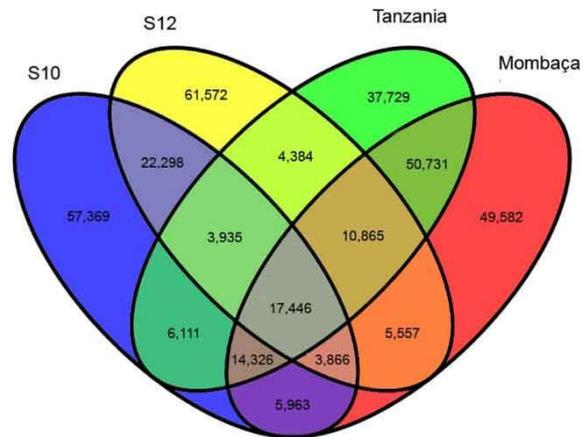
doi:10.1371/journal.pone.0070781.t005

Table 6. Summary of putative SSRs found in *P. maximum* unigenes.

Nucleotide repeat	Different motifs	Number of SSRs	% of total SSRs
Dinucleotide	4	451	8.9%
Trinucleotide	10	4333	86%
Tetranucleotide	30	184	3.6%
Pentanucleotide	29	44	0.8%
Hexanucleotide	23	23	0.4%
Total	96	5035	100%

doi:10.1371/journal.pone.0070781.t006

because of the smaller number of proteins in this more reliable protein bank. The Phytozome grass protein database comparison displayed a slightly higher number of hits (26,319, 68.91%) than the nr search, mostly because it contained *P. virgatum* and *S. italica* protein sequences that were not present in the nr database. The top hit species in searches of the grass database were *S. italica* (44.40%), *P. virgatum* (31.87%) and *S. bicolor* (10.11%). A homology search against the COG protein database returned 22,473 hits (58.84%). The COG database was used to define the orthologous functions of unigenes, as shown in Figure S2. The functional classification of the COG classes was inferred from a BLAST homology search against the COG protein database. COG ortholog classes were determined by comparing the protein sequences of complete genomes representing major phylogenetic lineages. Each COG class consisted of individual proteins or groups of paralogs from at least three lineages and thus corresponded to an ancient conserved domain [42]. Unigenes were classified into 25 functional categories and 8,850 COG terms. The top categories among the COG terms were general function prediction only (1,515) and DNA replication, recombination and repair (1,484). Within the general COG classification (Figure S3), unigenes with matched COG terms were distributed into information storage and processing (2,575; 29%), cellular processes (1,697; 19%), metabolism (2,503; 28%) and poorly characterized (2,075; 24%). The COG category classifications showed that the identified leaf transcripts were involved in a wide range of functions and used in a great number of processes that configure the tissue metabolism transcriptional machinery. The results of BLASTx searches against the nr protein database were imported into Blast2GO [27] for GO mapping and annotation. Based on the nr 20 top hits, the Blast2GO program obtained GO annotations for the unigenes, and WEGO software [28] was used to perform GO functional classification into the three major classes. Among the unigenes with nr hits, 18,995 (49.73%) were assigned to gene ontology classes with 100,440 functional terms. Biological processes comprised the majority of the functional terms (44,880; 45%), followed by cellular components (28,647; 28%) and molecular functions (26,913; 27%) (Figures 2 and S3). The distribution of GO functional classifications among unigenes was similar to the distribution in *P. virgatum* [43]. Within the biological processes category, cellular (10,556 unigenes) and metabolic (10,542 unigenes) processes were prominently represented. Additionally, response to stimulus (3,664 unigenes) was an abundant biological process term, and response to stress (2,658 unigenes) was the most highly represented child GO term. Of particular interest, 1,467 unigenes were annotated as related to reproductive processes. These unigenes may provide valuable information for further studies of reproduction in *P. maximum*. These sequences

**Figure 4.** Shared and unique putative SNPs found in *Panicum maximum* unigenes.

doi:10.1371/journal.pone.0070781.g004

provide new information about the genes involved in this process, which will aid future studies focusing on understanding Guinea grass reproduction and transferring *Panicum*-type apomixis and/or expressing this phenotype in grain crops. This type of apomixis can potentially be expressed in grain crops because it maintains the female to male genome ratio required for seed endosperm development [9]. In summary, 195 different terms were found among the unigenes, indicating that the genes expressed in Guinea grass leaves are involved in a wide variety of biological functions, as indicated by their COG classification. In the cellular component category, 230 terms were successfully mapped, with cell (14,765) and cell part (14,732) comprising the majority of this category. Among the molecular function categories, binding (10,235) and catalytic activity (9,435) were the most represented among the 33 terms found.

To correlate *P. maximum* leaf unigenes with known metabolic pathways, we used the KAAS server to assign sequences with KEGG orthology (KO) terms and their respective KEGG maps. A total of 4,110 (10.76%) assembled unigenes were associated with 2,665 KO terms and 297 pathways. Highly represented pathways included metabolic pathways (681 members) and secondary metabolite biosynthesis pathways (313 members). Similarly, pathways with the potential for further studies were detected

Table 7. Summary of putative SNPs found in *P. maximum* unigenes.

SNP type	Count
Transitions	228,468
A-G/G-A	113,075
C-T/T-C	115,393
Transversions	117,988
A-C/C-A	29,403
A-T/T-A	26,310
T-G/G-T	29,035
G-C/C-G	33,240
Total	346,456

doi:10.1371/journal.pone.0070781.t007

using KAAS annotation; these pathways included glycolysis/gluconeogenesis (31 members) (Figure S4), photosynthesis (28 members) (Figure S5), carbon fixation in photosynthetic organisms (23 members) (Figure S6), phenylpropanoid biosynthesis (15 members) (Figure S7) and many others. KAAS analysis showed that the assembled transcripts were distributed among several metabolic pathways, which provided the first overview of the Guinea grass transcriptome. Gathering all of the information obtained using the different annotation strategies, we provided initial information about the whole transcriptome of *P. maximum* leaves. Finally, 12,450 unigenes were not annotated in any of the databases compared in this study. Many of these 9,581 sequences were short contigs, of less than 500 bp. We searched for ORFs in this set of non-annotated unigenes and found 243 putative coding regions. Future analysis of this small dataset may reveal potential unknown genes in *P. maximum*.

Abundance Estimation

The assembled transcripts were mapped using Bowtie [25], and their respective read abundances (FPKM) were estimated by RSEM [30]. Reads from each genotype were mapped individually and in combination against the reference transcriptome. We used a FPKM cut-off value of 0.5 to define the unigenes expressed in different samples sequenced in this study. Based on these values, we determined which unigenes were unique or shared among the genotypes, as shown in Figure 3. The samples shared a total of 19,664 unigenes, and the S10 genotype displayed a higher number of unique transcripts (2,216). We also selected the ten most abundant transcripts for a brief description (Table 3). The mean FPKM value of the unigenes was approximately 21. The first and third most represented unigenes were ribosomal RNA proteins, members of the translational machinery required for general protein synthesis. Among the more abundant unigenes, four were associated with photosynthesis and carbon fixation in leaves; carbonic anhydrase was the second most represented unigene (10,021 FPKM). Carbonic anhydrase catalyzes the interconversion of CO₂ and HCO₃⁻ and is assumed to play an important role in photosynthesis [44]. Phosphoenolpyruvate carboxylase (PEPC) and pyruvate phosphate dikinase (PPDK) are also key enzymes in the C₄ photosynthetic pathway. Chlorophyll a-b binding proteins are components of light-harvesting complexes in plants that are crucial for the photosynthesis process [45]. We expected these carbon fixation enzymes to be abundant because the mRNA was extracted from the leaves. Two metallothionein-like proteins were also well represented. Metallothioneins (MTs) are cysteine-rich proteins that coordinate heavy metal atoms [46]. Plant MTs bind to metals with high affinity, and the respective genes are up-regulated in the presence of metal molecules. Because of these characteristics, plant MTs are believed to be involved in cellular metal homeostasis and tolerance mechanisms [47]. Furthermore, two undescribed proteins detected in other grasses were identified among the ten most abundant transcripts among unigenes: a stem-specific protein and a protein from the chloroplastic ycF76 family. Further studies are needed to determine the biological functions of these proteins.

Genes of Interest

To describe the genes of interest, we used sequences with associated KO terms derived from the KAAS [29] analysis. Most of the enzymes in the C₄ photosynthetic pathway (Table 4) and in cellulose and lignin production (Table 5) were identified. These pathways are important for efficient beef cattle grazing and the potential for biomass accumulation for energy generation. However, further studies, including specific molecular and

proteomic analysis procedures, are required to validate these predictions.

C₄ photosynthesis pathway. C₄ plants such as grasses express high levels of carbonic anhydrase and PEP carboxylase (PEPC) for initial CO₂ fixation in the cytoplasm [48]. These two enzymes showed high FPKM values (10,021 for CA and 5,043 for PEPC) and were the second and fifth most abundant transcripts, respectively, among the unigenes. The carbonic anhydrases form a family of enzymes that catalyze the interconversion of CO₂ and H₂O to bicarbonate (HCO₃⁻) and protons. PEPC catalyzes the addition of available bicarbonate to phosphoenolpyruvate (PEP) to form oxaloacetate (OAA) [49]. PEPC has high transcript abundance in *P. virgatum* [39]. OAA is either reduced to malate by NADP-malate dehydrogenase (NADP-MDH) or transaminated to aspartate by aspartate aminotransferase (AspAT). We were only able to find the cytosolic AspAT based on KAAS analysis. A similar distribution was observed in *P. virgatum*; the majority of the transcripts were cytosolic, and only one mitochondrial transcript was identified [39]. The resulting C₄ acid formed from NADP-MDH or AspAT is then decarboxylated to release CO₂ in the presence of Rubisco, which is the key enzyme in the C₃ photosynthetic pathway [49,50]. The FPKM value for Rubisco was also high (2,530). The decarboxylation reaction is catalyzed by one or more of the following three enzymes: NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME) and phosphoenolpyruvate carboxykinase (PEPCK). C₄ plants are classified into three subtypes depending on their major decarboxylation enzyme [49]. *P. maximum* is considered a PEPCK-type enzyme, and as expected, we found a higher (2,714) FPKM value for this enzyme than for the other two enzymes. The pyruvate generated by decarboxylation is used by pyruvate orthophosphate dikinase (PPDK) to recover the phosphoenolpyruvate (PEP) levels in chloroplasts. PPDK is considered one of the most abundant enzymes in nature [48] and was the eighth most abundant transcript here, with an FPKM value of 4,130.

Cellulose biosynthesis. Plant cell walls are complex structures composed of polysaccharides, proteins and lignins [50]. Cell wall constituents are of great importance for livestock grazing and various industrial applications such as biofuel production. The ability to release the carbohydrates from the cell wall for both applications is becoming more important [51]. Among the plant cell wall polysaccharides, cellulose is considered the main component and is a key substrate for livestock foraging and industrial applications [50]. Usually, biofuels are derived from starch-abundant species such as corn or sugarcane. Cellulose-rich plants such as Miscanthus and switchgrass are currently being investigated as alternatives [52]. We searched for homologues of cellulose biosynthesis enzymes using KO terms associated with the unigenes assembled in this work. Cellulose is synthesized by large multimeric cellulose synthase (CesA) complexes. Currently, the only known components of these complexes are the cellulose synthase proteins [50]. We found 21 transcripts with CesA-related KO terms. Among these transcripts, the highest FPKM value was 213. CesA complexes use UDP-glucose (UDP-Glc) as the activated sugar donor for β-1,4 glucan chain polymerization. UDP-Glc can be produced by sucrose synthase (SuSy) or UDP-glucose pyrophosphorylase (UGPase) [50]. We found three representative SuSy and two UGPase transcripts, with FPKM values of 51 and 189, respectively. Because we extracted RNA from the leaves, we expected that SuSy and UGPase would be expressed, although SuSy is mainly expressed in sink tissues [50]. Steryl glycosides may act as initiators of cellulose polymerization [53] and/or alter membrane conditions for cellulose synthesis [54]. Among the unigenes, four sequences were assigned KO terms belonging to

sterol beta-glucosyltransferase enzymes. Invertases have been proposed to provide carbon for cellulose production in non-photosynthetic cells [55], and they were not found among the *P. maximum* leaf unigenes described here.

Lignin biosynthesis. Lignin is one of the most abundant organic polymers on Earth, exceeded only by cellulose, and it constitutes approximately one-third of the non-fossil organic carbon and one-fourth to one-third of the dry mass of existing wood [56]. Lignin limits cell-wall digestibility by livestock, modulating the energy availability of forage crops in beef and dairy production [57], and it is also a limiting factor in the conversion of plant biomass to pulp or biofuels [51]. A reduction in lignin concentration would have positive effects on both grazing and the energetic use of grasses because the lignin and cellulose/hemicellulose within the lignocellulose constitute major obstacles to these uses [50]. Lignin is mainly synthesized from hydroxycinnamyl alcohols (or monolignols), coniferyl alcohol, sinapyl alcohol and p-coumaryl alcohol [51]. The monolignols, i.e., guaiacyl (G), syringyl (S) and p-hydroxyphenyl (H) units, are then incorporated into the lignin polymer. H-units are slightly more prevalent in grasses [56]. Among the unigenes of Guinea grass, we associated KO terms with all enzymes involved in monolignol formation (Table 5), with the exception of p-coumarate 3-hydroxylase (C3H) and caffeic acid O-methyltransferase (COMT). After the monolignols are transported to the cell wall, lignin is formed through the dehydrogenation of these molecules. The dehydrogenation reaction has been associated with peroxidases and laccases [56]. Among the unigenes, 18 sequences were associated with peroxidase function, but no laccases were found. The abundances of most enzymes in the lignin pathway found in this study were similar (mean of ~87 FPKM). However, phenylalanine ammonia-lyase (PAL) was more abundant, at 544 FPKM. This enzyme is up-regulated by high levels of phenylalanine in cells of *Pinus taeda* [58].

Putative Marker Discovery

Transcriptome sequencing provides valuable resources for the development of molecular markers, mostly because of the high quantity of the generated data, in which different types of polymorphisms (e.g., SNPs and SSRs) can be observed. These markers can be tested for potential utility as molecular markers for population genetics, linkage mapping and comparative genomics studies [59]. Here, we investigated two types of putative markers from *P. maximum* leaves: microsatellites and single nucleotide polymorphisms. Both types of markers need future validation for practical use in *P. maximum* breeding and research.

SSR discovery. Assembled Guinea grass contigs were analyzed to identify SSRs. The distributions of tandem repeats are shown in Table 6. Among the 38,192 unigenes, MISA [31] found 4,270 sequences (11.1%) containing 5,035 SSRs using a minimum length criterion of 12 bp, or approximately 1 SSR per 6 kb. This SSR value is slightly higher than that in switchgrass (8.8%) [39], similar to that in *Dendrocalamus latiflorus* (12.8%) [38] and higher than those in expressed sequences in other grasses (1.5–4.7%) [60] or plants in general (1–5%) [61]. A survey of bacterial artificial chromosome libraries of *P. virgatum* revealed one SSR per 5.2-kb [43]. The repeat size distribution analysis indicated that trinucleotides represented 86% of the microsatellite motifs found in *P. maximum* unigenes. This SSR value was higher than those of *P. virgatum* (48%–55%) [39,43], Ma bamboo (56%) [38] and barley (56%) [31] and similar to that of the red algae *Pyropia haitanensis* (87.1%) [62]. The CCG/CGG motif was the most common trinucleotide repeat (28.8% of trinucleotide SSRs). This particular motif is also the most abundant trinucleotide repeat in *P. virgatum*

[39,43], *P. haitanensis* [62] and *H. vulgare* [31]. Regarding the dinucleotide repeats (8.9% of total SSRs), the AG/CT motif represented 54.3% of this class. Tetra-, penta- and hexanucleotide repeats were less abundant (Table 6).

SNP markers. For SNP calling, BWA [33] was used to map the reads of each sample to the reference transcriptome. Freebayes [35] detected a total of 346,456 putative SNP positions in the unigenes using previously detailed parameters, as shown in Table 7. This value corresponds to approximately one SNP per 90 bp of unique transcript sequence (1/87), similar to the ratio in rice [63]. Transitions were approximately 1.93 times more abundant than transversions (Table 7). Transitions are generated by oxidative deamination and tautomerization. Although there are twice as many possible transversions than transitions, transitions are usually more common in genomes, possibly because of the molecular mechanisms from which they originate [64]. Similarly, transitions are more easily tolerated in natural selection because they are more likely to generate synonymous mutations in coding sequences than are transversions [65]. Among the detected variants, 231,589 (69.09%) were in predicted open reading frames. Based on the RSEM distinction of shared and unique transcripts among the four *P. maximum* genotypes, we determined the intersection of the putative SNPs found using VCFtools [36], which is shown in Figure 4. Considering the unique SNP positions, we found 49,582 variants for Mombaça, 57,369 for S10, 61,572 for S12 and 37,729 for Tanzania. These unique putative markers will be important for Guinea grass breeding, as the sampled genotypes are distinct for several economically important agronomic characteristics. We identified 17,446 SNPs that were present in all genotypes (Figure 4). Additionally, we determined the numbers of SNPs in the key pathways described in this work. The C₄ pathway contained 1,159 SNPs, whereas cellulose biosynthesis accounted for 491 SNPs, and lignin biosynthesis accounted for 924 SNPs. Prior to this study, there were no SNPs available for *P. maximum*; these markers may thus represent a useful new tool for marker-assisted selection in Guinea grass breeding.

Supporting Information

Figure S1 Top-hit species distribution among NCBI nr BLASTx hits.

(TIFF)

Figure S2 Functional classification of *Panicum maximum* unigenes based on COG classes.

(TIFF)

Figure S3 General COG and GO classification of *Panicum maximum* unigenes. (A) General COG classification (B) General GO classification.

(TIFF)

Figure S4 Glycolysis and gluconeogenesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

(TIFF)

Figure S5 Photosynthesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

(TIFF)

Figure S6 Carbon fixation KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

(TIFF)

Figure S7 Phenylpropanoid biosynthesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

(TIF)

Table S1 Statistics of sequenced data for each *P. maximum* genotype.

(DOC)

References

- FAOSTAT (2011) Live Animals data. Available: <http://faostat.fao.org/site/573/DesktopDefault.aspx?PageID=573#ancor>. Accessed 2013 Feb 4.
- IBGE (2011) Pesquisa pecuária municipal. Available: <http://www.sidra.ibge.gov.br/bda/tabela/listabl.asp?c=73&z=p&o=23>. Accessed 2013 Feb 8.
- Jank L, Valle CB, Resende RMS (2011) Breeding tropical forages. *Crop Breeding and Applied Biotechnology* 11: 27–34.
- Batistoti C, Lempp B, Jank L, Morais MDG, Cubas AC, et al. (2012) Correlations among anatomical, morphological, chemical and agronomic characteristics of leaf blades in *Panicum maximum* genotypes. *Animal Feed Science and Technology* 171: 173–180.
- Muir JP, Jank L (2004) Guineagrass. In: Moser LE, Burson BL, Sollenberger LE, editors. Warm-season (C_4) grasses. Madison: American Society of Agronomy, 589–621.
- Savidan Y (1983) Genetics and utilization of apomixis for the improvement of guineagrass (*Panicum maximum* Jacq.). In: Smith JA, Hayes VW, editors. Proceedings of the 14th International Grasslands Congress. Lexington: Westview Press, 182–184.
- Sousa ACB, Jungmann L, Campos T, Sforça DA, Boaventura LR, et al. (2011) Development of microsatellite markers in Guineagrass (*Panicum maximum* Jacq.) and their transferability to other tropical forage grass species. *Plant Breeding* 130: 104–108.
- Bogdan AV (1977) *Panicum maximum*. In: Bogdan AV, editor. Tropical pasture and fodder plants. London: Longmead, 181–191.
- Savidan Y (2000) Apomixis: genetics and breeding. In: Janick J, editor. Plant breeding reviews. Oxford: John Wiley & Sons, Vol. 18, 13–86.
- Sage RF (2004) The evolution of C_4 photosynthesis. *New Phytologist* 161: 341–370.
- Okada M, Lanzatella C, Saha MC, Bouton J, Wu R, et al. (2010) Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. *Genetics* 185: 745–760.
- Bouton JH (2007) Molecular breeding of switchgrass for use as a biofuel crop. *Current Opinion in Genetics & Development* 17: 553–558.
- Sokhansanj S, Mani S, Turhollow A, Kumar A, Bransby D, et al. (2009) Large-scale production, harvest and logistics of switchgrass (*Panicum virgatum* L.) - current technology and envisioning a mature technology. *Biofuels, Bioproducts and Biorefining* 3: 124–141.
- Jank L, Valle CB, Resende RMS (2005) Grass and forage plant improvement in the tropics and sub-tropics. In: McCalloway DA, editor. Grassland: a global resource. Wageningen: Wageningen Academic Publishers, 69–81.
- Sousa ACB, Jank L, Campos T, Sforça DA, Zucchi MI, et al. (2011) Molecular diversity and genetic structure of guineagrass (*Panicum maximum* Jacq.), a tropical pasture Grass. *Tropical Plant Biology* 4: 185–202.
- Yamada-Akiyama H, Akiyama Y, Ebina M, Xu Q, Tsuruta S-I, et al. (2009) Analysis of expressed sequence tags in apomictic guineagrass (*Panicum maximum*). *Journal of plant physiology* 166: 750–761.
- Ebina M, Kouki K, Tsuruta S, Akashi R, Yamamoto T, et al. (2007) Genetic relationship estimation in guineagrass (*Panicum maximum* Jacq.) assessed on the basis of simple sequence repeat markers. *Grassland Science* 53: 155–164.
- Ebina M, Nakagawa H, Yamamoto T (2005) Co-segregation of AFLP and RAPD markers to apospory in Guineagrass (*Panicum maximum* Jacq.). *Grassland Science* 51: 71–78.
- Chandra A, Tiwari KK (2010) Isolation and characterization of microsatellite markers from guineagrass (*Panicum maximum*) for genetic diversity estimate and cross-species amplification. *Plant Breeding* 129: 120–124.
- Tiwari KK, Chandra A (2010) Use of degenerate primers in rapid generation of microsatellite markers in *Panicum maximum*. *Journal of Environmental Biology* 31: 965–968.
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10: 57–63.
- Kistner C, Matamoros M (2005) RNA isolation using phase extraction and LiCl precipitation. In: Márquez AJ, editor. *Lotus japonicus Handbook*. Dordrecht: Springer Netherlands, 123–124.
- Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7: e30619.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10: R25.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, et al. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21: 3674–3676.
- Ye J, Fang L, Zheng H, Zhang Y, Chen J, et al. (2006) WEGO: a web tool for plotting GO annotations. *Nucleic Acids Research* 34: W293–W297.
- Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* 35: W182–W185.
- Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323.
- Thiel T, Michalek W, Varshney RK, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411–422.
- Pupko T, Graur D (1999) Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. *Journal of Molecular Evolution* 48: 313–316.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv: 12073907*: 9.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
- Liu S, Li W, Wu Y, Chen C, Lei J (2013) *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. *PLoS One* 8: e48156.
- Liu M, Qiao G, Jiang J, Yang H, Xie L, et al. (2012) Transcriptome sequencing and *de novo* analysis for Ma bamboo (*Dendrocalamus latiflorus* Munro) using the Illumina platform. *PLoS One* 7: e46766.
- Wang Y, Zeng X, Iyer NJ, Bryant DW, Mockler TC, et al. (2012) Exploring the switchgrass transcriptome using second-generation sequencing technology. *PLoS One* 7: e34225.
- Luhin H, Xiao Y, Pei S, Wen T, Shangqin H (2012) The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7: e38653.
- Victoria FC, Da Maia LC, De Oliveira AC (2011) In silico comparative analysis of SSR markers in plants. *BMC Plant Biology* 11: 15.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4: 41.
- Sharma MK, Sharma R, Cao P, Jenkins J, Bartley LE, et al. (2012) A genome-wide survey of switchgrass genome structure and organization. *PLoS One* 7: e33892.
- Badger MR, Price GD (1994) The role of carbonic anhydrase in photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 45: 369–392.
- Li XP, Björkman O, Shih C, Grossman R, Rosenquist M, et al. (2000) A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* 403: 391–395.
- Mir G, Domènech J, Huguet G, Guo W-J, Goldsbrough P, et al. (2004) A plant type 2 metallothionein (MT) from cork tissue responds to oxidative stress. *Journal of Experimental Botany* 55: 2483–2493.
- Cobbett C, Goldsbrough P (2002) Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annual Review of Plant Biology* 53: 159–182.
- Kanai R, Edwards GE (1999) The biochemistry of C_4 photosynthesis. In: Sage RF, Monson RK, editors. C_4 Plant Biology. San Diego: Academic Press, Vol. 4, 49–87.
- Matsuoka M, Furbank RT, Fukayama H, Miyao M (2001) Molecular engineering of C_4 photosynthesis. *Annual Review of Plant Physiology and Plant Molecular Biology* 52: 297–314.
- Endler A, Persson S (2011) Cellulose synthases and synthesis in *Arabidopsis*. *Molecular Plant* 4: 199–211.
- Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin biosynthesis and structure. *Plant Physiology* 153: 895–905.
- Carroll A, Somerville C (2009) Cellulosic biofuels. *Annual Review of Plant Biology* 60: 165–182.
- Peng L, Kawagoe Y, Hogan P, Delmer D (2002) Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. *Science* 295: 147–150.
- Schrick K, Fujioka S, Takatsuto S, Stierhof Y, Stransky H, et al. (2004) A link between sterol biosynthesis, the cell wall, and cellulose in *Arabidopsis*. *The Plant Journal* 38: 227–243.
- Barratt DHP, Derbyshire P, Findlay K, Pike M, Wellner N, et al. (2009) Normal growth of *Arabidopsis* requires cytosolic invertase but not sucrose synthase. *Proceedings of the National Academy of Sciences of the United States of America* 106: 13124–13129.
- Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annual Review of Plant Biology* 54: 519–546.

Author Contributions

Conceived and designed the experiments: IJ APS. Performed the experiments: GTS. Analyzed the data: GTS CBCS. Wrote the paper: GTS. Coordinated the mapping field that selected the samples used: IJ.

57. Jung HG, Allen MS (1995) Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. *Journal of Animal Science* 73: 2774–2790.
58. Anterola AM, Jeon JH, Davin LB, Lewis NG (2002) Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. *The Journal of biological chemistry* 277: 18272–18280.
59. Parchman TL, Geist KS, Grahnen J a, Benkman CW, Buerkle CA (2010) Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* 11: 180.
60. Kantety R V, La Rota M, Matthews DE, Sorrells ME (2002) Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* 48: 501–510.
61. Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* 23: 48–55.
62. Xie C, Li B, Xu Y, Ji D, Chen C (2013) Characterization of the global transcriptome for *Pyropia haitanensis* (Bangiales, Rhodophyta) and development of cSSR markers. *BMC Genomics* 14: 107.
63. Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, et al. (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research* 9: 163–171.
64. Ebersberger I, Metzler D, Schwarz C, Pääbo S (2002) Genomewide comparison of DNA sequences between humans and chimpanzees. *The American Journal of Human Genetics* 70: 1490–1497.
65. Wakeley J (1996) The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends in Ecology & Evolution* 11: 158–162.

MATERIAL SUPPLEMENTAR

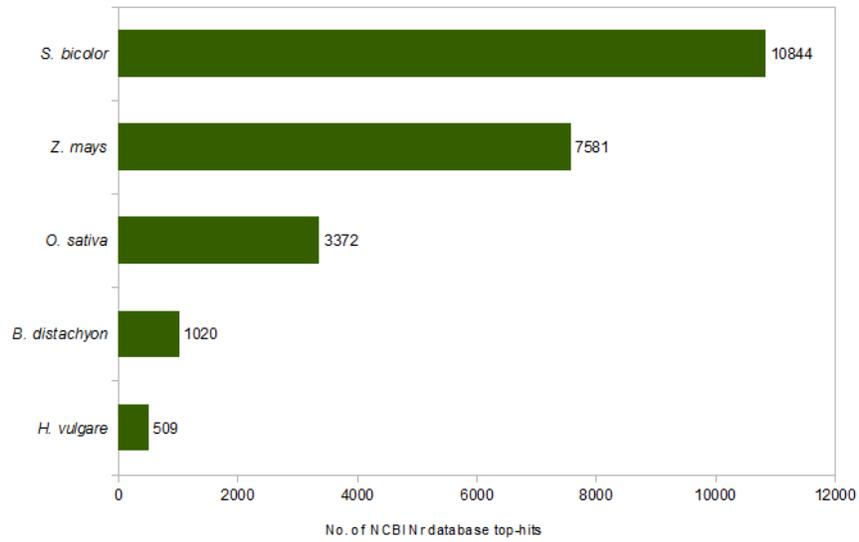


Figure S1. Top-hit species distribution among NCBI nr BLASTx hits.

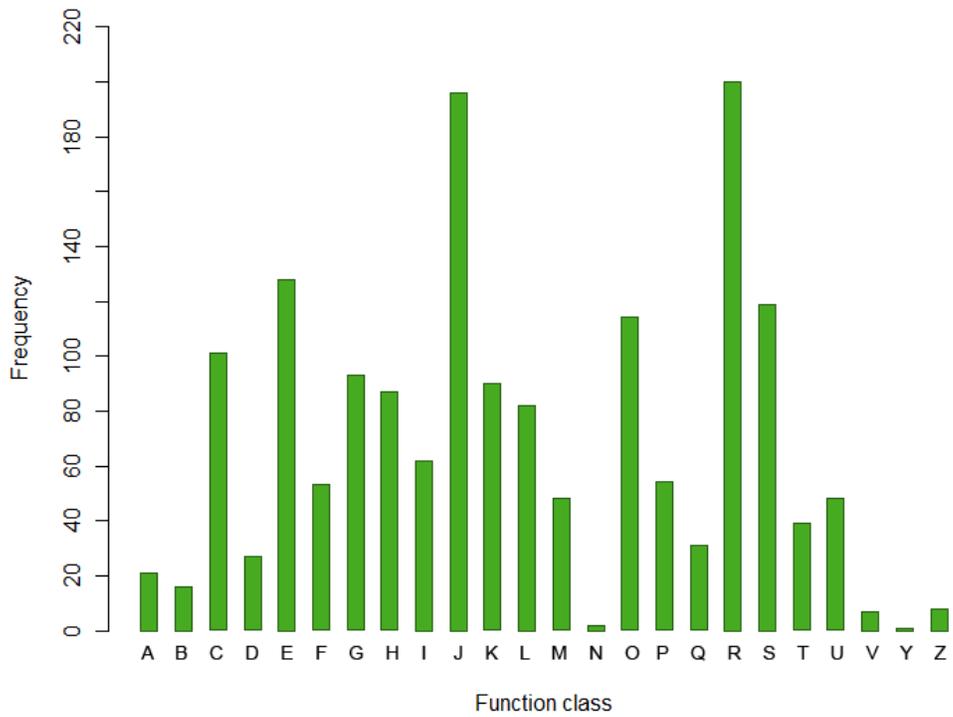


Figure S2: Functional classification of *P. maximum* unigenes based on COG classes.

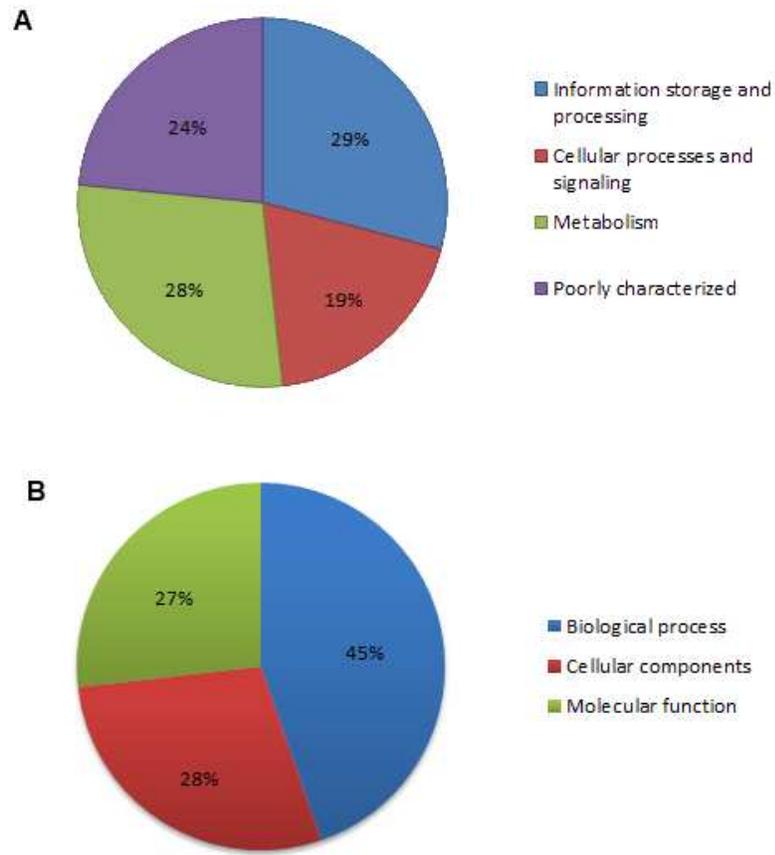
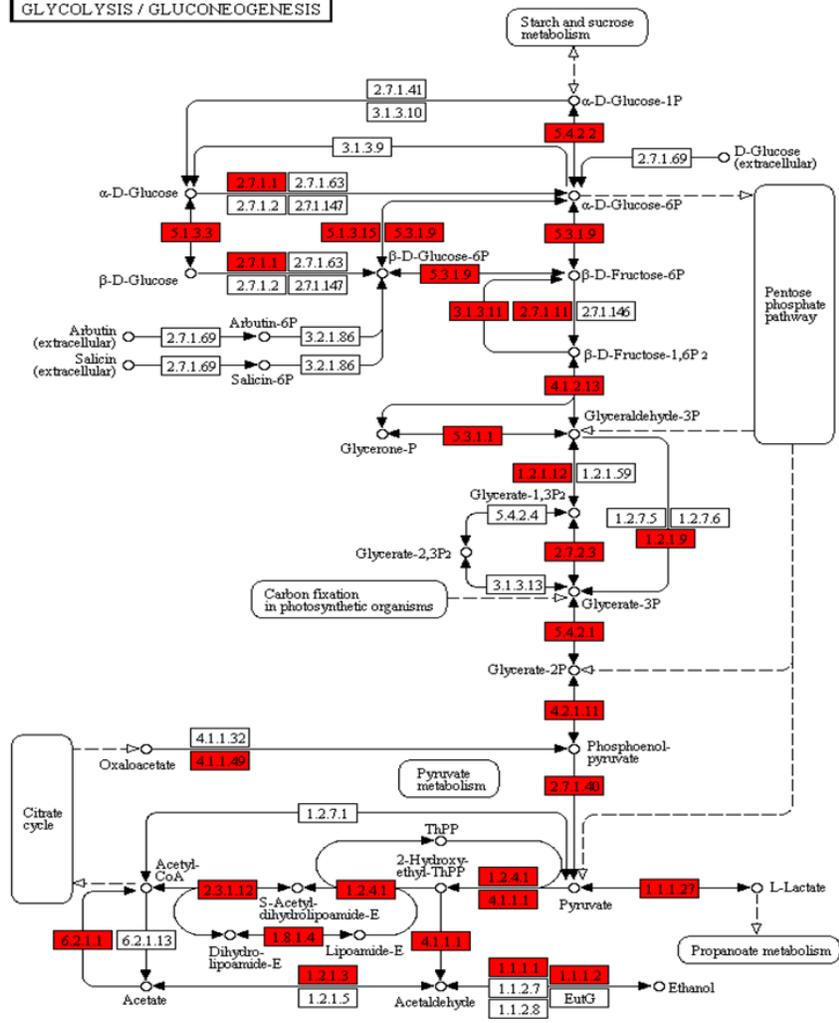


Figure S3. General COG and GO classification of *Panicum maximum* unigenes. (A) General COG classification (B) General GO classification.

GLYCOLYSIS / GLUCONEOGENESIS



00010 531/12
 (c) Kanehisa Laboratories

Figure S4. Glycolysis and gluconeogenesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

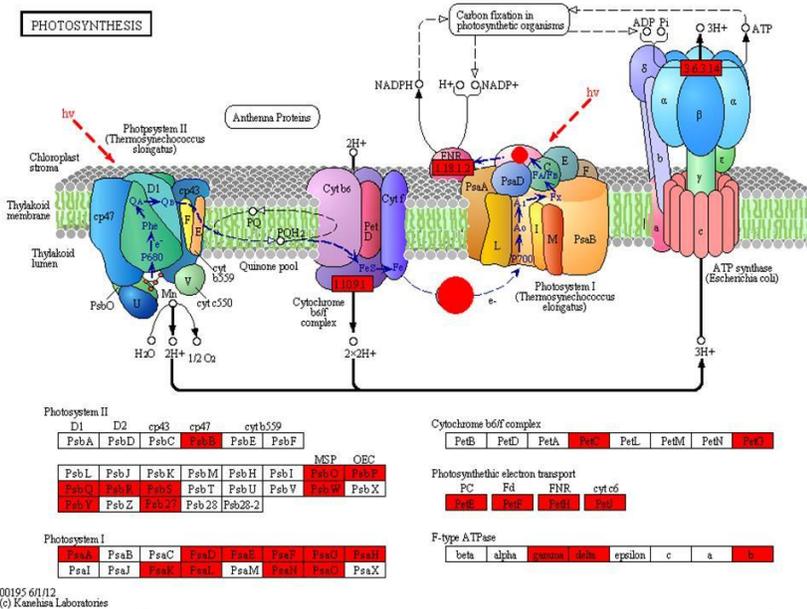


Figure S5. Photosynthesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

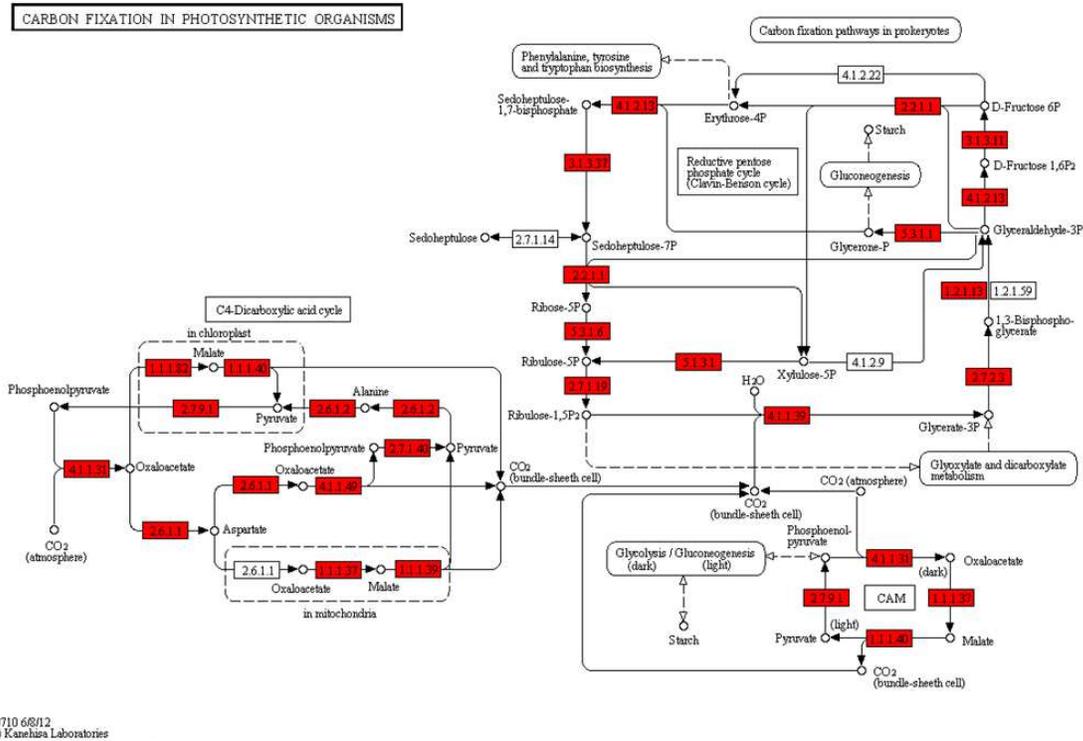


Figure S6. Carbon fixation KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

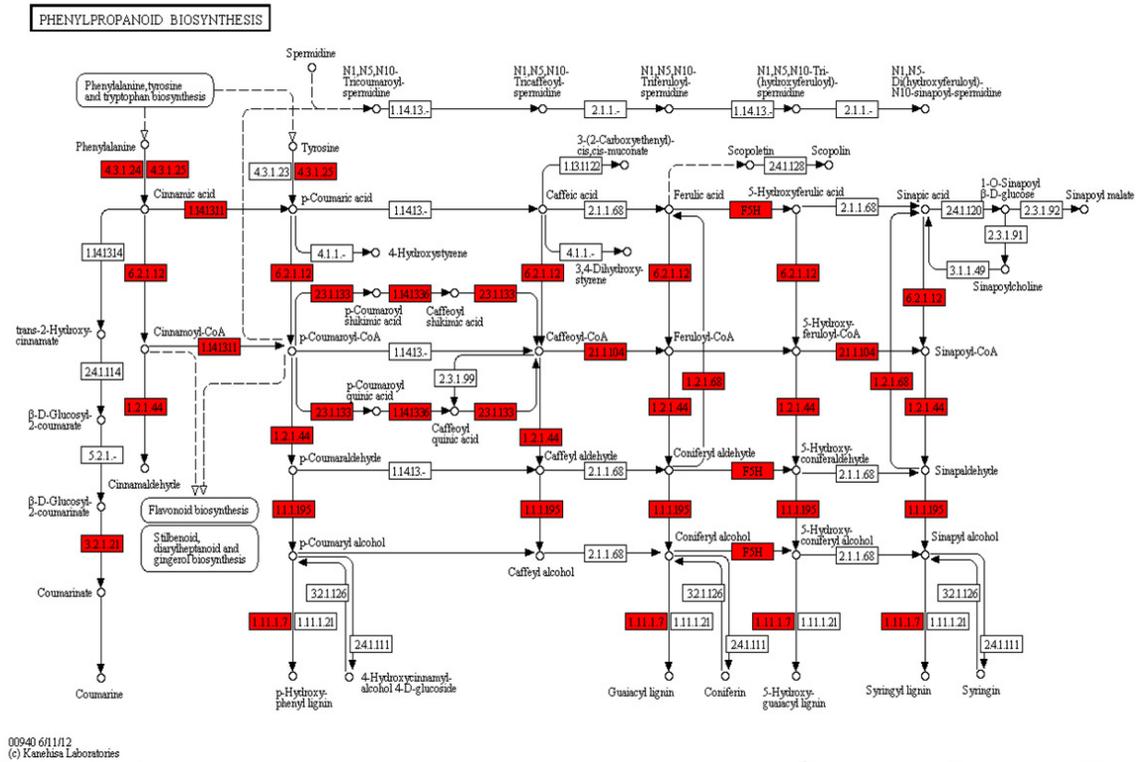


Figure S7. Phenylpropanoid biosynthesis KEGG pathway. Unigenes with associated KEGG orthology terms are shown in red.

Table S1: Statistics of sequenced data for each *P. maximum* genotype.

	S10	S12	Mombaça	Tanzania
Total raw reads	41,459,861	37,275,090	47,523,787	41,794,980
Total data	3.31 Gb	2.98 Gb	3.80 Gb	3.34 Gb
Total high-quality reads	30,799,030	27,633,430	33,178,507	29,227,369
High-quality reads %	74.29%	74.13%	69.81%	69.93%

**Mapeamento genético-molecular de *Panicum
maximum* Jacq., gramínea forrageira tropical
autotetraploide**

**Mapeamento genético-molecular de *Panicum maximum* Jacq., gramínea forrageira
tropical autotetraploide**

Guilherme Toledo-Silva¹, Livia Moura de Souza¹, Sthéphanie Louise Cavalcanti de Brito¹,
Marcelo Mollinari², Antônio Augusto Franco Garcia², Liana Jank³, Anete Pereira de Souza^{1,4}

¹ Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas,
Campinas, SP, 13083-875, Brasil

² Departamento de Genética, Escola Superior de Agricultura "Luiz de Queiroz", Universidade
de São Paulo, Piracicaba, SP, 13400-970, Brasil

³ Empresa Brasileira de Agropecuária, Embrapa Gado de Corte, Campo Grande, MS, 79106-
550, Brasil

⁴ Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de
Campinas, Campinas, SP, 13083-875, Brasil

Correspondência: Anete P. de Souza, fax: +55-19-3521-1089, email: anete@unicamp.br,
Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas,
Campinas, São Paulo, Brasil, CEP 13083-862

Resumo

Panicum maximum, popularmente conhecido no Brasil como capim colonião, é uma gramínea de origem africana, que apresenta distribuição por todo território nacional. O plantio de *P. maximum* se expandiu devido à sua grande adaptação, produtividade e qualidade, fácil propagação por sementes e por ser altamente palatável ao gado. Encontra-se entre as espécies mais utilizadas nas pastagens cultivadas, e devido a sua importância para a pecuária, existe um programa de melhoramento da espécie na Embrapa Gado de Corte (Campo Grande - MS). No presente trabalho, foram construídos mapas de ligação para dois genótipos usados no programa de melhoramento da espécie (S10 e Mombaça), a partir dos quais foi obtida uma população F1 segregante. Para tal fim, foram utilizados 228 marcadores microssatélites, sendo 43 obtidos da literatura e 185 desenvolvidos neste trabalho. Destes marcadores, 174 resultaram em amplificação esperada, dos quais 90 apresentaram polimorfismo entre os genitores e híbridos da população. Marcadores foram analisados utilizando metodologia apropriada para espécies autotetraploides, e marcadores SSR puderam ser genotipados como marcadores codominantes e multi-alélicos. Os mapas de ligação cobriram 672,2 cM (S10) e 802,2 cM (Mombaça). Estes resultados trazem grande contribuição genética para a pesquisa associada a *P. maximum* e forrageiras tropicais, assim como para os programas de melhoramento genético.

Introdução

Panicum maximum Jacq. é a espécie forrageira tropical propagada por sementes mais produtiva no mercado brasileiro. Apesar da alta exigência em fertilidade do solo, oferece pastejo de qualidade e é adaptada às diversas regiões do Brasil e da América Latina. O uso desta gramínea é fator responsável pela engorda de bovinos nestes países (JANK et al., 2008). Esta espécie tem origem africana, mais precisamente na região da África oriental (MUIR e JANK, 2004). *P. maximum* é caracterizada como uma espécie apomítica, assim como a maioria das gramíneas forrageiras tropicais (JANK, VALLE e RESENDE, 2011; WARMKE, 1954). A maioria dos genótipos encontrados são autotetraploides ($2n=4x=32$), e se reproduzem por apomixia aposporica. Alguns acessos com modo de reprodução exclusivamente sexual foram encontrados ao leste da África, sendo estas plantas identificadas como diploides ($2n=2x=16$) (NAKAJIMA et al., 1979).

Panicum maximum é considerada uma das mais importantes forrageiras em regiões tropicais e semitropicais, e ainda assim existe pouco conhecimento genético associado a esta espécie (SOUSA et al., 2011a). A produção animal nas regiões tropicais e subtropicais do mundo é altamente dependente de pastagens nativas ou cultivadas, levando à necessidade de desenvolver forrageiras produtivas, adaptadas e de alta qualidade (JANK, VALLE e RESENDE, 2011). Neste sentido, através do desenvolvimento e uso de marcadores moleculares do tipo microssatélite (também conhecidos como *single sequence repeats* - SSR), a coleção de acessos do banco de germoplasma da Embrapa Gado de Corte (Campo Grande – MS) teve seu perfil genético estudado (SOUSA et al., 2011a; SOUSA et al., 2011b). Em relação ao mapeamento genético da espécie, Ebina e colaboradores (2005), construíram o único mapa da espécie, utilizando apenas marcadores dominantes, gerando um mapa parcial de uma cultivar lançada pelo programa de melhoramento genético japonês.

O mapeamento genético-molecular de espécies poliploides, como no caso de *P. maximum*, se torna mais complexo frente à maior diversidade de genótipos e sua identificação precisa, e ao escasso conhecimento referente ao tipo de poliploidia em muitas espécies (RIPOL et al., 1999). Wu e colaboradores (1992) desenvolveram uma metodologia de mapeamento genético em poliploides, baseada na análise da segregação de marcadores em dose única (MDU). Desta forma, o marcador analisado está presente em uma única cópia, e em apenas um genitor do cruzamento estudado. Um loco nesta condição irá segregar na progênie em uma proporção de 1:1, independente do nível e tipo de ploidia. Marcadores apresentando segregação de 3:1 na progênie também podem ser utilizados, representando marcadores em dose única presentes em ambos os genitores. Em gramíneas, diversos mapas genéticos foram construídos por meio da metodologia de Wu et al. (1992), incluindo espécies como *Poa pratensis* (PORCEDDU et al., 2002), *Pennisetum ciliare* (JESSUP et al., 2003), *Festuca arundinacea* (SAHA et al., 2005), *Panicum maximum* (EBINA et al., 2005), *Cynodon* sp. (BETHEL et al., 2006), *Paspalum notatum* (STEIN et al., 2007), *Saccharum* spp. (OLIVEIRA et al., 2007) e *Panicum virgatum* (OKADA et al., 2010).

Com relação a espécies autotetraploides, classe de poliploide em que *P. maximum* se enquadra, estudos sobre segregação gênica e construção de mapas de ligação foram realizados (HACKETT, PANDE e BRYAN, 2003; HACKETT et al., 1998; LUO et al., 2000, 2001; MEYER et al., 1998) e culminaram no desenvolvimento de um programa apropriado para o mapeamento genético-molecular de espécies autotetraploides, a partir de populações F1-segregantes, denominado TetraploidMap (HACKETT et al., 2007).

A apomixia torna o mapeamento genético uma tarefa ainda mais complexa para *P. maximum*, pois não permite a obtenção de linhagens endogâmicas. Dessa forma, o mapeamento deve ser feito em uma progênie de irmãos-completos, originada a partir de genitores heterozigotos, que segregam muitos alelos por loco e a fase de ligação dos locos nos genitores

é geralmente desconhecida (MALIEPAARD, JANSEN e OOIJEN, VAN, 1997). A metodologia de Hackett e colaboradores (HACKETT et al., 2007) permite o uso de diversos marcadores no mapeamento genético, incluindo os MDU presentes em somente um genitor ou ambos (segregação 1:1 ou 3:1, respectivamente), marcadores em dose dupla, presentes em um ou ambos os genitores (segregação de 5:1 e 35:1) e marcadores codominantes. Desta forma é possível construir um mapa genético-molecular baseado em marcadores dominantes e codominantes, onde são formados inicialmente grupos de ligação e estes são então separados em cromossomos homólogos baseado na informação de fase de ligação entre os marcadores codominantes e dominantes (HACKETT e LUO, 2003; HACKETT et al., 2007).

O objetivo deste trabalho foi desenvolver um mapa genético de *P. maximum* com marcadores microssatélites utilizando metodologia apropriada para espécies autotetraploides implementada no programa TetraploidMap, a partir de cruzamento F1 intraespecífico realizado na Embrapa Gado de Corte (Campo Grande – MS). O mapa genético-molecular poderá ser usado em estudos genéticos básicos sobre a apomixia, evolução do genoma da espécie e o comportamento genômico em híbridos, e fornecer informações para o programa de melhoramento genético da espécie e também do gênero *Panicum*. Tal mapa também pode ser usado no futuro para trabalhos de mapeamento de *quantitative trait loci* (QTL), identificando locos que controlam a variação de caracteres quantitativos.

Material e Métodos

População de mapeamento

A população de mapeamento utilizada, uma progênie de irmãos completos que consta de 137 indivíduos F1, foi gerada a partir de um cruzamento intraespecífico entre o acesso sexual S10 e a cultivar apomítica Mombaça, ambos tetraplóides ($2n=4x=32$). Este cruzamento

foi realizado pela pesquisadora Dra. Liana Jank (Embrapa Gado de Corte, Campo Grande – MS) que também é responsável pela manutenção desta população que vem sendo cultivada desde Janeiro de 2009. O genótipo sexual foi tetraploidizado anteriormente através do uso de colchicina, estando apto para geração de híbridos em conjunto com acesso autotetraploide. Estes genitores apresentam características de produção contrastantes como produção total de matéria seca (MST), produção de matéria seca de folhas (MSF), produção de matéria seca de colmos (MSC), digestibilidade *in vitro* de matéria orgânica (DIVMO) e valor de sílica biogênica (SIL) (BRAZ, 2012). Amostras de folhas jovens de cada um dos híbridos, bem como de seus genitores, foram coletadas e liofilizadas na Embrapa Gado de Corte. Este material foi então transportado até o Laboratório de Análise Genética e Molecular (Unicamp – Campinas – SP) para processamento em moinho mecânico. O DNA foi extraído segundo Sousa (SOUSA et al., 2011a), e sua integridade e quantidade verificada em gel de agarose 1% corado com brometo de etídio.

Desenvolvimento de novos marcadores do tipo microssatélite

Uma biblioteca enriquecida de microssatélites foi construída durante o processo de avaliação da diversidade genética do banco de germoplasma de *P. maximum* presente na Embrapa Gado de Corte, por Sousa e colaboradores (SOUSA et al., 2011a; SOUSA et al., 2011b) onde foram publicados 86 locos microssatélites, dos quais 43 foram utilizados para o mapeamento genético-molecular de *Panicum maximum* no presente trabalho. A partir desta biblioteca enriquecida de SSRs, 166 novos pares de oligos foram desenvolvidos para genotipagem, além de 19 pares de iniciadores baseados no transcriptoma da espécie *P. maximum* (TOLEDO-SILVA et al., 2013). Todos iniciadores foram desenhados utilizando o programa BatchPrimer3 (YOU et al., 2008) seguindo os parâmetros: (a) *primer type: SSR screening and primers*; (b) *primer size: mínimo (min) 18, ótimo (opt) 20 e máximo (max) 24*;

(c) *primer Tm*: 55 (min), 60 (opt), 65 (max); (d) *primer GC%*: 45 (min), 50 (opt) e 55 (max); aceitando máximo de diferença de *Tm* (*Max Tm difference*) de dois graus Celsius. Em relação ao tamanho dos motivos, o tamanho mínimo adotado foi de 10 pb, sendo considerados os seguintes valores mínimos em relação ao tipo de repetição: (a) dinucleotídeos – 5 repetições (rep); (b) trinucleotídeos – 4 rep; (c) tetra e pentanucleotídeos – 4 rep. Os oligos selecionados foram testados em uma sub-amostra de sequências genômicas obtidas no sequenciamento da biblioteca enriquecida (SOUSA et al., 2011a), utilizando a função de PCR *in silico* do programa FastPCR (KALENDAR, LEE e SCHULMAN, 2009). Aqueles com amplificação *in silico* específica (somente o produto esperado), foram selecionados para síntese, com adição da sequência referente a cauda M13 (5'-CACGACGTTGTAAAACGA-3') no iniciador *forward*, conforme metodologia proposta por Schuelke (2000).

Os 185 novos marcadores foram amplificados por reação de PCR, utilizando um programa de gradiente de temperatura que variou de 45° a 65°C. As reações de PCR (10 µL no total) foram realizadas utilizando 5 ng de DNA molde, 0.2 µM de cada iniciador, 200µM de dNTP, 2mM de MgCl₂, 1X tampão de PCR e 1U Taq DNA polimerase. Os marcadores com metodologia de genotipagem baseada em coloração em prata (CRESTE, NETO e FIGUEIRA, 2001) foram amplificados a partir de um passo inicial de denaturação de 4 min a 94 °C, seguido por 30 ciclos de denaturação, anelamento e extensão (94°C por 45 s), temperatura de anelamento específica (*T_a*) por 1 min, 72°C por 1 min 15 s e um passo de extensão final a 72°C por 5 min. Para os marcadores baseados na metodologia de SCHUELKE (2000), a ciclagem foi baseada em programa de *touchdown*, com um passo inicial de denaturação de 4 min a 94°C, seguido por 10 ciclos de denaturação, anelamento e extensão com um decréscimo de 0.5°C na temperatura de anelamento, de 60°C a 55°C (94°C por 45 s, 55°C por 1 min, 72°C por 1 min 15 s), 25 ciclos com temperatura de anelamento a 50°C (94°C por 45 s, 50°C por 1 min, 72°C por 1 min), e um passo de extensão final a 72°C por 5 min. Alguns pares de marcadores

tiveram resultados melhores quando o programa de *touchdown* variou de 55°C a 50°C. Os resultados desta otimização inicial foram avaliados em gel de agarose 3% corado com brometo de etídio. Os locos nos quais o perfil de amplificação de banda foi considerado específico para o tamanho esperado, foram selecionados para teste de polimorfismo. Nesta etapa, todos os pares de marcadores otimizados foram amplificados nos dois genitores da população (S10 e Mombaça) e mais oito híbridos, utilizando a respectiva metodologia de genotipagem para cada marcador: coloração por nitrato de prata (CRESTE, NETO e FIGUEIRA, 2001) ou em genotipador Li-Cor 4300 DNA Analyzer (SCHUELKE, 2000). Em ambas as técnicas, os produtos de PCR foram denaturados por 5 min a 94°C antes da separação por tamanho em um gel de poliacrilamida 6.5%. Os marcadores avaliados como polimórficos foram genotipados no restante da população.

Construção do mapa genético

A partir do resultado da genotipagem dos genitores e população F1, foi construído um mapa de ligação para a espécie *P. maximum*, através do programa TetraploidMap (HACKETT et al., 2007). Cinco etapas principais foram empregadas para a construção do mapa genético utilizando esta metodologia:

(a) **Inferência dos genótipos e segregações:** Através da função *findgeno*, cada marcador foi analisado em relação à segregação observada na progênie, onde a dosagem mais provável de cada loco foi inferida na presença e ausência de dupla redução. Nos casos onde não foi possível estimar um genótipo mais provável para o marcador codominante, os marcadores foram então lidos como dominantes, e novamente submetidos a função *findgeno*. A segregação foi avaliada através de teste de qui-quadrado. Foram selecionados apenas os marcadores dominantes de segregação 1:1, 3:1 e 5:1, adotando a significância estatística de 0,001 para os marcadores 1:1 e 3:1, e 0,01 para os 5:1. Marcadores com distorções em relação a segregação esperada foram

desconsiderados das análises, assim como aqueles com segregação de 11:1 e 35:1, devido a menor confiança neste tipo de marca em relação ao tamanho da população.

(b) **Clusterização:** os marcadores selecionados na primeira etapa foram submetidos à função *cluster*, onde foi realizado um teste de qui-quadrado para segregação independente dos cromossomos, resultando em um dendograma baseado em distâncias de similaridade. Esta análise foi conduzida separadamente para cada genitor, utilizando todos os marcadores dominantes e codominantes multialélicos, considerando a análise de clusterização média para dividi-los em grupos de ligação mais prováveis.

(c) **Estimação da frequência de recombinação:** através da função *initial run* e *twopoint*, em conjunto com o algoritmo *ripple*, as frequências de recombinação e respectivos valores de *LOD score* (razão de verossimilhança em logaritmo de base 10) entre todos os marcadores dentro de um grupo de ligação (GL) foram estimadas. Apenas os resultados com valores de $LOD > 3.0$ foram considerados.

(d) **Ordenação:** a ordenação dos marcadores dentro dos GLs foi feita baseada nos dados obtidos da etapa anterior, onde a fase com maior verossimilhança (baseada nas frequências de recombinação e valores de LOD) foi selecionada para ordenação dos marcadores. O algoritmo *simulated annealing* foi utilizado para identificar a ordem final de cada grupo de ligação, pressupondo o menor valor dos mínimos quadrados para calcular a distância entre os marcadores do mapa (BRADSHAW et al., 2008; HACKETT e LUO, 2003; HACKETT et al., 2007).

(e) **Definição dos cromossomos homólogos:** Através da informação da fase de ligação entre os marcadores, cada grupo de ligação foi reconstruído em quatro cromossomos homólogos quando possível. A informação sobre a fase mais provável de ligação entre marcadores ordenados, juntamente com a frequência de recombinação e valores de LOD permitiram a inferência manual das fases dos marcadores (HACKETT et al., 2007).

Os mapas de ligação dos genitores, e a colineariedade entre eles, foram desenhados utilizando o programa MAPCHART 2.2 (VOORRIPS, 2002).

Resultados

Marcadores microssatélites

Neste trabalho, 228 marcadores microssatélites foram utilizados no total. Destes, 43 foram obtidos da literatura (SOUSA et al., 2011a; SOUSA et al., 2011b), 166 foram desenvolvidos a partir de biblioteca enriquecida de microssatélites (SOUSA et al., 2011a), e 19 a partir de sequências transcritas da espécie (TOLEDO-SILVA et al., 2013). Inicialmente, todos os marcadores microssatélites foram otimizados quanto a temperatura de anelamento, sendo que 174 (76,3%) foram amplificados com sucesso, resultando em banda de tamanho esperado e com perfil de bandas adequado para a genotipagem. Os 174 locos com amplificação bem sucedida, e seus respectivos motivos de repetição, temperaturas de anelamento, tamanho esperado dos produtos amplificados, sequências dos *primers*, número de alelos por loco, além de existência de polimorfismo e utilização no mapa de ligação estão apresentados na Tabela Suplementar 1. Destes, 84 (48,2%) apresentaram-se monomórficos entre os genitores e os indivíduos F1, não gerando informações úteis para a construção do mapa e 90 (51,8%) locos apresentaram polimorfismo e amplificaram 247 alelos, com uma média de 2.74 alelos por loco.

Análise dos marcadores

O mapa de ligação para a espécie *P. maximum* foi construído através do programa TetraploidMap (HACKETT et al., 2007). O primeiro passo na construção do mapa foi a inferência dos genótipos dos genitores a partir dos dados da progênie, e assim definir aqueles marcadores como codominantes ou dominantes. Dos 90 marcadores polimórficos genotipados

na população de mapeamento, foi possível inferir com sucesso o genótipo mais provável dos genitores em 20 marcadores moleculares, e estes foram utilizados como marcadores multi-alélicos e codominantes, o que permitiu a correlação entre os grupos de ligação dos dois diferentes genitores. Os 70 marcadores microssatélites restantes foram genotipados como 189 marcadores dominantes, e divididos em marcadores uni-parentais, apresentando heterozigiosidade (dosagem simplex-1:1 ou duplex 5:1) em um único genitor, e marcadores bi-parentais (simplex-3:1, duplex-35:1, e duplex e simplex-11:1) apresentando heterozigiosidade em ambos os pais. Todos os tipos de segregação foram avaliados através do teste de qui-quadrado com a hipótese nula de acordo com a razão esperada de segregação mendeliana, utilizando valores de significância 0,001 para os marcadores 1:1 e 3:1 e 0,01 para os marcadores 5:1. Desta maneira, 136 (71,9%) marcadores dominantes foram selecionados para construção do mapa, em conjunto com os 20 marcadores SSR multi-alélicos (Tabela 1). Aqueles com segregação de 11:1 e 35:1, resultando em 16 marcadores (8,4%) foram descartados, juntamente com 37 (19,5%) apresentando distorções nas segregações esperadas. Quando comparadas todas as segregações, foi observado maior número de marcadores dominantes com segregação 1:1, com 41 para o genitor S10 e 50 para o genitor Mombaça (Tabela 1). Um total de 29 marcadores bi-parentais com segregação 3:1, em conjunto com os 20 marcadores multi-alélicos co-dominantes (Tabela 1), foram integrados ao mapa de ligação, permitindo comparações entre os mapas construídos para cada genitor.

Construção do mapa

A clusterização dos marcadores, a uma similaridade média de 0.3 resultou em 10 grupos de ligação para o genitor S10 (Figura 1A), e a uma similaridade de 0.36, em 13 grupos de ligação para o genitor Mombaça (Figura 1B). Para identificar possíveis erros nesta etapa de clusterização, foi realizada uma análise combinando análise de dois-pontos (função *twopoint*)

com função *initial-run*, seguidos de análise pelo algoritmo *ripple*. Esta análise identificou 38 (18,6%) marcadores posicionados de forma errônea, os quais foram checados novamente quanto a frequência de recombinação e valores de LOD em relação a outros marcadores do mesmo GL. Se através destes parâmetros utilizados, um determinado marcador aparentava não pertencer a um grupo de ligação, o mesmo foi excluído (21 ou 10%) ou então movido manualmente para um diferente GL (17 ou 8,1%) onde existissem menores frequências de recombinação e maiores valores de LOD.

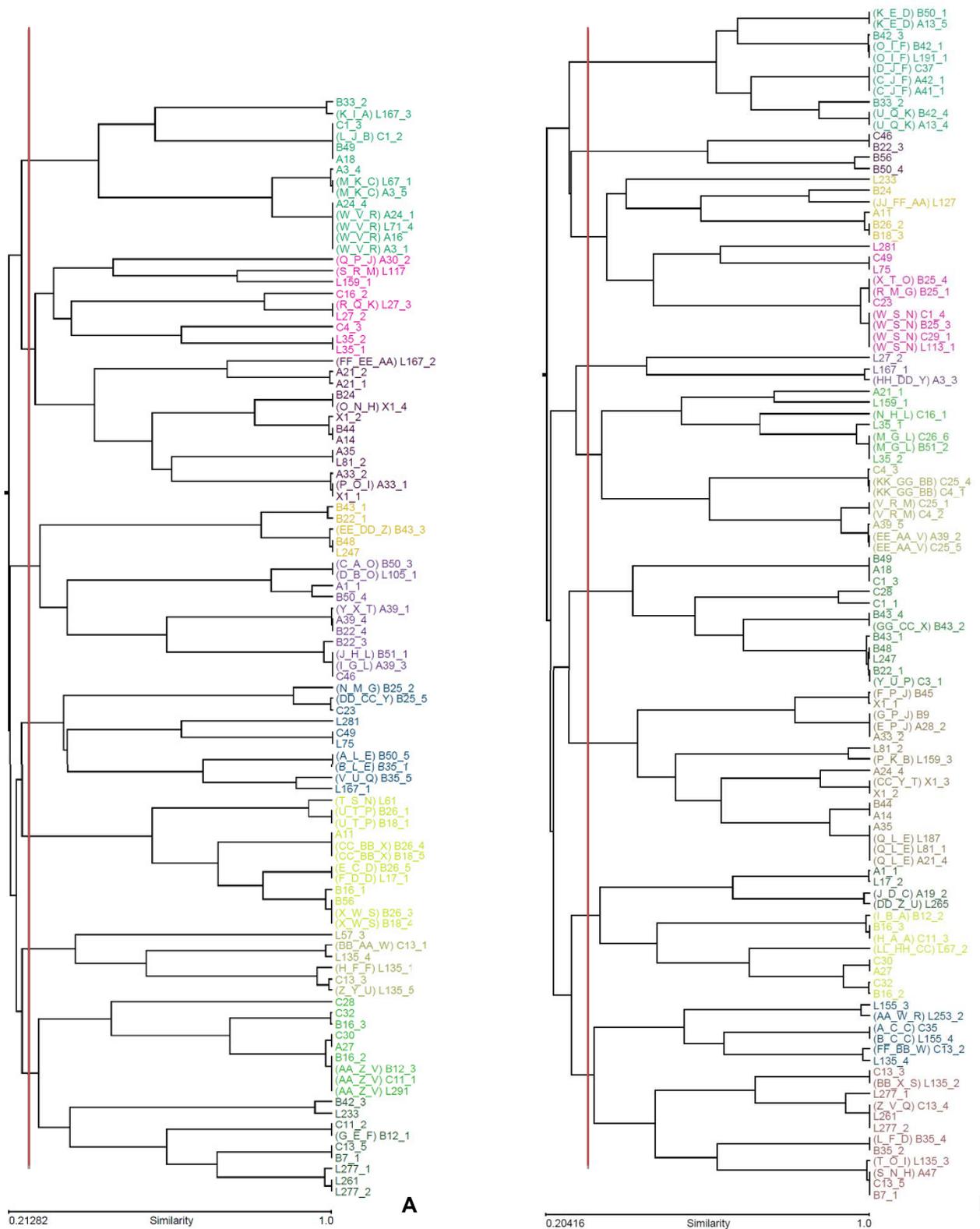


Figura 1: Dendrogramas das distâncias de similaridade a partir dos dados de marcadores selecionados. *Threshold* de similaridade indicado em vermelho. Cada grupo de ligação formado corresponde a uma diferente cor utilizada nos nomes dos marcadores. (A) Dendrograma gerado a partir dos marcadores do genitor S10. (B) Dendrograma gerado a partir do genitor Mombaça.

Tabela 1: Resumo dos marcadores SSR utilizados neste trabalho, discriminados quanto a segregação e genitores.

Tipo de marcador	Marcadores totais	Marcadores selecionados*
Codominantes	20	20
Dominantes	189	136
S10		
1:1	52	41
5:1	16	8
Mombaça		
1:1	53	50
5:1	10	8
Ambos genitores		
3:1	42	29
11:1	14	-
35:1	2	-
Total	209	156

*Marcadores selecionados utilizando valor de $p < 0,001$ para os marcadores 1:1 e 3:1, e $p < 0,01$ para os marcadores 5:1. Os marcadores com segregação de 11:1 e 35:1 foram descartados.

Tabela 2: Informações sobre os mapas de ligação construídos para os genitores S10 e Mombaça.

	S10	Mombaça
Grupos de ligação	10	13
Total de marcadores	88	96
Marcadores codominantes	19	18
Marcadores dominantes	69	78
1:1	37	46
5:1	7	7
3:1	25	25
Distância coberta pelo mapa (cM)	672.1	802.2
Distância média entre marcadores (cM)	7.63	8.35

A etapa de ordenação foi realizada utilizando o algoritmo de *simulated annealing* (HACKETT, PANDE e BRYAN, 2003) que explora as melhores ordens possíveis dentro de um GL, e assim, os mapas parciais dos genitores S10 (Figura 2) e Mombaça (Figura 3) foram obtidos. Um total de 135 marcadores foram integrados aos mapas dos genitores, considerando todos os tipos de marcadores utilizados (Tabelas 1, Suplementar 2, Suplementar 3 e Suplementar 4). Para o genitor S10, o mapa de ligação cobriu 672.1 cM, com 88 marcadores apresentando uma distância média de 7.63 cM, sendo que a maior distância entre dois marcadores (*gap*) foi de 46,4 cM (Tabela 2). Dez marcadores permaneceram não ligados ao mapa genético. O menor grupo de ligação apresentou 0.8 cM, enquanto o maior cobriu 188.5 cM. O mapa do genitor Mombaça cobriu 802.2 cM, com 96 marcadores posicionados, a uma média de distância entre os marcadores de 8.35 cM e o maior *gap* de 41,6 cM. Dos 107 marcadores utilizados para este genitor, 11 permaneceram sem ligação ao mapa genético. O menor GL apresentou 1,4 cM e o maior 122,4 cM.

Separação dos cromossomos homólogos

O mapa genético de uma espécie autopoliplóide apresenta dois componentes: os grupos de ligação e os conjuntos de cromossomos homólogos (RIPOL et al., 1999). O segundo componente foi obtido a partir dos grupos de ligação originais. Utilizando os resultados da análise dos pares de marcadores, frequência de recombinação, LOD scores e da codificação dos marcadores simplex, fornecida pelo programa TetraploidMap (HACKETT et al., 2007), foi possível determinar manualmente a fase de ligação dos marcadores ordenados na maioria dos grupos de ligação (Figuras Suplementares 1 e 2), separando-os nos quatro cromossomos homólogos para o grupo respectivo, com exceção de nove grupos (S10 – GL2, GL6, GL9 e GL10; Mombaça – GL5, GL8, GL11, GL12 e GL13), que por serem formados por poucos

marcadores dominantes possibilitaram a identificação de apenas um, dois ou três cromossomos homólogos.

Discussão

O desenvolvimento de mapas genético-moleculares para *P. maximum* é uma etapa essencial para a compreensão da arquitetura genômica e identificação de QTLs associados a características de interesse. Através do cruzamento de genitores heterozigotos, uma população F1 foi utilizada para construção do mapa de ligação, sendo tal delineamento recomendado para o mapeamento de espécies autotetraploides (HACKETT et al., 1998). A população de 137 indivíduos F1 segregantes foi originada a partir de cruzamento entre o genótipo sexual S10 e a cultivar apomítica Mombaça. Os genitores da população são contrastantes para diversas características de interesse para o melhoramento genético, e estão sendo avaliadas em campo, onde estes dados serão utilizados para busca de QTLs utilizando o mapa de ligação da espécie.

Os marcadores escolhidos para este trabalho foram os microssatélites, devido a disponibilidade de locos polimórficos para análise na literatura, além de serem codominantes, transferíveis e com alta taxa de reprodutibilidade. Foram utilizados marcadores publicados (SOUSA et al., 2011a; SOUSA et al., 2011b) e novos marcadores desenvolvidos a partir de biblioteca enriquecida de SSRs (SOUSA et al., 2011a) e transcriptoma da espécie (TOLEDO-SILVA et al., 2013). Cerca de 50% dos marcadores analisados apresentaram polimorfismo entre os genitores da população e a progênie. Contudo, os marcadores avaliados como monomórficos podem ser utilizados em outros contextos genéticos, visto que a variabilidade analisada neste experimento é parte ínfima da diversidade genética total de *P. maximum*. Os 174 marcadores amplificados representam um novo e valioso conjunto significativo de ferramentas para o estudo da biologia de *P. maximum* e espécies próximas, assim como o uso

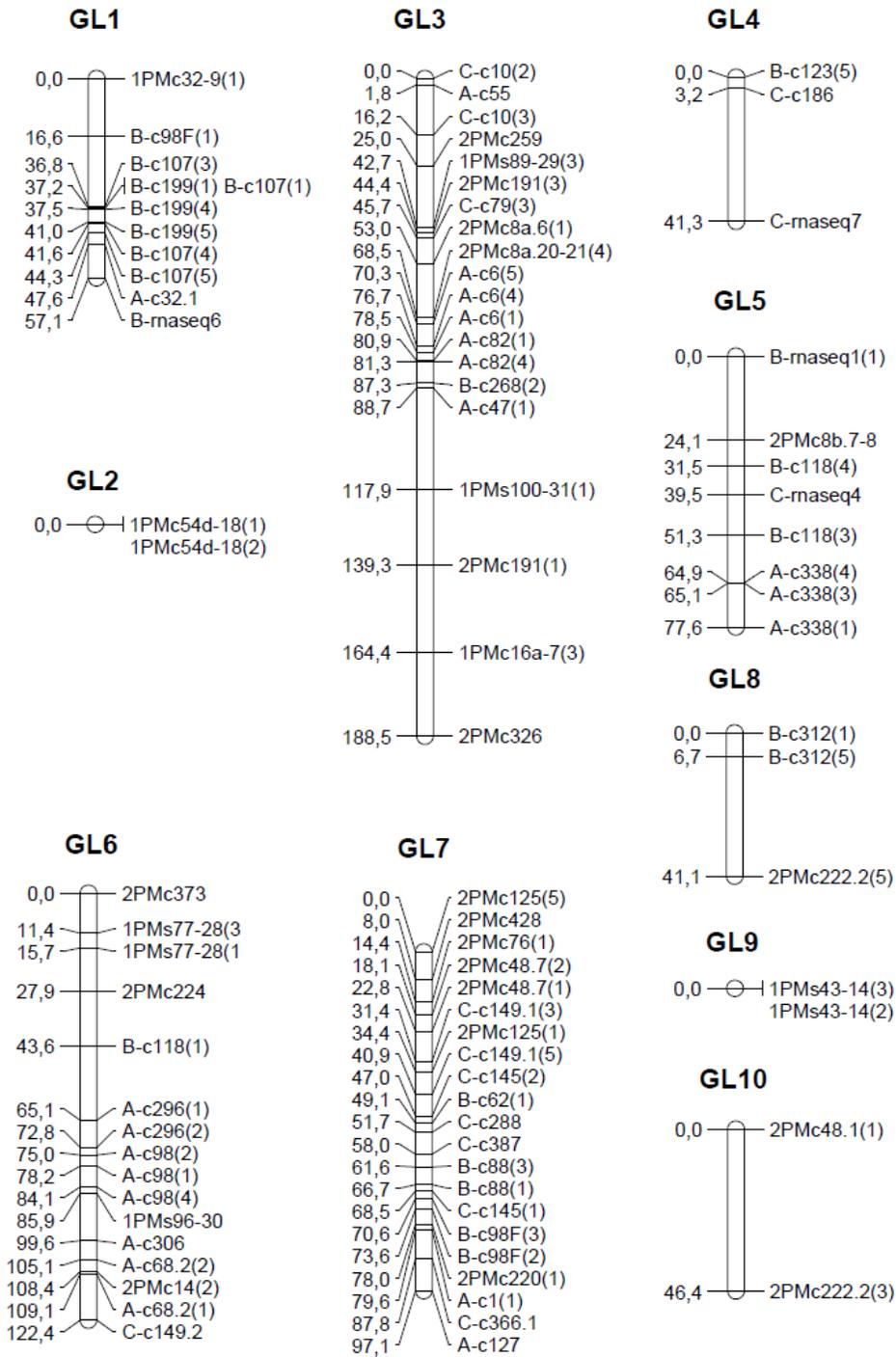


Figura 2: Mapa de ligação referente ao genitor S10. Dez grupos de ligação (GL) são representados, com a distância entre marcadores em centimorgans (cM).

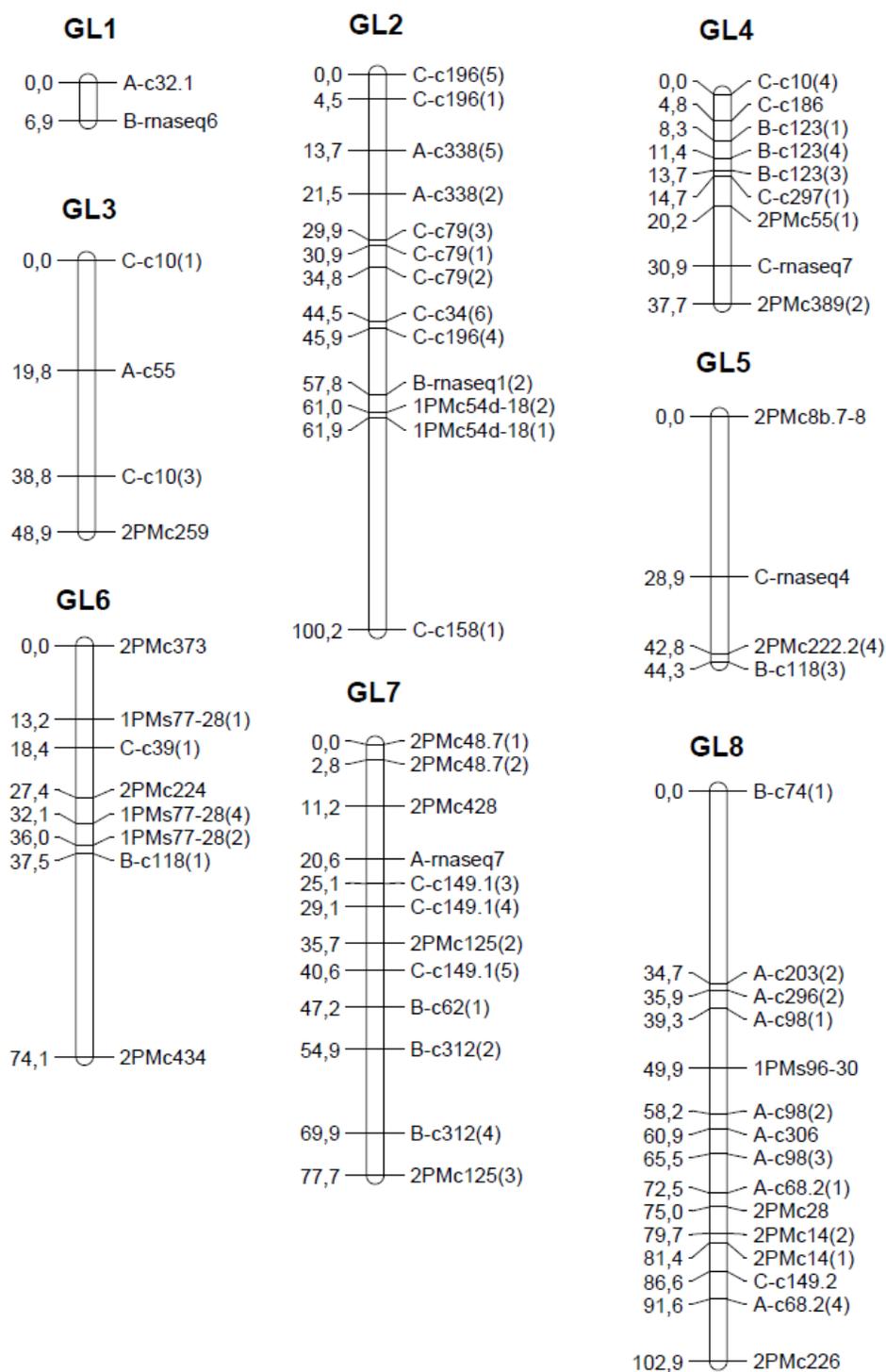


Figura 3: Mapa de ligação referente ao genitor Mombaça. Treze grupos de ligação (GL) são representados, com a distância entre marcadores em centimorgans (cM).

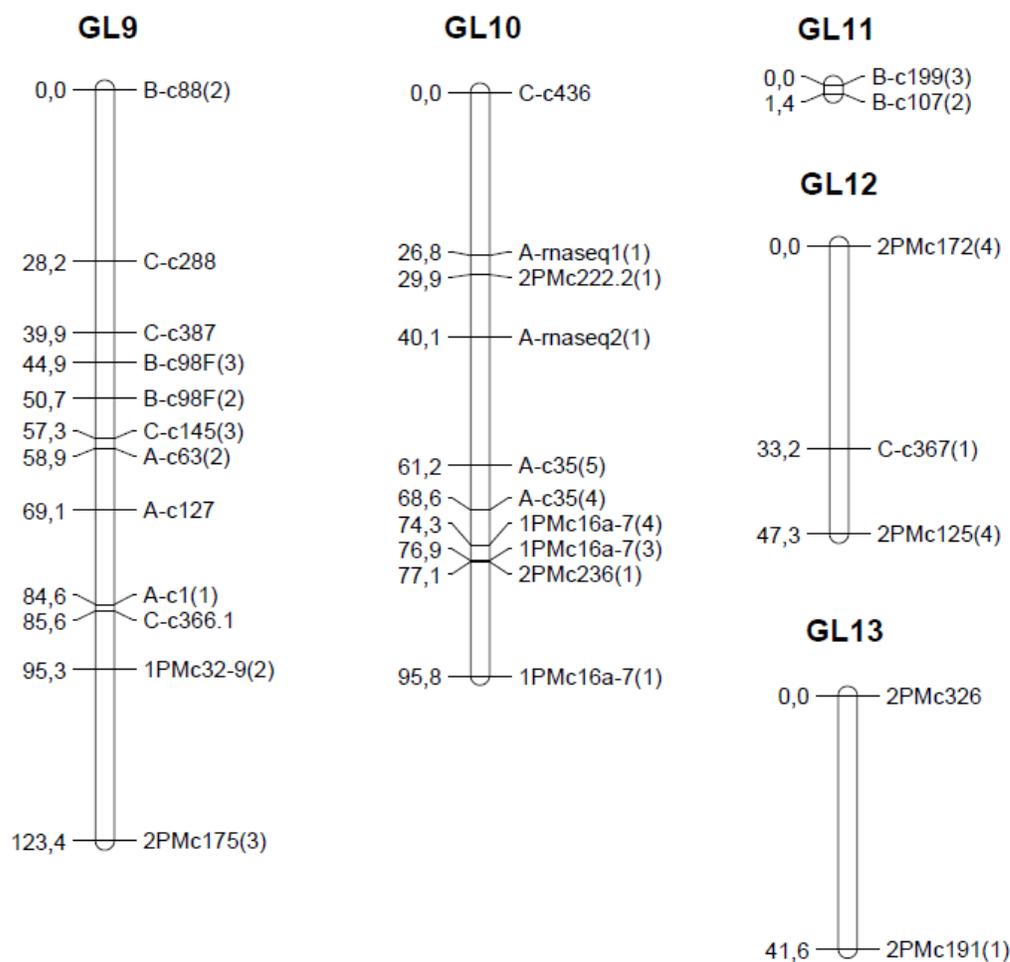


Figura 3: continuação.

nos programas de melhoramento. A proporção encontrada de marcadores dominantes simplex e duplex entre os dois genitores foi praticamente a mesma, com um número levemente maior no genótipo S10 (Tabela 1), assim como número de alelos dos marcadores codominantes (2,7 alelos por loco/genitor). No presente trabalho, alguns locos apresentaram frequência de dupla-redução significativa, e aqueles nos quais os genótipos possíveis eram explicados exclusivamente na presença de dupla-redução (10 locos) foram removidos da análise, uma vez que este tipo de situação não pode ser considerada pelo modelo utilizado no programa TetraploidMap (HACKETT et al., 2007). Não existem dados sobre as taxas de dupla-redução

em *P. maximum*, e a detecção destes valores com maior precisão necessita de populações maiores para uma análise mais robusta. Onze marcadores desenvolvidos apresentaram amplificação multi-loco em *P. maximum*, fato observado também em alfafa (JULIER et al., 2003) e *Brachiaria humidicola* (VIGNA, 2010). Uma hipótese é associar esses locos a regiões genômicas duplicadas com diferenciação na composição, devido a diferentes caminhos evolutivos ao longo da história biológica da espécie. Estudos mais aprofundados nestes locos poderiam auxiliar na compreensão da evolução do genoma de *P. maximum*. Em relação aos métodos de genotipagem testados, foi constatada uma maior taxa de sucesso utilizando a metodologia de SCHUELKE (2000), devido principalmente ao maior grau de sensibilidade e detecção do método. Dois marcadores considerados polimórficos pela metodologia de coloração por prata (CRESTE, NETO e FIGUEIRA, 2001) mostraram-se monomórficos quando adaptados para metodologia de SCHUELKE (2000). Isso deve-se ao maior poder de detecção da técnica, que nestes casos, puderam identificar alelos específicos que eram ausentes na coloração por prata. Também, foi notado um elevado grau de sucesso na transferência da técnica de coloração por prata (CRESTE, NETO e FIGUEIRA, 2001) para genotipagem no Li-Cor 4300 DNA Analyzer, na qual 13 marcadores foram transferidos com sucesso, seguindo a metodologia de Schuelke (2000).

O mapa foi construído utilizando o programa TetraploidMap (HACKETT et al., 2007), que permite a construção de mapas a partir de marcadores dominantes e codominantes (multi-alélicos), em populações de irmãos-completos provenientes de cruzamento de espécies autotetraploides. A construção de mapas utilizando o TetraploidMap em espécies autotetraploides foi bem sucedida em batata (*Solanum tuberosum*, BRADSHAW et al., 2008), rosa (*Rosa hybrida*, GAR et al., 2011) e alfafa (*Medicago sativa*, JULIER et al., 2003). Foram gerados dois mapas parciais, um para cada genitor da população (Figuras 2 e 3). O mapa de ligação do genótipo S10 apresenta 88 marcas distribuídas em 10 grupos de ligação, com

cobertura de 672.1 cM. O mapa da cultivar Mombaça apresenta 96 marcadores entre 13 grupos de ligação, cobrindo 802.2 cM. Ambos os mapas ainda não estão saturados, visto os nove grupos com apenas dois ou três marcadores (S10 – GL2, GL4, GL8, GL9 e GL10; Mombaça – GL1, GL11, GL12 e GL13) e os oito grupos de ligação com gaps maiores do que 30 cM (S10 – GL4, GL8 e GL10; Mombaça – GL2, GL6, GL8, GL12 e GL13). Seria esperado encontrar 8 GLs nesta espécie ($2n=4x=32$), entretanto, a identificação de 10 e 13 GLs (para S10 e Mombaça, respectivamente) provavelmente é decorrência de uma pequena amostragem do genoma, resultando em GLs separados que não se ligam por falta de um ou mais marcadores que consigam estabelecer relações de ligação entre dois GLs. Considerando os cromossomos homólogos reconstruídos a partir dos GLs, foram obtidos 31 homólogos para S10 e 36 para o genitor Mombaça, em comparação com os 32 esperados segundo o número básico da espécie. Os GLs formados por poucos marcadores dominantes não oferecem informação suficiente para separação nos quatro cromossomos homólogos de um mesmo grupo. Com uma maior saturação do mapa esperamos alcançar o número esperado de homólogos para a espécie.

Para validação dos mapas gerados para cada um dos genitores (S10 e Mombaça), os marcadores codominantes foram fundamentais pois, através dessas marcas informativas e que segregam em ambos os genitores é que foi possível identificar GLs homólogos entre os dois mapas (Figuras 4 e 5). No total foram genotipados 20 marcadores multialélicos, com o estabelecimento de 18 relações possibilitando correlação e uma maior taxa de ligação entre os diferentes mapas (HACKETT et al., 2007). Ainda que menos informativas, as 25 marcas do tipo 3:1 permitiram o estabelecimento de 20 relações entre os mapas de cada genitor, o que vai possibilitar no futuro a construção de um mapa genético integrado (GARCIA et al., 2006). No total foi possível mostrar a colineariedade de sete diferentes grupos de ligação entre os dois mapas genéticos, sendo que em duas situações houve colineariedade de dois GLs de Mombaça para um GL de S10 (S10-GL6: Mombaça-GL6 e GL8; S10-GL7: Mombaça-GL7 e GL9)

(Figura 5). Um total de quatro GLs de S10 não pôde ser correlacionado a outros quatro GLs de Mombaça. Com uma maior saturação e obtenção de marcadores em comum, esperamos que estes grupos possam ser integrados aos restantes. Entre os 38 marcadores em comum, 29 (76,3%) mantiveram a mesma ordem quando comparados entre os mapas de S10 e Mombaça. Este valor foi maior do que obtido no mapa de ligação da batata (BRADSHAW et al., 2008) mas menor do que observado em mapa genético de *R. hybrida* (GAR et al., 2011). Ainda, 37 (97,3%) marcadores em comum foram relacionados ao mesmo grupo de ligação em ambos genitores, quando considerados que GL8 e GL9 de Mombaça como pertencentes a GL6 e GL7 de S10 respectivamente, conforme demonstrado na Figura 5. Considerando que a ordem na maioria dos marcadores em comum foi similar nos dois mapas, é possível pressupor uma maior confiança na ordem estabelecida para os mapas de ligação de *P. maximum* construídos neste trabalho. Contudo, a ordenação das marcas mostrará muito mais consistência e será melhorada com o aumento na densidade dos marcadores moleculares. Uma amostra de 19 locos provenientes do transcriptoma da espécie (EST-SSR) (TOLEDO-SILVA et al., 2013) foram avaliados na população, visando integração ao mapa de ligação. Destes 19 locos, 12 foram avaliados como polimórficos e 6 foram integrados com sucesso aos mapas de ligação. Estes marcadores são derivados de regiões genômicas transcritas e, portanto, mais conservadas.

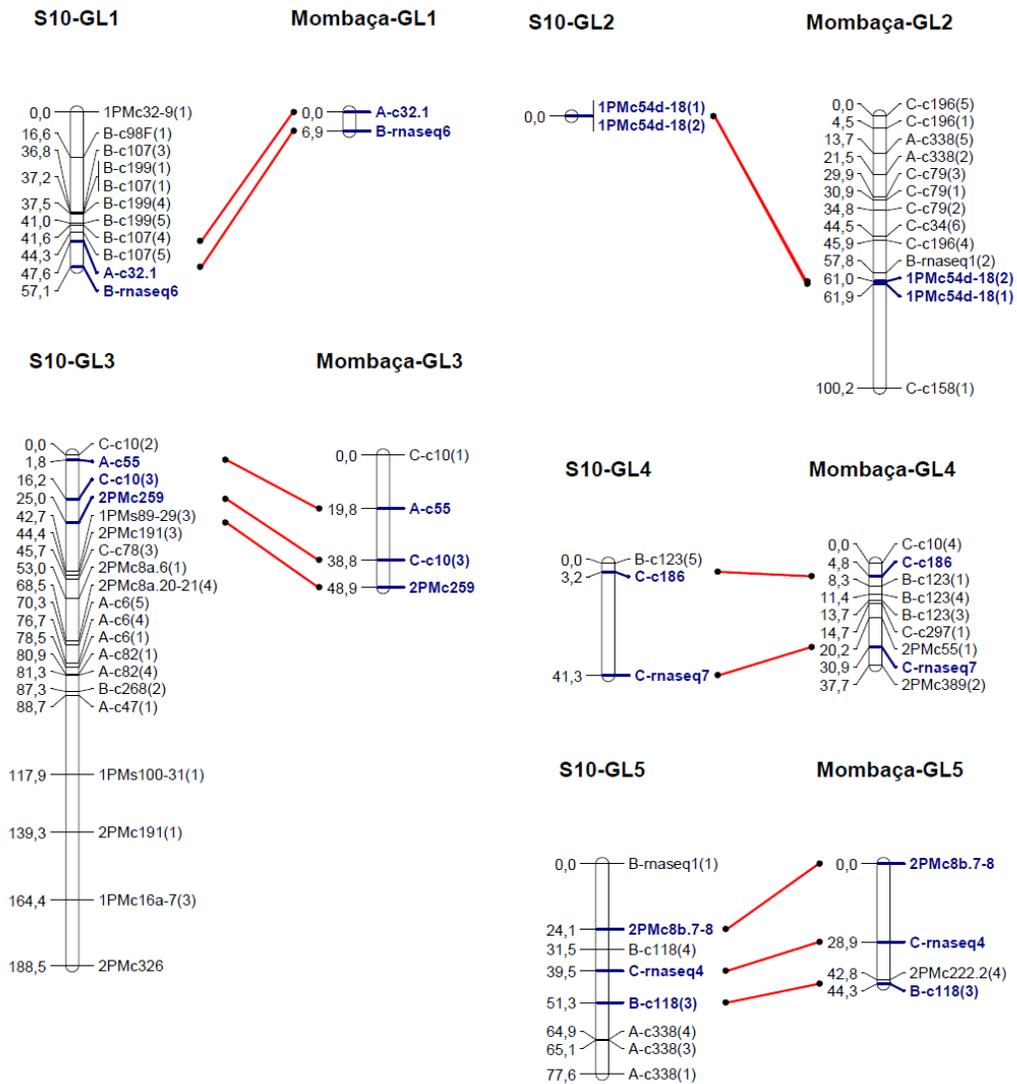


Figura 4: Ligação entre os mapas dos genitores S10 e Mombaça através de marcadores SSR em comum.

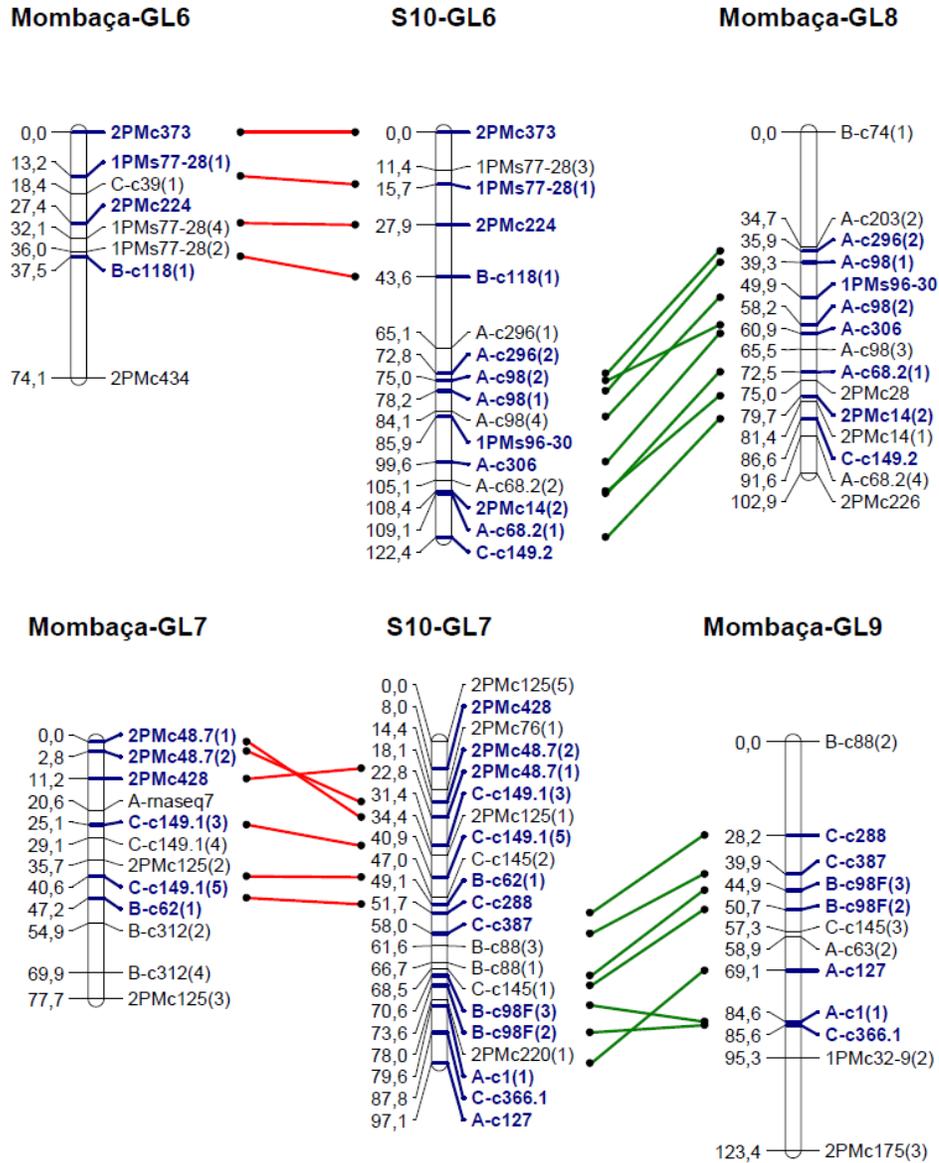


Figura 5: Ligação entre os mapas dos genitores S10 e Mombaça através de marcadores SSR em comum.

Desta forma, são mais prováveis de serem transferidos com sucesso para outras populações de mapeamento e espécies de gramíneas (SCOLES et al., 2009; VARSHNEY, GRANER e SORRELLS, 2005), além de serem recomendados como loci de ancoragem intra e interespecífico para estudos filogenéticos entre diferentes espécies (MIAN et al., 2005). Os EST-SSRs são considerados importantes para os mapas de ligação, visto que podem ser associados funcionalmente a variações de características de interesse (STUDER et al., 2010). Estes resultados também serviram como validação de alguns marcadores putativos encontrados no transcriptoma da espécie (TOLEDO-SILVA et al., 2013).

Em comparação com o único mapa genético disponível de *P. maximum* (EBINA et al., 2005), houve diferenças esperadas devido às metodologias e tipos de marcadores envolvidos. O mapa anterior utilizou uma combinação de marcadores AFLP e RAPD para gerar um mapa genético de apenas um genitor de uma população F1-segregante. Foram utilizados apenas marcadores em dose única, seguindo metodologia de WU et al. (1992). O trabalho encontrou 39 grupos de ligação, com uma maior cobertura de 1706.5 cM em 39 GLs, e densidade média também foi maior, com um marcador a cada 4.7 cM. O trabalho conseguiu mapear uma provável região ligada a característica reprodutiva de apomixia, algo que será realizado também no mapa gerado neste presente trabalho. Apesar dos resultados positivos, o mapa de EBINA et al. (2005) tem sua utilização restrita devido aos tipos de marcadores utilizados, de difícil reprodutibilidade e transferibilidade. A ausência de marcadores SSRs impediu uma comparação entre os mapas gerados neste trabalho e o mapa desenvolvido por EBINA et al. (2005), e uma possível integração e construção de mapa de ligação sintético. Ainda, nenhuma informação sobre cromossomos homólogos foi gerada, assim como nenhuma forma de validação foi possível devido a construção do mapa utilizando apenas um genitor no trabalho anteriormente publicado.

Segundo RIPOL e colaboradores (1999), a problemática envolvida na construção de mapas genéticos em poliploides é referente à ordenação dos locos ao longo dos cromossomos individuais, e à definição destes cromossomos a um grupo de homologia. O primeiro problema pode ser resolvido com a utilização de algoritmos de ordenação apropriados, também representando um problema comum em mapas de ligação de espécies diploides (MOLLINARI, 2007), contudo o segundo problema é único de poliploides (GAR et al., 2011). Neste trabalho nós conseguimos superar de forma satisfatória o problema referente a ordenação dos GL (Figuras 2 e 3) e a definição dos cromossomos homólogos (Figuras Suplementares 1 e 2).

O mapeamento completo dos 32 cromossomos da forrageira tropical *P. maximum* é uma etapa crucial para a análise de QTLs, que será obtido com a maior saturação do mapa em conjunto com dados fenotípicos. O mapa parcial que apresentamos neste trabalho, portanto, é um importante passo dado na compreensão de como os alelos interagem geneticamente para exibir diferentes fenótipos de modo que dois mapas parciais de *P. maximum* foram obtidos. O mapa do genitor S10 foi formado por 10 grupos de ligação, dos quais foram gerados 31 cromossomos homólogos. O mapa da cultivar Mombaça foi formado por 13 grupos de ligação, reconstruindo 36 cromossomos homólogos. A principal contribuição em relação ao mapa genético de *P. maximum* existente (EBINA et al., 2005) reside no fato da utilização de uma metodologia de mapeamento apropriada a autotetraploides, na qual marcadores SSRs são interpretados como codominantes, proporcionando grande vantagem em relação aos marcadores dominantes em dose única comumente utilizados em mapas de espécies poliploides. Ainda, o mapa aqui apresentado foi construído a partir de genitores atualmente utilizados no programa de melhoramento genético de *P. maximum*, na Embrapa Gado de Corte, garantindo futuras aplicações do conhecimento gerado. Estes mapas serão saturados com marcadores do tipo SNPs, que serão desenvolvidos a partir de sequências do transcriptoma de *P. maximum* (TOLEDO-SILVA et al., 2013) e genotipados através de espectrometria de massa

utilizando tecnologia Sequenom Mass Array e o método implementado no programa SuperMASSA (SERANG, MOLLINARI e GARCIA, 2012) para análise dos dados. Pretendemos também aplicar a metodologia de genotipagem-por-sequenciamento (*genotyping-by-sequencing* ou GBS) (ELSHIRE et al., 2011) nos genitores e nas progênes da população, identificando e genotipando milhares de locos SNPs. A utilização de marcadores do tipo SNP pode trazer grandes contribuições ao mapa genético, dada sua co-dominância, mesmo em poliplóides, e a possibilidade de utilizar múltiplas doses. Estudos complementares de citogenética, como hibridização *in situ*, também são necessários pois permitiriam o estudo da evolução do genoma da espécie. Os genitores do cruzamento são contrastantes para caracteres produtivos, e ainda se diferenciam quanto ao modo de reprodução, permitindo estudos futuros de QTLs ligados a estas características. O mapa de ligação gerado neste estudo traz informações importantes acerca do genoma da espécie, podendo ser usado como base para estudos básicos e aplicados em *Panicum*.

Referências bibliográficas

- BETHEL, C. M.; SCIARA, E. B.; ESTILL, J. C.; BOWERS, J. E.; HANNA, W. e PATERSON, A. H. A framework linkage map of bermudagrass (*Cynodon dactylon transvaalensis*) based on single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 112, n. 4, p. 727–737, 2006.
- BRADSHAW, J. E.; HACKETT, C. A.; PANDE, B.; WAUGH, R. e BRYAN, G. J. QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). **Theoretical and Applied Genetics**, v. 116, n. 2, p. 193–211, 2008.
- BRAZ, T. **Parâmetros genéticos, repetibilidade e diversidade genética em híbridos de *Panicum maximum***. Tese (Doutorado) - Universidade Federal de Lavras, 2012.
- CRESTE, S.; NETO, A. T. e FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v. 19, n. 4, p. 299–306, 2001.
- EBINA, M.; NAKAGAWA, H.; YAMAMOTO, T.; ARAYA, H.; TSURUTA, S. I.; TAKAHARA, M. e NAKAJIMA, K. Co-segregation of AFLP and RAPD markers to apospory in Guineagrass (*Panicum maximum* Jacq.). **Grassland Science**, v. 51, n. 1, p. 71–78, 2005.
- ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S. e MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.
- GAR, O.; SARGENT, D. J.; TSAI, C. J.; PLEBAN, T.; SHALEV, G.; BYRNE, D. H. e ZAMIR, D. An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. **PloS one**, v. 6, n. 5, p. e20463, 2011.
- GARCIA, A. A. F.; KIDO, E. A.; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; DA SILVA, J. A. G.; ULIAN, E. C. e FIGUEIRA, A. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, v. 112, n. 2, p. 298–314, 2006.
- HACKETT, C. A.; BRADSHAW, J. E.; MEYER, R. C.; MCNICOL, J. W.; MILBOURNE, D. e WAUGH, R. Linkage analysis in tetraploid species: a simulation study. **Genetical research**, v. 71, n. 02, p. 143–153, 1998.
- HACKETT, C. A.; MILNE, I.; BRADSHAW, J. E. e LUO, Z. TetraploidMap for Windows: linkage map construction and QTL mapping in autotetraploid species. **The Journal of heredity**, v. 98, n. 7, p. 727–9, 2007.
- HACKETT, C. A. e LUO, Z. W. TetraploidMap: construction of a linkage map in autotetraploid species. **Journal of Heredity**, v. 94, n. 4, p. 358–359, 2003.

- HACKETT, C. A.; PANDE, B. e BRYAN, G. J. Constructing linkage maps in autotetraploid species using simulated annealing. **Theoretical and Applied Genetics**, v. 106, n. 6, p. 1107–1115, 2003.
- JANK, L.; RESENDE, R. M. S.; VALLE, C. B.; RESENDE, M. D. V.; CHIARI, L.; CANÇADO, L. J. e SIMONI, C. Melhoramento genético de *Panicum maximum* Jacq. In: **Melhoramento de forrageiras tropicais**. Campo Grande: Embrapa Gado de Corte, p. 55–87, 2008.
- JANK, L.; VALLE, C. B. e RESENDE, R. Breeding tropical forages. **Crop Breeding and Applied Biotechnology**, v. 11, n. spe, p. 27–34, 2011.
- JESSUP, R. W.; BURSON, B. L.; BUROW, G.; WANG, Y-W.; CHANG, C. LI, Z.; PATERSON, A. H. e HUSSEY, M. A. Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping. **Genome**, v. 46, n. 2, p. 304–313, 2003.
- JULIER, B.; FLAJOULOT, S.; BARRE, P.; CARDINET, G.; SANTONI, S.; HUGUET, T. e HUYGHE, C. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. **BMC plant biology**, v. 3, p. 9, 2003.
- KALENDAR, R.; LEE, D. e SCHULMAN, A. H. FastPCR software for PCR primer and probe design and repeat search. **Focus on Bioinformatics/Mansour, A.(Ed)**, 2009.
- LUO, Z. W.; HACKETT, C. A.; BRADSHAW, J. E.; MCNICOL, J. W. e MILBOURNE, D. Predicting parental genotypes and gene segregation for tetrasomic inheritance. **Theoretical and Applied Genetics**, v. 100, n. 7, p. 1067–1073, 2000.
- LUO, Z. W.; HACKETT, C. A.; BRADSHAW, J. E.; MCNICOL, J. W. e MILBOURNE, D. Construction of a genetic linkage map in tetraploid species using molecular markers. **Genetics**, v. 157, n. 3, p. 1369–1385, 2001.
- MALIEPAARD, C.; JANSEN, J. e OOIJEN, J. W. V. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, v. 70, n. 3, p. 237–250, 1997.
- MEYER, R. C.; MILBOURNE, D.; HACKETT, C. A.; BRADSHAW, J. E.; MCNICHOL, J. W. e WAUGH, R. Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). **Molecular & general genetics**, v. 259, n. 2, p. 150–60, 1998.
- MIAN, M. A. R.; SAHA, M. C.; HOPKINS, A. A.; WANG, Z-Y. Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. **Genome**, v. 48, n. 4, p. 637–647, 2005.
- MOLLINARI, M. **Comparação de algoritmos usados na construção de mapas genéticos**. Dissertação (Mestrado) - Universidade Estadual de São Paulo, 2007.
- MUIR, J. P. e JANK, L. Guineagrass. In: **Warm-season (C4) grasses**, p. 589–621, 2004.

- NAKAJIMA, K.; KOMATSU, T.; MOCHIZUKI, N. e SUZUKI, S. Isolation of diploid and tetraploid sexual plants in guinea grass (*Panicum maximum* Jacq.). **Japanese Journal of Breeding**, v. 29, 1979.
- OKADA, M.; LANZATELLA, C.; SAHA, M. C.; BOUTON, J.; WU, R. e TOBIAS, C. M. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. **Genetics**, v. 185, n. 3, p. 745–60, 2010.
- OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MARGARIDO, G. R. A.; PASTINA, M. M.; TEIXEIRA, L. H. M.; FIGUEIRA, A. V.; ULIAN, E. C.; GARCIA, A. A. F. e SOUZA, A. P. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, v. 20, n. 3, p. 189–208, 2007.
- PORCEDDU, A.; ALBERTINI, E.; BARCACCIA, G.; FALISTOCCO, E.; FALCINELLI, M. Linkage mapping in apomictic and sexual Kentucky bluegrass (*Poa pratensis* L.) genotypes using a two way pseudo-testercross strategy based on AFLP and SAMPL markers. **Theoretical and Applied Genetics**, v. 104, n. 2-3, p. 273–280, 2002.
- RIPOL, M. I.; CHURCHILL, G. A.; DA SILVA, J. A. G. e SORRELLS, M. Statistical aspects of genetic mapping in autopolyploids. **Gene**, v. 235, n. 1, p. 31–41, 1999.
- SAHA, M. C.; MIAN, R.; ZWONITZER, J. C.; CHEKHOVSKIY, K. e HOPKINS, A. A. An SSR-and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). **Theoretical and applied genetics**, v. 110, n. 2, p. 323–336, 2005.
- SCHUELKE, M. An economic method for the fluorescent labeling of PCR fragments. **Nature biotechnology**, v. 18, n. 2, p. 233–4, doi:10.1038/72708, 2000.
- SCOLES, G.; SIM, S-C.; YU, J-K.; JO, Y-K.; SORRELLS, M. E. e JUNG, G. Transferability of cereal EST-SSR markers to ryegrass. **Genome**, v. 52, n. 5, p. 431–437, 2009.
- SERANG, O.; MOLLINARI, M. e GARCIA, A. A. F. Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. **PLoS one**, v. 7, n. 2, p. e30906, 2012.
- SOUSA, A. C. B.; JUNGSMANN, L.; CAMPOS, T.; SFORÇA, D. A.; BOAVENTURA, L. R.; SILVA, G. M. B.; ZUCCHI, M. I.; JANK, L. e SOUZA, A. P. Development of microsatellite markers in Guineagrass (*Panicum maximum* Jacq.) and their transferability to other tropical forage grass species. **Plant Breeding**, v. 130, n. 1, p. 104–108, 2011a.
- SOUSA, A. C. B.; JANK, L.; CAMPOS, T.; SFORÇA, D. A.; ZUCCHI, M. I. e SOUZA, A. P. Molecular Diversity and Genetic Structure of Guineagrass (*Panicum maximum* Jacq.), a Tropical Pasture Grass. **Tropical Plant Biology**, v. 4, n. 3-4, p. 185–202, 2011b.
- STEIN, J.; PESSINO, S. C.; MARTÍNEZ, E. J.; RODRIGUEZ, M. P.; SIENA, L. A.; QUARIN, C. L. e ORTIZ, J. P. A. A genetic map of tetraploid *Paspalum notatum* Flüge (bahia grass) based on single-dose molecular markers. **Molecular Breeding**, v. 20, n. 2, p. 153–166, 2007.

- STUDER, B.; KÖLLIKER, R.; MUYLLE, H.; ASP, T.; FREI, U. e ROLDÁN-RUIZ, I. EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). **BMC plant biology**, v. 10, p. 177, 2010.
- TOLEDO-SILVA, G.; CARDOSO-SILVA, C. B.; JANK, L. e SOUZA, A. P. De novo transcriptome assembly for the tropical grass *Panicum maximum* Jacq. **PloS one**, v. 8, n. 7, p. e70781, 2013.
- VARSHNEY, R. K.; GRANER, A. e SORRELLS, M. E. Genic microsatellite markers in plants: features and applications. **Trends in biotechnology**, v. 23, n. 1, p. 48–55, 2005.
- VIGNA, B. B. Z. **Estudos genético-moleculares em *Brachiaria humidicola* (Rendle) Schweick. (Poaceae)**. Tese (Doutorado) - Universidade Estadual de Campinas, 2010.
- VOORRIPS, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. **Journal of Heredity**, v. 93, n. 1, p. 77–78, 2002.
- WARMKE, H. E. Apomixis in *Panicum maximum*. **American Journal of Botany**, p. 5–11, 1954.
- WU, K. K.; BURNQUIST, W.; SORRELLS, M. E.; TEW, T. L.; MOORE, P. H. e TANKSLEY, S. D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 83, n. 3, p. 294–300, 1992.
- YOU, F. M.; HUO, N.; GU, Y. Q.; LUO, M-C.; MA, Y.; HANE, D.; LAZO, G. R.; DVORAK, J. e ANDERSON, O. D. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. **BMC bioinformatics**, v. 9, p. 253, 2008.

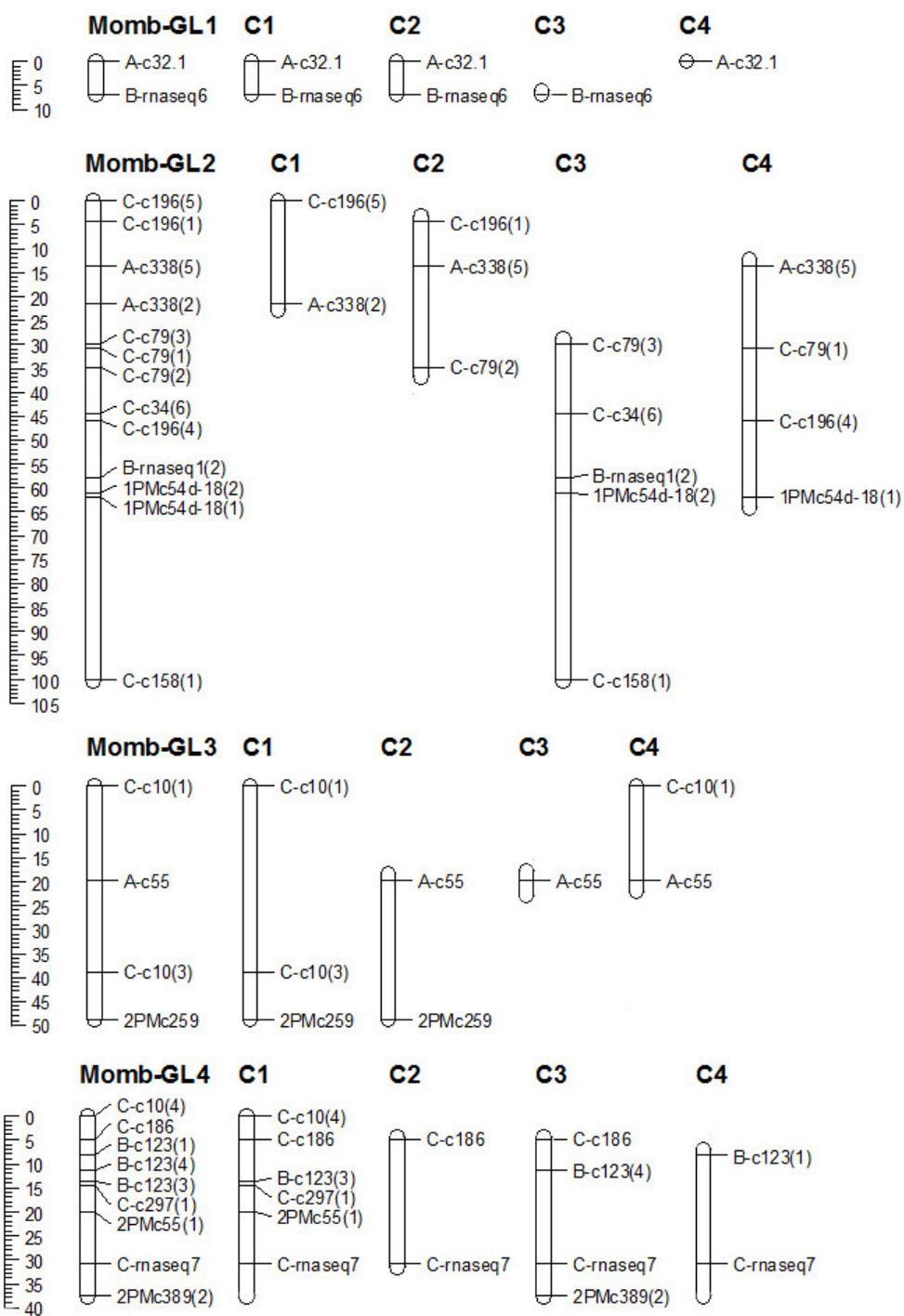


Figura Suplementar 1: Representação dos cromossomos homólogos reconstruídos a partir de cada grupo de ligação do genitor Mombaca. A distância em cM é apresentada na régua a esquerda de cada grupo. As siglas C1, C2, C3 e C4 correspondem aos cromossomos hipotéticos de cada grupo de ligação.

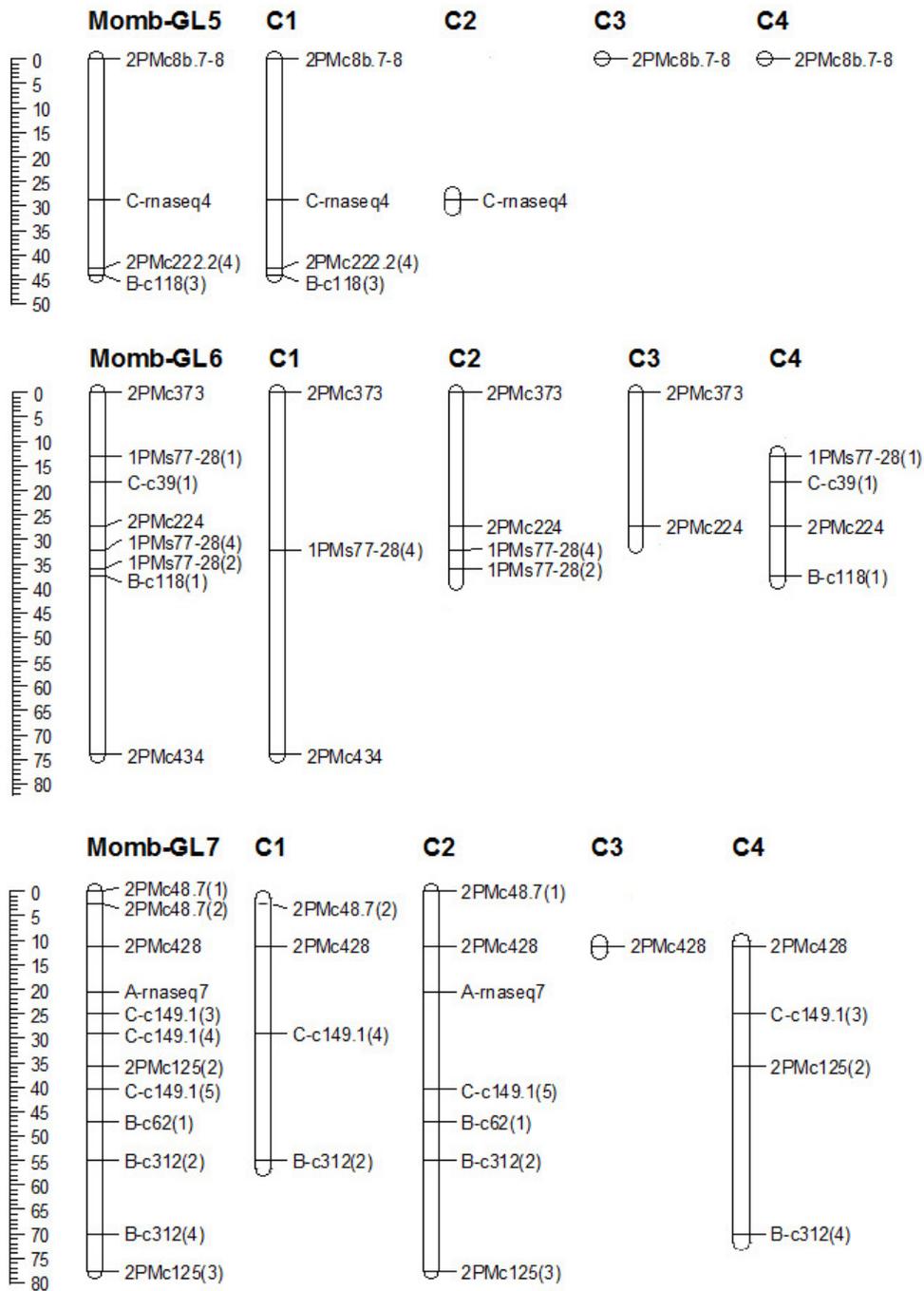


Figura Suplementar 1: continuação.

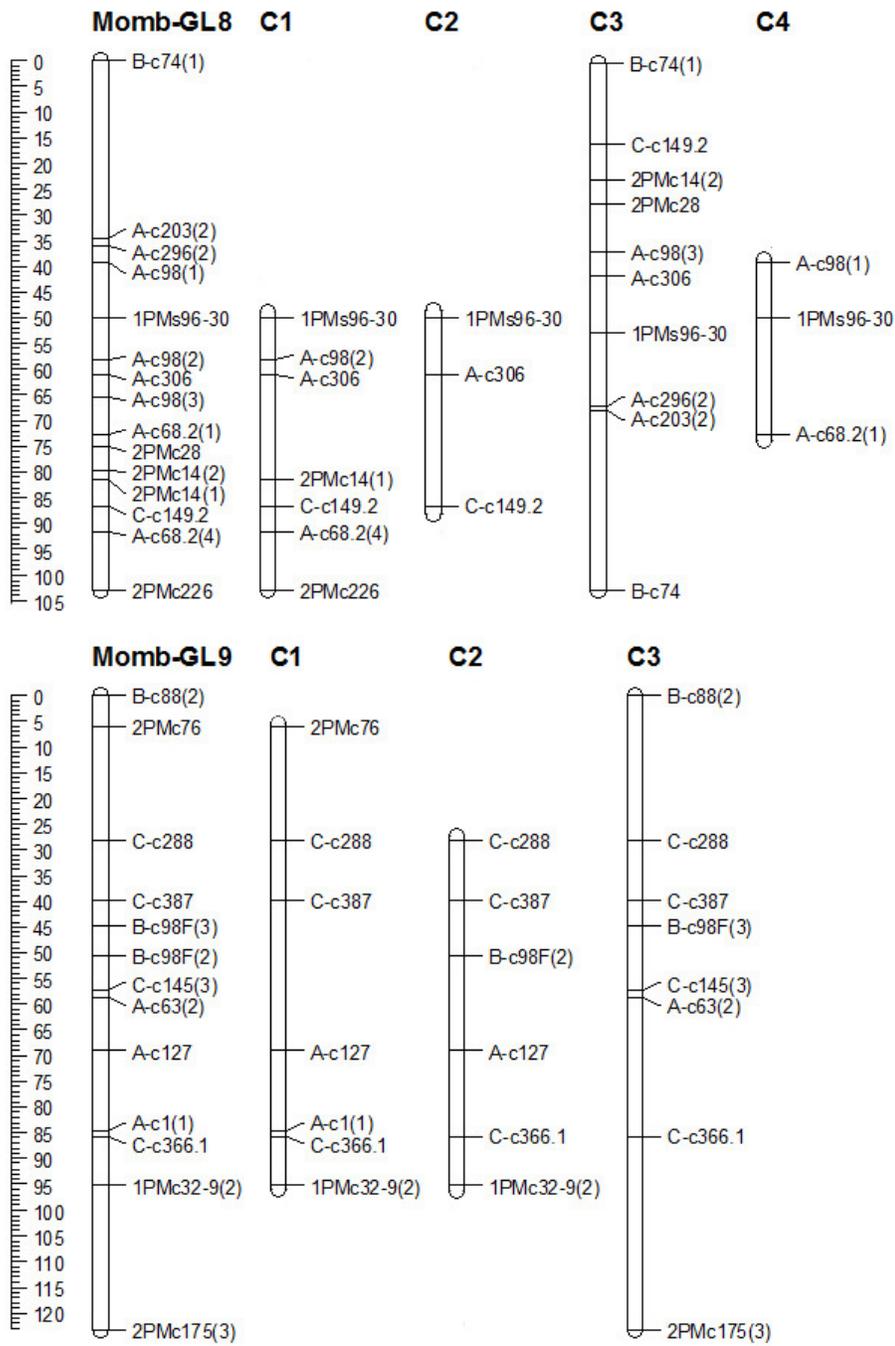


Figura Suplementar 1: continuação.

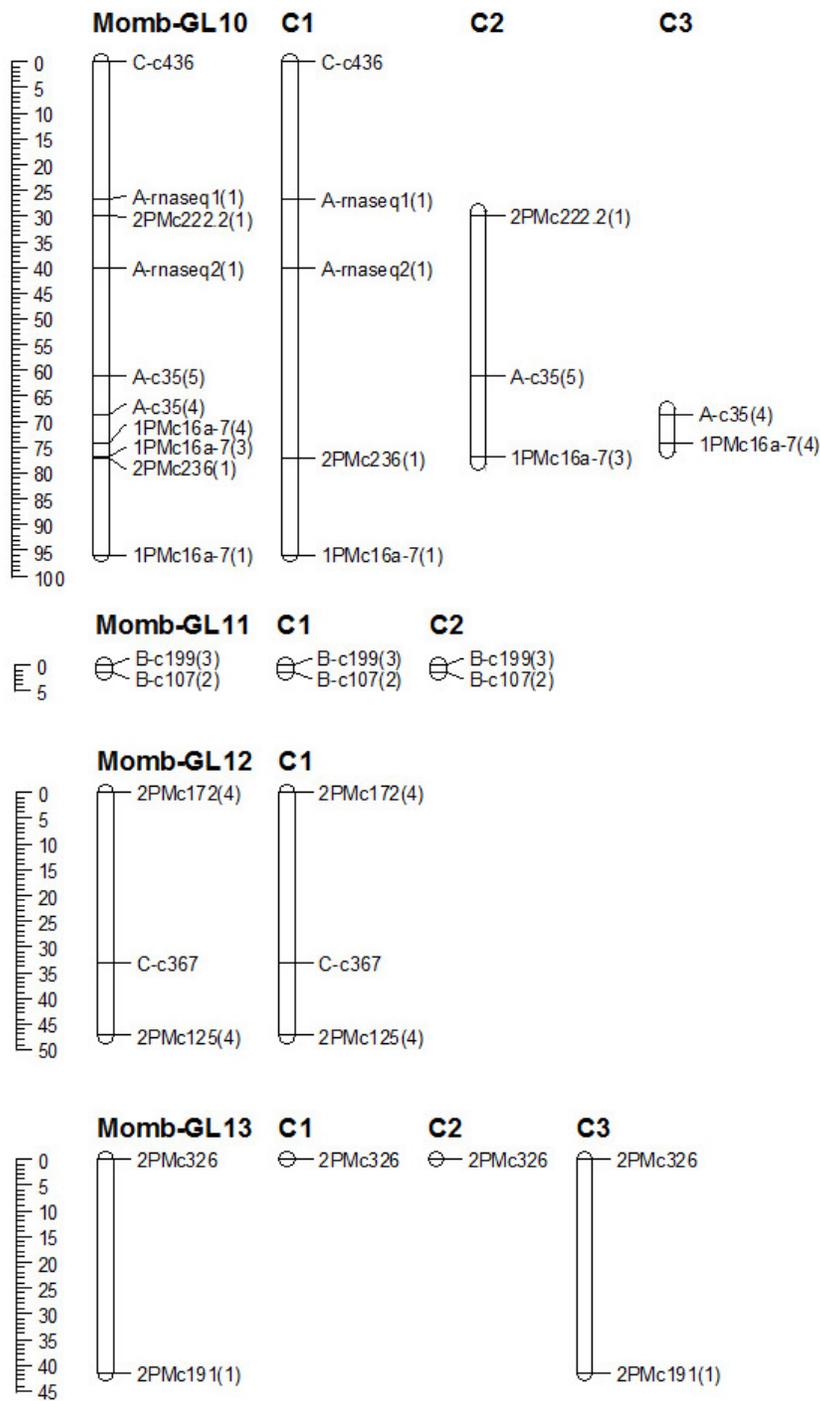


Figura Suplementar 1: continuação.

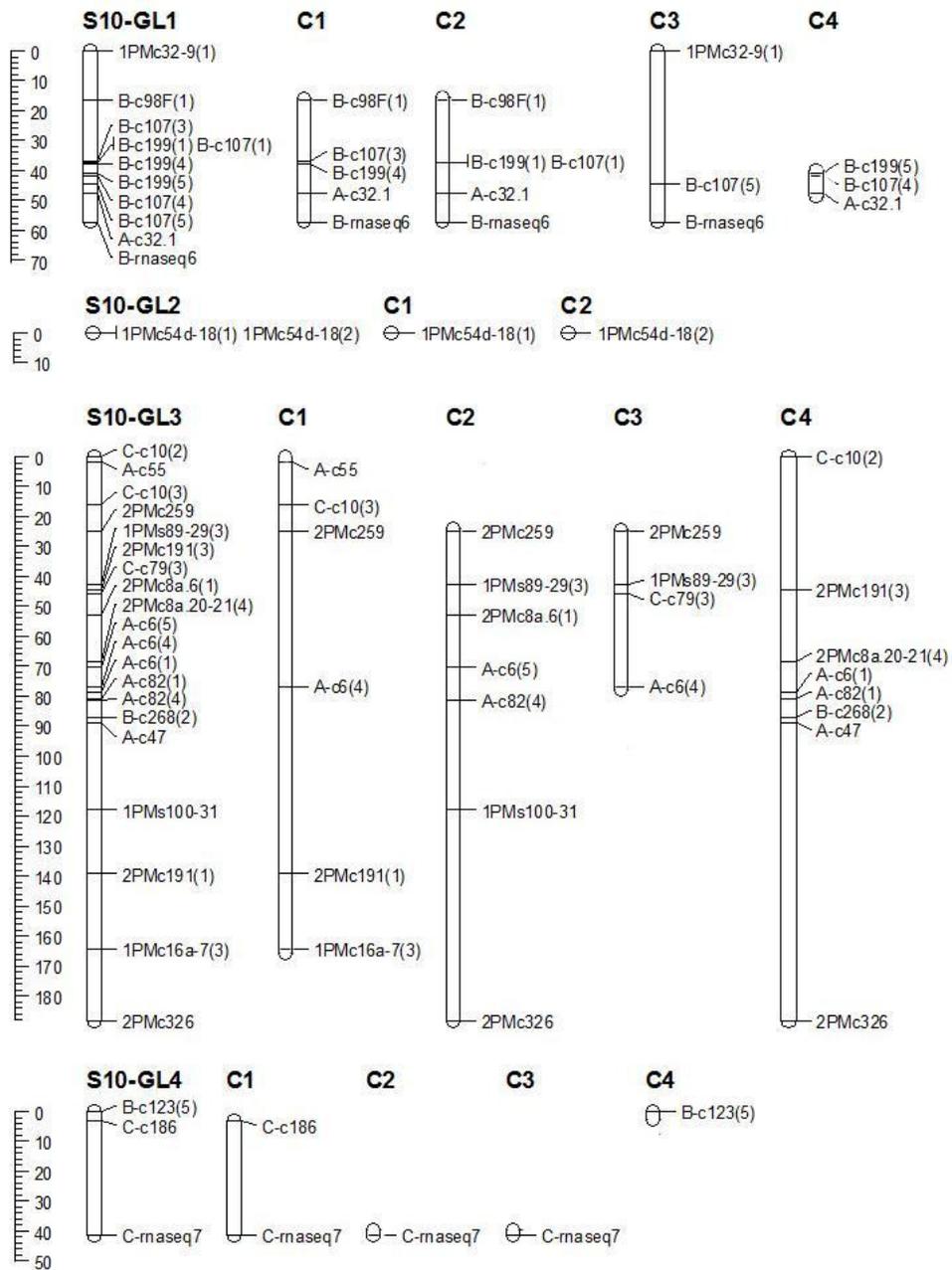


Figura Suplementar 2: Representação dos cromossomos homólogos reconstruídos a partir de cada grupo de ligação do genitor S10. A distância em cM é apresentada na régua a esquerda de cada grupo. As siglas C1, C2, C3 e C4 correspondem aos cromossomos hipotéticos de cada grupo de ligação.

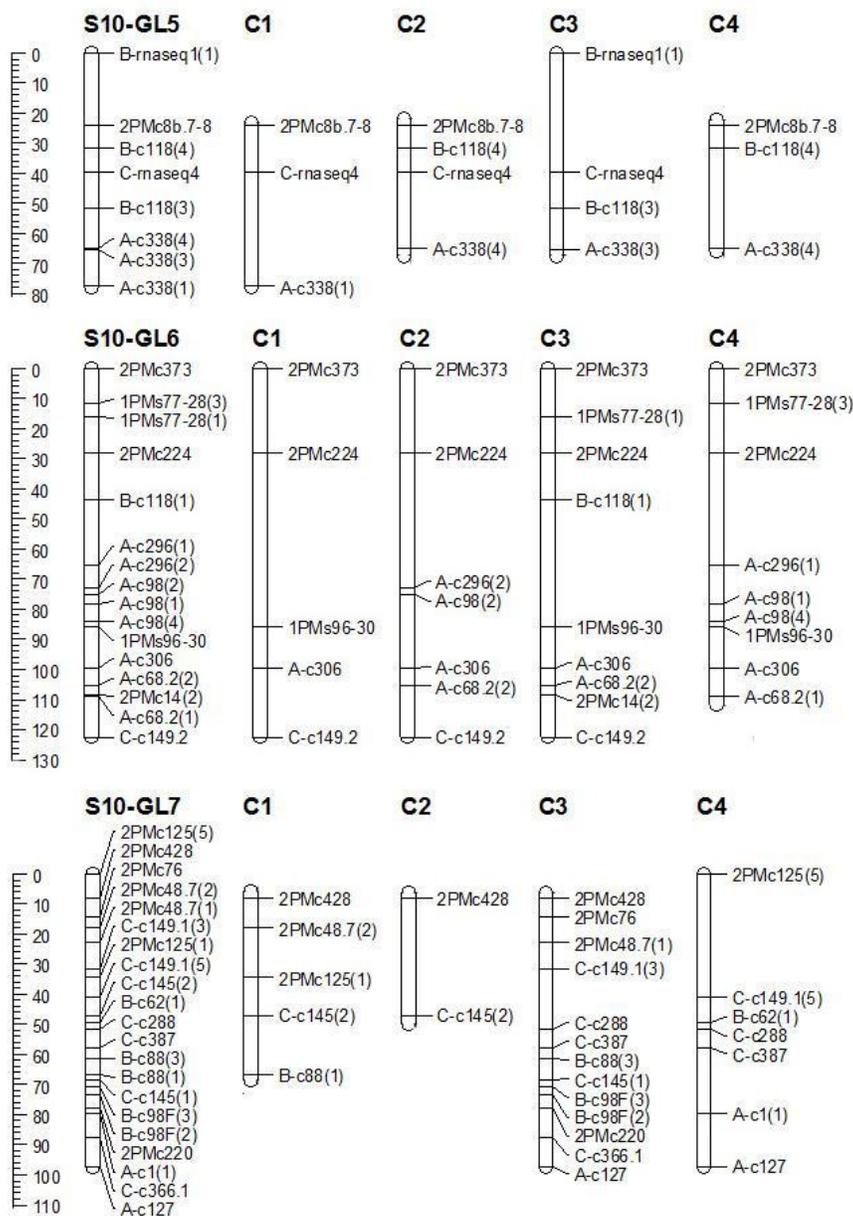


Figura Suplementar 2: continuação.

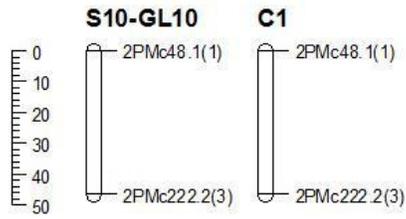
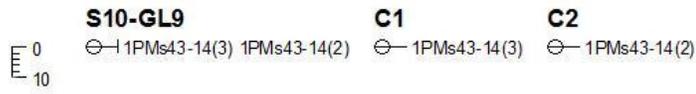
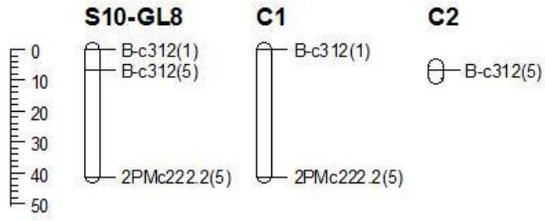


Figura Suplementar 2: continuação. Tabela Suplementar 1: Tabela informativa acerca de todos os marcadores utilizados com amplificação bem sucedida, com os respectivos motivos microssatélites, tamanho esperado (pb), temperatura de anelamento em °C (Ta), número de alelos encontrados, sequência dos iniciadores, origem do marcador, existência de polimorfismo, técnica de genotipagem, tipo de dominância envolvida e ligação ao mapa genético.

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
1PMc16a-7	(TG)13	60	234	4	CACGACGTTGTAAAACGAGGCTTGTTCCTCCATAG TCCATAAAGTAGGTAATCAT	Não-publicado	P	A	D	L
1PMc2.1-3	(CGC)3	60	257	1	GACCTTTTCTCCT GGGTATCGTATTTTA	Não-publicado	M	B	-	NL
1PMc21-8	(TC)21	55	180	3	ATCATTACTGTTCCTTCATACC ATTCATTTGGCGTCATTT	Não-publicado	P	B	-	NL
1PMc32-9	(TG)6	55	278	3	AACAGTTTGCAGATGGTAG TTGAGGATTAATGAGAAGTC	Publicado ¹	P	B	D	L
1PMc45d-15	(GT)8	50	202	1	CAAATACCCAAGTTACAATG GCCAGAAGCAGTGATGA	Não-publicado	M	B	-	NL
1PMc48-16	(CA)6	60	240	1	GAAATCTGGTGCATCATCCT ATCCGGTTGGCAGAGATATC	Publicado ¹	M	B	-	NL
1PMc54d-18	(GT)5	55	336	2	TATATTGTTTTACCTGAGAA AAGAAGTTTATTTGTTTATC	Não-publicado	P	B	D	L
1PMc55-19	(GT)13	55	160	1	TCCTCTAGAACCAAGCACA ATCAAGACACATCAAGAACACAT	Publicado ²	M	B	-	NL
1PMc61.1-21	(GA)11	50	213	1	AGAGTAGAACCAATG GGACAATAATGAACAA	Não-publicado	M	B	-	NL
1PMc72-27	(CA)6	60	195	1	GAAATCCGCCTCCACCAA TCCGGCGCCACTTCAT	Publicado ²	M	B	-	NL
1PMs100-31	(TG)7	55	154	2	TGTTAAGGGCAGCATCT CCGAACGAGCTGGACATT	Não-publicado	P	B	D	L
1PMs35.1-11	(TG)6	55	198	1	TACACTACGCCATTTTG CTAATAGCTTCCTCAGTAATAG	Publicado ²	M	B	-	NL
1PMs35.2-12	(TG)6	60	208	1	AATTTTGTTATCCTGCTCCAC ACCCAAAGATAATTAGAACCTG	Não-publicado	M	B	-	NL
1PMs43-14	(TC)5	55	200	3	ATGAAGCGGGCGGTAGTATT TGGTGGCGGTAAGAGTAAAG	Publicado ¹	P	B	D	L
1PMs60-20	(GC)3(A)7	55	257	1	ACATTAAGTTGCTGAAG TAAAGTACTGAGGTCGTC	Não-publicado	M	B	-	NL
1PMs77-28	(AC)7	60	180	4	CACGACGTTGTAAAACGAGTTCATGTTAAATCTGTGTTA GGCCAAGCTTATTTATTCTA	Não-publicado	P	A	D	L
1PMs89-29	(CA)8	55	181	4	GGTGTGCAGCCTATCC GACCCAGCCATTTT	Não-publicado	P	B	D	L
1PMs96-30	(CA)5A(CA)2	50	252	4	CACGACGTTGTAAAACGAACAAAGATGGGGCGTGAAGAC CTAGGTAGGCCGACAACAATGA	Publicado ²	P	A	CD	L
2PMc118	(CA)5	60	189	1	GGCCAGAATG CGCTACAACGAG	Não-publicado	M	B	-	NL
2PMc125	(TGT)9	60	268	5	CTTTGCCATCTCTTATCCACA GCTGCCCATTTCTTAGTCCA	Publicado ²	P	B	D	L
2PMc14	(CA)7	60	190	3	AGAGACAGGCAGGAG CGGTGTATGAAATGAA	Não-publicado	P	B	D	L
2PMc14b	(GT)7	55	190	1	CAGCTCCGTCCCGTATCTCTAA CCGCAGGAAGCACTATGGT	Publicado ¹	M	B	-	NL

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
2PMc15	(CA)5	60	297	1	AAAAATAAAAGGAGT TGAAGCAAGAATAGTA	Não-publicado	M	B	-	NL
2PMc151.1	(TA)4	50	146	1	TCGATGACAGG CAAACCACACTA	Não-publicado	M	B	-	NL
2PMc158	(TC)6(CA)7	55	227	1	GGAATAGCCCCAGATA GGCTACCTTCATTGTTT	Publicado ²	M	B	-	NL
2PMc168	(TG)5	60	213	1	CCTCGCATTTTTCTGGATTTA CATAGACGCACGCACACTCAC	Publicado ²	M	B	-	NL
2PMc169	(AC)7	60	166	1	TGGGGTGCAGGATAATGGT TCGAGGCTGAGTTCTTGGATA	Não-publicado	M	B	-	NL
2PMc172	(CT)20	60	194	4	AAGCTAGCAGTTTGAT CGTAGGTATTGGAGTG	Publicado ²	P	B	D	L
2PMc175	(CA)6	60	243	3	TTCACGGTCAGATTCA TGCAGCTCATTGTTT	Publicado ²	P	B	D	L
2PMc178	(CA)6	55	226	1	ACCTGCTTGTTTTGCTTGTGTTG AGGCTGGCTCTGATTGG	Publicado ²	M	B	-	NL
2PMc181	(CTC)5	55	197	1	TTTCTCCTCAAG CAGAATCCACTC	Não-publicado	M	B	-	NL
2PMc19	(AC)9	50	248	1	ATGGTTAAAGATGTTGTGAGTG GAGGCTGAGTTCTTGGATAG	Publicado ¹	M	B	-	NL
2PMc191	(GT)5	60	125	4	ACACTCAACTGGCAACCT TTCACAATTTCTTTTCACAC	Publicado ²	P	B	D	L
2PMc198	(TC)3TGA(CT)11	60	254	1	CACGACGTTGTAAAAACGAACTAGAGCAATTCAGGTGTT AAGACGCATAAAGAAAAATA	Não-publicado	M	A	-	NL
2PMc198.1	(CT)7	55	255	1	CAGAAAGAAGGAAAGGAAAGGAA TCTAGCTGCATGCATAAACACT	Publicado ¹	M	B	-	NL
2PMc220	(GT)6	50	227	2	GCTCCTGCTAGTTGTGTA AGTCTTTTGCAATTATTGG	Não-publicado	P	B	D	L
2PMc222.2	(CT)8	60	210	5	CACGACGTTGTAAAAACGATAATCCCTTTGCCTTTTC CCATTCATCTGCTGCTAC	Não-publicado	P	A	D	L
2PMc224	(CA)5	60	207	4	CACGACGTTGTAAAAACGAAAGGGGTGGGGTTCC TGGTCATGGTGGGTTC	Publicado ²	P	A	CD	L
2PMc226	(AC)5	55	340	2	TCGAATATGCAGGAGAAC GTAAAGTGTAACCCAGTG	Publicado ²	P	B	D	L
2PMc236	(TG)7	60	214	3	AACGTGCACATATTTCACTTT GGATCCCTCTTCATTAG	Publicado ¹	P	B	D	L
2PMc239.1	(GA)8	60	216	1	TAACAAGAGAAAATAAACAA GGAGTAAAGGACCAC	Publicado ²	M	B	-	NL
2PMc239.2	(GT)6	60	251	1	CAAACAATTCAGCACAAACTA CTATGGAGAAATTAAGCAACCT	Publicado ¹	M	B	-	NL
2PMc259	(GT)8	55	207	3	CACGACGTTGTAAAAACGAGGGCCGGGGAGTAGGTTAG GGTTCAGGGGGAGAGCAGAGTG	Publicado ²	P	A	CD	L
2PMc261	(GT)6	55	205	1	TTGTATTGAAACCA TCCTATACCCACAT	Não-publicado	M	B	-	NL
2PMc28	(AC)6	60	241	2	CACGACGTTGTAAAAACGAAACCCGCGCATTACTACA ATGGTTGCAGAGAAGAGATGAC	Publicado	P	A	D	L
2PMc326	(TC)18	60	251	4	CAATTCGTCCTCGTCTA GGTTCCATGCACAAATAA	Não-publicado	P	B	CD	L
2PMc34	(TG)7	60	211	1	AGCACTGTGCACTAACCAATG CGTCTCCGTCCACCGATAG	Publicado	M	B	-	NL
2PMc342	(TG)7	60	240	1	CACGACGTTGTAAAAACGAGGGATCCGGTATGGTCA GTCGGTCTCGTGCTA	Publicado	M	A	-	NL

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
2PMc350	(GA)11	55	252	1	AACGACTCTCTGATT AGAACCTCGACAAG	Não-publicado	M	B	-	NL
2PMc373	(AG)9	60	244	5	CAGGCGAATGAGAGCAAAA TTAAAGGCATTCGGGTGCA	Publicado ²	P	B	CD	L
2PMc389	(CA)9	50	177	5	CAGGTAACATCACAAGTA CTATAGGTAAGCCAGTA	Publicado ²	P	B	D	L
2PMc399	(ACG)7	60	251	1	GAGAGCAGCAGTGGA GCAACGCATCTTTA	Não-publicado	M	B	-	NL
2PMc40	(TG)5CA(TG)4	60	254	1	ATATTTCTCGAGATTTGTGTT AAGCTTTGGGGATTAGTAGAA	Publicado ²	M	B	-	NL
2PMc401	(GT)5	50	199	1	GCAAAGGCAGTTCAC GCAAAGTTGGACACA	Não-publicado	M	B	-	NL
2PMc428	(CA)11	60	206	3	CTCTCAGTCCCACAGCACAC TATTTGGGGATTGGGAGTAGTT	Publicado ¹	P	B	CD	L
2PMc434	(CA)8	60	246	2	GTCCAAAGTGAACAA AAATCGGAGTAACAT	Não-publicado	P	B	D	L
2PMc47	(AT)4	55	190	1	CCTTCGGATTTACA AAGCGCAACACTAT	Não-publicado	M	B	-	NL
2PMc48.1	(CA)13	60	231	2	CCTGTCAAAAATAATGC GGGGAGACCTAACCA	Publicado ¹	P	B	D	L
2PMc48.7	(TG)13	60	208	2	GAACGGATCAGGGAGTAT TAAAGGGAAAGCAGAACA	Não-publicado	P	B	D	L
2PMc51	(GA)23	60	244	1	TCAGCAAGAAACATCCTCA TTCCATAACCCAAATCCTG	Publicado ²	M	B	-	NL
2PMc52	(TG)7	60	235	2	AGAATGGCACCTGGAGATAG GGATAGGCCGAAAGAACAT	Publicado ²	P	B	-	NL
2PMc55	(AC)10	60	220	3	GGTAGCGCTCTGTCTCTTG GACGGCCTTTCGCTTATTTT	Publicado ¹	P	B	D	L
2PMc60	(CA)8	55	237	2	ACAGTTAGCTTAGTGGTTG TATGAAGGAGTAAAAAGACA	Publicado ²	P	B	-	NL
2PMc62	(AG)10	55	228	1	TGCTGTTTCATACTCTCATT ACTGTCTGTTGCTTCACTG	Publicado ²	M	B	-	NL
2PMc7.12-15	(GT)7	55	269	1	TAAACTAGAGGACCCGTGTG TGTAGGCTCAAGAAAGGATT	Publicado ²	M	B	-	NL
2PMc73	(CA)5	60	233	1	TAGTTATGTCATTATTTAGCA AAGTCTTATTTAGTCATTTTG	Publicado ²	M	B	-	NL
2PMc76	(CA)3(CACG)4	60	188	2	GAGATTGTTTCGGTGGTCAG TATATGCATCCTACGCTTTTC	Não-publicado	P	B	D	L
2PMc8a.20-21	(TG)8	50	174	4	GTGCGTCTGTGAACCTGTGAT ACGATGCCGCCTCTTGTGTTGTT	Não-publicado	P	B	D	L
2PMc8a.28-29	(GC)3	55	251	3	CATGCGATACCTGCTGTG AAATGCGTTGAAAATAAAAAT	Não-publicado	P	B	-	NL
2PMc8a.6	(TG)8	55	266	4	GCGTTGCTGCATGCGATACCT GGGGACAAATGCGTTGAAATTAATAAATA	Publicado ¹	P	B	D	L
2PMc8b.7-8	(CATG)3	55	181	3	CTCGAACTGTAGCAACC TCCAACGAAAAGACCA	Não-publicado	P	B	CD	L
2PMc9.17	(CA)5	60	230	1	ATCAACGCTTTAATCCCTGTCC CATCGTCGTCCTCATCGTAGTC	Publicado ²	M	B	-	NL
2PMc90	(CA)8	60	178	1	AACGGTAGCTGGTGAAGA ATGTCGATGTGGCAAGTG	Publicado ¹	M	B	-	NL
2PMc96	(CA)7	50	237	2	TCCTCCCTTCTTTGTA TCCTTCAGGTCTCCAC	Publicado ²	P	B	-	NL

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
A-c1	(CT)17	55	114	2	CACGACGTTGTA AAAACGACAGCCATGCATGAGCATTGT CCCAGAGAAAAGAGAGGGACA	Não-publicado	P	A	D	L
A-c127	(TC)11	55	149	4	CACGACGTTGTA AAAACGAGGTGTGTGTCTTTTGGAC CCAGCGTGATGCATGTACTAA	Não-publicado	P	A	CD	L
A-c17.3	(CATG)3	60	146	1	CACGACGTTGTA AAAACGACTTTCTCATGGCAGCAGTACG CGATGCCTCAATGTGTGACG	Não-publicado	M	A	-	NL
A-c2	(CAT)4	60	118	1	CACGACGTTGTA AAAACGACTTTGTGTGATCACCATTGGA TGGACAATGAGTACGTGAACAA	Não-publicado	M	A	-	NL
A-c203	(AG)13	60	113	2	CACGACGTTGTA AAAACGATGCAGTTTCTACTTCTGCCTTG CCCACATTGGTTTGTGGTT	Não-publicado	P	A	D	L
A-c23	(TGA)5	60	100	1	CACGACGTTGTA AAAACGACTGGGTGATTTCGATTACAGGA CCGAAATCGTTCACACGAG	Não-publicado	M	A	-	NL
A-c233	(GA)19	55	128	1	CACGACGTTGTA AAAACGAGCAAAATGATGGAGGAGGTG GATCGGCATCACACTCTCTG	Não-publicado	M	A	-	NL
A-c234	(CA)12	60	114	3	CACGACGTTGTA AAAACGACCGCCAACTAAAACAGGAAA CAAAGTGTACCGGAAGTTCTCT	Não-publicado	P	A	-	NL
A-c243	(GT)10	60	117	1	CACGACGTTGTA AAAACGACCATGGAAGAAGCCATCTTTA TTGTGTTTCATGGCTAGAGAGAGA	Não-publicado	M	A	-	NL
A-c25.2	(TTC)4	55	125	1	CACGACGTTGTA AAAACGACTGTCCAAACTCACCACATTCA GGAGATGGCCCTATTTGTGA	Não-publicado	M	A	-	NL
A-c26	(TCC)4	50	124	1	CACGACGTTGTA AAAACGACCCCTTCCACTCTCATGCTGT GTTGGACTGTGTGGCAGTA	Não-publicado	M	A	-	NL
A-c296	(GA)12	60	114	4	CACGACGTTGTA AAAACGAGAGAGAAAACAGCGGGAGACA TCGGAGAGGAGAAAGATGAGC	Não-publicado	P	A	D	L
A-c30.1	(CTG)4	60	100	1	CACGACGTTGTA AAAACGACTCTTCCTTGCGAAAAAGCAT CCTAAAGGACCCTAATGTCCAA	Não-publicado	M	A	-	NL
A-c305	(GA)17	60	103	1	CACGACGTTGTA AAAACGAAACCGCGGTGTGTATAAT ATGTCACCCGCTCTCTGTCT	Não-publicado	M	A	-	NL
A-c306	(CACG)5	60	112	3	CACGACGTTGTA AAAACGAAATGATGGCGGTACACACTG CATGCTCTGTGGGAGTACA	Não-publicado	P	A	CD	L
A-c32.1	(CCT)4	60	104	2	CACGACGTTGTA AAAACGACGACTCCCGTCTCTTCTTC TCAGAGAGCTCGTTCAGCAG	Não-publicado	P	A	CD	L
A-c326	(TG)14	50	142	1	CACGACGTTGTA AAAACGATCCTGTGTGCCATAAGAAA CACATACTAAGGGGGTGTGC	Não-publicado	M	A	-	NL
A-c338	(CT)11	55	103	5	CACGACGTTGTA AAAACGAGATAGCATCCAGTACACAG GCAGAGTGAAGGATAGCAGGA	Não-publicado	P	A	D	L
A-c35	(CCT)5	55	108	2	CACGACGTTGTA AAAACGACTCGAACAAGTAGCGTTTGA TGAGGAGCTGTTCTTGTGC	Não-publicado	P	A	D	L
A-c46	(CT)5	60	112	1	CACGACGTTGTA AAAACGACCCCTCGTCTCTTTTCTTA TGCTGTGCTTCTGTCTCT	Não-publicado	M	A	-	NL
A-c47	(TCA)6	60	102	3	CACGACGTTGTA AAAACGACTCATCATCGTCTTTCGCATC CCGTCGTAGAAGTGCATCAA	Não-publicado	P	A	D	L
A-c52	(TCA)4	60	121	1	CACGACGTTGTA AAAACGACAGTCTTCAATATCGCCATACA CGTGTCTGTGTGGTTTGTG	Não-publicado	M	A	-	NL
A-c55	(ATC)6	55	144	3	CACGACGTTGTA AAAACGACATGATCCAGAGCACCCCTCAG TGTCTCATTTGCAAGAGCTG	Não-publicado	P	A	CD	L
A-c6	(CA)9	60	128	5	CACGACGTTGTA AAAACGACCAAAAATGAACATCACTCTGGTG CCTCATTGACGTGCCTAAA	Não-publicado	P	A	D	L
A-c63	(TCA)6	55	124	4	CACGACGTTGTA AAAACGACAAAGAACGGCAGTGTCTCGT TGAGGTTCAAGTCTCCGAGTT	Não-publicado	P	A	D	L
A-c68.2	(CAT)5	60	105	2	CACGACGTTGTA AAAACGACAACAAAGTCCAGGCATGAGG GGTTTTATTAGAGGGCGTTGA	Não-publicado	P	A	D	L

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
A-c68.2	(CAT)5	60	105	2	CACGACGTTGTA AAAACGACAACAAAAGTCCAGGCATGAGG GGTTTTATTTAGAGGGCGTTGA	Não-publicado	P	A	D	L
A-c70	(GAG)5	60	111	1	CACGACGTTGTA AAAACGACAGGAGATTCTCGCCATCCTT TAGTGGCCAGGTCCATATCC	Não-publicado	M	A	-	NL
A-c75	(TGA)5	55	105	1	CACGACGTTGTA AAAACGACGGCCTGGATCCTTTTTGATT GAGCAACTCCCGGTTGTA AAA	Não-publicado	M	A	-	NL
A-c82	(CT)8	60	102	4	CACGACGTTGTA AAAACGACCGTTGTTAAGCAAAGGGAAA AGATGTTGCTTAGCTCCAATTT	Não-publicado	P	A	D	L
A-c84	(AG)10	60	116	1	CACGACGTTGTA AAAACGACCTCGCTCTAGATCCGCTAA CGCACATGCATCCTTTTACA	Não-publicado	M	A	-	NL
A-c98	(CA)4	50	108	4	CACGACGTTGTA AAAACGAATGCTGATGCGATGCTGAT TTTGAGGCGAACTGATGTTTAG	Não-publicado	P	A	D	L
A-rnaseq1	(TCA)4	55	119	2	CACGACGTTGTA AAAACGAGGCGAGCTCTACCTCTTCAA GATCCGTAGCCGTTGTTGAC	Não-publicado	P	A	D	L
A-rnaseq2	(GGC)7	60	91	2	CACGACGTTGTA AAAACGACGATGACGATGACAGAAGGA AGTTGATGGTTGCGGGATT	Não-publicado	P	A	D	L
A-rnaseq4	(GCA)5	50	103	1	CACGACGTTGTA AAAACGAGAAGCGGACAGTGTGAAGGT GGCTGAGGTTTTCTGTGCT	Não-publicado	M	A	-	NL
A-rnaseq6	(GCT)5	60	118	1	CACGACGTTGTA AAAACGATGCTTCTATGCTAGTTCTACCG TCAGAAATAATGAGGGGGTA	Não-publicado	M	A	-	NL
A-rnaseq7	(GCG)5	60	82	2	CACGACGTTGTA AAAACGATAGGAGGAGGACGACCAGAA ACCCAGTACCTGTACGACGA	Não-publicado	P	A	D	L
B-c104	(GT)5	60	206	1	CACGACGTTGTA AAAACGACCCATGATGAATGAACCA CATGGCAGGATCGCTTAAAT	Não-publicado	M	A	-	NL
B-c107	(CA)9	50	197	5	CACGACGTTGTA AAAACGACACTGATGATGAATGAACCA TGCATTGTGCCATACTGTGA	Não-publicado	P	A	D	L
B-c114	(GA)14	60	196	1	CACGACGTTGTA AAAACGACAACCATCGCTTGATACTCC GGAAACAGCAGACGAGGAGA	Não-publicado	M	A	-	NL
B-c115	(TC)5	55	222	1	CACGACGTTGTA AAAACGACTGTGATGCCTGAAAAATTGAA ATGCGACCTTGTTCGGTAA	Não-publicado	M	A	-	NL
B-c118	(TC)9	55	233	4	CACGACGTTGTA AAAACGACGTAGTACGGCGATCCAGAGC CGGGGTGAATGATGTCAGAT	Não-publicado	P	A	D	L
B-c120	(CT)11	60	204	4	CACGACGTTGTA AAAACGACTACATGAGGCGCACTACCAA CTGCTGCCATTGTTTCATCTC	Não-publicado	P	A	-	NL
B-c123	(AG)11	60	217	5	CACGACGTTGTA AAAACGACGCGCCATATAGCATGAAGGT CAAGCGCATGAGAATCAAAA	Não-publicado	P	A	D	L
B-c15	(CT)9	60	201	1	CACGACGTTGTA AAAACGACGCAATCCCTCAACAACTCGT ATGTGATTTGTTGCGAGCAC	Não-publicado	M	A	-	NL
B-c199	(CA)10	60	200	2	CACGACGTTGTA AAAACGACCCATGATGAATGAACCA TGCATTGTGCCATACTGTGA	Não-publicado	P	A	D	L
B-c214	(GT)8	55	199	1	CACGACGTTGTA AAAACGAATGGTGGCAGTAAGCGTGTT AAGCGGTTGGAAATTAGGAGA	Não-publicado	M	A	-	NL
B-c25.1	(ATC)5	60	191	1	CACGACGTTGTA AAAACGACCTTCATGTCCGCTAACTTCTCA CCTTCTTCTCCAAGCCCTTT	Não-publicado	M	A	-	NL
B-c268	(AC)8	60	189	2	CACGACGTTGTA AAAACGAGCTACTCCAGTTGGCAAGGA TTGGTTCCGATAAAGCAAGG	Não-publicado	P	A	D	L
B-c312	(TC)13	50	200	5	CACGACGTTGTA AAAACGACCTGTCTTTCCCTCCTC AATCCTCCATGGTTGGTGAG	Não-publicado	P	A	D	L
B-c385	(CCATT)4	60	203	2	CACGACGTTGTA AAAACGAAAGGAGACATGCCCTGAC GAGGCATCAGAGGAAGTTGG	Não-publicado	P	A	-	NL
B-c44	(TCC)4	60	202	1	CACGACGTTGTA AAAACGACAGACGAACCGGAAGCAGAG GGAGTCGAAAACCTGGTGAAA	Não-publicado	M	A	-	NL

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
B-c450	(GT)17	60	178	1	CACGACGTTGTA AAAACGAAGCCTTGCACAGAGAAAACA AGCACCTGAAGTCCTGAACC	Não-publicado	M	A	-	NL
B-c49	(ATC)6	60	198	1	CACGACGTTGTA AAAACGACCCCTTGATTTCGGATGCCTAAA TGGAGACGATGGAGATGACA	Não-publicado	M	A	-	NL
B-c62	(CCT)4	60	204	2	CACGACGTTGTA AAAACGACTCTGAATCATCGTCGCCATA ATTCATGAACGAAGGGCTTG	Não-publicado	P	A	D	L
B-c74	(GAT)5	55	208	2	CACGACGTTGTA AAAACGACGGAGAACCCTTTCAGGAACC GCACGTAGAAATGTTGGCAAT	Não-publicado	P	A	D	L
B-c77.2	(CT)5	55	205	1	CACGACGTTGTA AAAACGACAATGCGGAGTGGGATTCTTA CCCTACTGCACCAAATGTCA	Não-publicado	M	A	-	NL
B-c85	(CA)27	60	193	1	CACGACGTTGTA AAAACGACCCAAATGCCACACCTTATT AAGACACTAGCCACATTGA	Não-publicado	M	A	-	NL
B-c88	(CT)12	60	223	3	CACGACGTTGTA AAAACGACCTCCTCCCCTCTCTCGTCTC CAAAAGCGCATTCAATAGCA	Não-publicado	P	A	D	L
B-c90	(GA)19	60	198	1	CACGACGTTGTA AAAACGACGTACGCCCTGTTGACATTT ATGAACGCCAGAGATCGAGT	Não-publicado	M	A	-	NL
B-c92	(TAC)4	50	221	1	CACGACGTTGTA AAAACGACCATCGCCTTATGAGCAGTCA ACGCGACTTGGTTGAAGAAA	Não-publicado	M	A	-	NL
B-c97	(GA)9	60	194	1	CACGACGTTGTA AAAACGACAGATAAACGCGCCCACTG CACAACCGTGGGGTTTTACT	Não-publicado	M	A	-	NL
B-c98	(AC)6	55	195	3	CACGACGTTGTA AAAACGACATCGGTCTCACGGCTTGTAT TAAGGAGCTCTGCCACGAAT	Não-publicado	P	A	D	L
B-rnaseq1	(CATA)4	60	212	3	CACGACGTTGTA AAAACGACTGGGTGAAGGAAGGTGTGT AGCGGCCGAAGATAGGTAAT	Não-publicado	P	A	D	L
B-rnaseq2	(AGC)4	60	185	1	CACGACGTTGTA AAAACGACTACCGTACTGTCGGCAATG GTACACGCTGGTGTCAAGGA	Não-publicado	M	A	-	NL
B-rnaseq3	(TTG)4	55	186	1	CACGACGTTGTA AAAACGAATAACAGGGCTGGTTGAAGC AGCAGAGATCCAACGCAACT	Não-publicado	M	A	-	NL
B-rnaseq6	(CTC)4	55	206	3	CACGACGTTGTA AAAACGACGCTCGTCAAATCAAGAAAC AGCGGGTATGTCACACGTA	Não-publicado	P	A	CD	L
C-c10	(GT)5	60	294	4	CACGACGTTGTA AAAACGACTTTGTTGGCTGCAAAAACCTG TTGCTTCATGGAATGTTACCAC	Não-publicado	P	A	D	L
C-c108.2	(CCAGC)3	60	305	1	CACGACGTTGTA AAAACGACTATTGCCCCGAGACTCTTGT ACGAAACAGAGGCACTCGAT	Não-publicado	M	A	-	NL
C-c122	(CA)8	60	300	1	CACGACGTTGTA AAAACGACGCAAGCGATTGTGTGAAAGA TGAGTTCCAAACTCCCCTA	Não-publicado	M	A	-	NL
C-c128.2	(GCAT)3	50	295	1	CACGACGTTGTA AAAACGACCCACGTC AAGAAATCATGG CAAACGTTGTTGGCTTGTG	Não-publicado	M	A	-	NL
C-c139	(GT)4	55	303	1	CACGACGTTGTA AAAACGACGGCAGCCATGACCTAAGTGT ACACCTGAGCCACGTTTTA	Não-publicado	M	A	-	NL
C-c145	(TG)6	60	307	2	CACGACGTTGTA AAAACGACGATCGATATGCGTGTGTTGC GGATTGCATGTCAGTCTTGG	Não-publicado	P	A	D	L
C-c148	(GT)6	55	303	1	CACGACGTTGTA AAAACGACTGGCGTAGGAGGAGATAGGA TTTGCTATGATGCGAAGTGC	Não-publicado	M	A	-	NL
C-c149.1	(CT)10	60	301	5	CACGACGTTGTA AAAACGACTGAAGCCATTGGGTACATCA GGGCTGCAGCAAAGTTAC	Não-publicado	P	A	D	L
C-c149.2	(CA)8	60	307	2	CACGACGTTGTA AAAACGACTAACGCATCACAAGGATGGA GCATGCATGGCGATAATCTA	Não-publicado	P	A	CD	L
C-c149.2	(CA)8	60	307	4	CACGACGTTGTA AAAACGACTAACGCATCACAAGGATGGA GCATGCATGGCGATAATCTA	Não-publicado	P	A	-	NL
C-c155	(CA)7	60	300	1	CACGACGTTGTA AAAACGACCGATGTTTTTGTGCTGGTG TTCCGGCCAATAATCCATAG	Não-publicado	M	A	-	NL

Loco SSR	Motivo	Ta	Tamanho	Alelos	Iniciador foward/reverse	Origem	Polimorfismo ³	Técnica de genotipagem ⁴	Dominância ⁵	Ligação ao mapa ⁶
C-c158	(GT)11	55	295	3	CACGACGTTGTA AAAACGACCTAGTAGCGGCCCGTAGTCA GGAATGGGAGAAAGTCGAAG	Não-publicado	P	A	D	L
C-c175.1	(AC)8	50	305	2	CACGACGTTGTA AAAACGACCCGATAGGAGACACGACGAT ATGCCAATCCCATAAGCAG	Não-publicado	P	A	-	NL
C-c186	(GT)6	60	318	2	CACGACGTTGTA AAAACGACCCAGGTGGGTGCAATTATGT TCATTGTCGTGCCACTAAC	Não-publicado	P	A	CD	L
C-c192	(TG)6	60	299	1	CACGACGTTGTA AAAACGACGTGTCGTGCGCACTTCTGTGT GCTCGGTTGGACCTGTTAA	Não-publicado	M	A	-	NL
C-c196	(AG)11	60	290	6	CACGACGTTGTA AAAACGACCCATGGAGAAAGGAGACCAG AAACTGTTCCACCCATTGCT	Não-publicado	P	A	D	L
C-c251	(GA)7	55	283	1	CACGACGTTGTA AAAACGAATTGTCGACTCCGCTGTCTT GAGGCTTGC AACCTTTGAT	Não-publicado	M	A	-	NL
C-c288	(AG)8	60	290	2	CACGACGTTGTA AAAACGAGTCATCGGTACGCTGACAAA CTGTGTGTTGGTTGGAGTGC	Não-publicado	P	A	CD	L
C-c297	(GT)9	55	311	3	CACGACGTTGTA AAAACGAGGGAATGCGGAACAATACTC GGTAGCGCTCTGCTCTCTTG	Não-publicado	P	A	D	L
C-c30.2	(GAG)4	60	308	2	CACGACGTTGTA AAAACGACACTACAGCTCTCGGTGGAA GGCCATGTGTTGCTTTGTAA	Não-publicado	P	A	-	NL
C-c366.1	(GA)10	60	305	2	CACGACGTTGTA AAAACGACACTGACCTCTTCCCTGAG AGTCACCCGGCCTCTAGTTT	Não-publicado	P	A	CD	L
C-c367	(GT)7	60	294	2	CACGACGTTGTA AAAACGAATTGAACA ACTGCCGAGATG CCAGATCAGCCCCAAAAGTA	Não-publicado	P	A	D	L
C-c368	(ATC)6	60	294	1	CACGACGTTGTA AAAACGAGAGCTCCAATGATGCCTTC ATAATAGTCCGTGCCAGCA	Não-publicado	M	A	-	NL
C-c375	(TG)7	60	286	1	CACGACGTTGTA AAAACGAGAGCAGCTCTGGTTGTTTTG GCCAAGGATAGACATGGGTAA	Não-publicado	M	A	-	NL
C-c387	(GT)9	55	294	2	CACGACGTTGTA AAAACGACAAGGAAAACCCAGCCAGGTA TAGTGTGCGCCGACATATTG	Não-publicado	P	A	CD	L
C-c39	(CTC)6	55	307	3	CACGACGTTGTA AAAACGACTTCCAGCTGTCCACATACCA CGTCTGTGTTGGCACTCTGT	Não-publicado	P	A	D	L
C-c436	(CA)7	60	304	2	CACGACGTTGTA AAAACGATAAAGAAGCAAACGGCATCG GTCGTCTTGCTGGCAGTACA	Não-publicado	P	A	D	L
C-c437	(CA)7	60	308	1	CACGACGTTGTA AAAACGATGTATCACGTGGACCTGCTT AGGTGCTGAAGTTTGCTGGT	Não-publicado	M	A	-	NL
C-c454	(TCA)5	50	295	1	CACGACGTTGTA AAAACGAGATCCGGAGATCGAGTTGTT GTTACGCCCATCTCTGATT	Não-publicado	M	A	-	NL
C-c79	(TC)5	60	298	2	CACGACGTTGTA AAAACGACAAAAGTGGCTGTGTTGGA CTTGCTGTTTTGCTGCAT	Não-publicado	P	A	D	L
C-rnaseq4	(TAG)5	60	306	2	CACGACGTTGTA AAAACGATGGTAGCATCGGGTATCAT ATGACCGAGGAGTGTGTGTG	Não-publicado	P	A	CD	L
C-rnaseq5	(AAGGG)4	55	295	1	CACGACGTTGTA AAAACGAGGGCTATGAAGAAGTCCGAGA GGGTTTTGCTGTGAGTGGA	Não-publicado	M	A	-	NL
C-rnaseq7	(CCT)6	60	299	3	CACGACGTTGTA AAAACGAGTCCCACTGCCAGCAATTA CCGGATATCCCTCCTTGAT	Não-publicado	P	A	CD	L

¹ (SOUSA, JUNGMANN, et al., 2011); ² (SOUSA, JANK, et al., 2011); ³ (P) polimórfico, (M) monomórfico; ⁴ (A) (SCHUELKE, 2000), (B) (CRESTE, NETO e FIGUEIRA, 2001); ⁵ (C) dominante, (D) codominante; ⁶ (L) ligado, (NL) não ligado.

Tabela Suplementar 2: Marcadores SSR classificados quanto aos grupos de ligação, tipo de marcador, e relação dos marcadores associados a cada genitor e aqueles compartilhados por ambos.

Grupos de ligação (GL) /tipos de marcador	Marcadores genitor S10	Marcadores compartilhados	Marcadores genitor Mombaça
GL 1 – GL 1	11	2	2
Codominantes	2	2	2
Dominantes	9	-	-
1:1	8	-	-
5:1	1	-	-
3:1	-	-	-
GL 2 – GL 2	2	2	13
Codominantes	-	-	-
Dominantes	2	2	13
1:1	-	-	9
5:1	-	-	1
3:1	2	2	3
GL 3 – GL3	20	3	4
Codominantes	3	2	2
Dominantes	17	1	2
1:1	9	-	-
5:1	2	-	1
3:1	6	1	1
GL 4 – GL4	3	2	9
Codominantes	2	2	2
Dominantes	1	-	7
1:1	1	-	7
5:1	-	-	-
3:1	-	-	-
GL 5- GL5	8	3	4
Codominantes	2	2	2
Dominantes	6	1	2
1:1	3	-	-
5:1	2	-	-

3:1	1	1	2
GL 6 – GL6	16	4	8
Codominantes	5	2	2
Dominantes	11	2	6
1:1	3	-	3
5:1	1	-	1
3:1	7	2	2
GL 6 – GL8	16	7	15
Codominantes	5	3	3
Dominantes	11	4	12
1:1	3	-	7
5:1	1	-	-
3:1	7	4	5
GL 7 – GL 7	21	6	12
Codominantes	5	1	1
Dominantes	16	5	11
1:1	6	-	5
5:1	1	-	1
3:1	9	5	5
GL 7 – GL 9	21	7	13
Codominantes	5	4	4
Dominantes	16	3	9
1:1	6	-	4
5:1	1	-	1
3:1	9	3	4

Tabela Suplementar 3: Marcadores SSR classificados quanto aos grupos de ligação e tipo de marcador, relativos ao genitor S10.

Grupos de ligação (GL) /tipos de marcador	Marcadores
GL 8 – S10	3
Codominantes	-
Dominantes	3
1:1	3
5:1	-
3:1	-
GL 9 – S10	2
Codominantes	-
Dominantes	2
1:1	2
5:1	-
3:1	-
GL 10 – S10	2
Codominantes	-
Dominantes	2
1:1	2
5:1	-
3:1	-

Tabela Suplementar 4: Marcadores SSR classificados quanto aos grupos de ligação e tipo de marcador, relativos ao genitor Mombaça.

Grupos de ligação (GL) /tipos de marcador	Marcadores
GL 11 - Mombaça	2
Codominantes	-
Dominantes	2
1:1	-
5:1	2
3:1	-
GL 12 - Mombaça	3
Codominantes	1
Dominantes	2
1:1	2
5:1	-
3:1	-
GL 13 - Mombaça	2
Codominantes	1
Dominantes	1
1:1	1
5:1	-
3:1	-
3:1	-

CONSIDERAÇÕES FINAIS

A principal motivação para a realização deste trabalho foi a possibilidade de contribuir para a construção do conhecimento sobre a espécie *Panicum maximum*, assim como as forrageiras tropicais de modo geral. A literatura sobre este grupo de gramíneas, de destacada importância econômica em nosso país, possui poucas informações no âmbito molecular. Acreditamos que as informações obtidas durante a realização deste trabalho possam fornecer os conhecimentos básicos para fomentar o crescimento da pesquisa em genética e genômica, necessário ao incremento do melhoramento genético das espécies de interesse.

Nesse sentido, a disponibilização de milhares de sequências inéditas, representando os transcritos expressos nas folhas de *P. maximum*, pode ser considerada como uma contribuição direta para os grupos de pesquisa em forrageiras tropicais (Capítulo I). O uso do transcriptoma como referência permitirá encontrar evidências de expressão diferencial entre diferentes genótipos, em diferentes ambientes e condições fisiológicas, elucidando os mecanismos de regulação gênica, e ao mesmo tempo, abrindo caminho para o desenvolvimento de marcadores funcionais nestas regiões, que podem ser aplicados nos programas de melhoramento genético. Proteínas de interesse podem ser clonadas em vetores de expressão para melhor compreensão das suas propriedades bioquímicas e vias de ação.

Os aproximadamente 13 Gb de *reads* gerados pelo sequenciamento das folhas foram depositados no arquivo público de *short reads* do NCBI (SRR821833). Os processos envolvidos em uma montagem *de novo* estão em constante desenvolvimento, o que torna possível revisitar os dados resultantes do Capítulo I, por diferentes grupos de pesquisa. Devido a enorme diversidade e complexidade existente na natureza, é arriscado afirmar que uma determinada metodologia de montagem de *short reads* seja recomendada para a maioria das espécies. Uma estratégia é testar a montagem dos *short reads* em diferentes montadores, que utilizem premissas distintas em seus respectivos *pipelines*. Fatores intrínsecos à espécie e

genótipos sequenciados, como diversidade genética e quantidade de variantes como os SNPs, podem influenciar no resultado final da montagem. Desta forma, exibiremos aqui alguns resultados nesta etapa de seleção da montagem dos *reads* de *P. maximum*, onde as limitações de espaço impostas pelo periódico impossibilitaram a incorporação de todas as etapas envolvidas no processo. Além do montador Trinity (GRABHERR et al., 2011), foi utilizado também o programa CLC Genomics Workbench, ou apenas CLC para fins desta discussão. O programa CLC é um pacote pago, disponível em nosso laboratório, com premissas e metodologias diferentes de montagem quando comparadas ao programa Trinity (GRABHERR et al., 2011). Em síntese, montadores como os programas Trinity (GRABHERR et al., 2011) e Oases (SCHULZ et al., 2012), tendem a resolver todas as possibilidades resultantes dos grafos de *Brujin* gerados a partir dos diferentes k-mers. Isto geralmente acarreta em uma maior quantidade de contigs formados, os quais representam variantes de um mesmo transcrito, como aquelas resultantes de *splicing* alternativo ou cópias recentemente duplicadas no genoma, são disponibilizadas no conjunto final de transcritos da respectiva montagem *de novo*. Desta forma, existe mais informação disponível acerca dos transcritos, porém isto acarreta em maior esforço para garantir um conjunto de genes não-redundantes (unigenes) e maior probabilidade de formação de artefatos (sequências sem significado biológico). O montador presente no pacote CLC Genomics Workbench resolve essas variantes encontradas selecionando apenas uma para representar o transcrito, com critérios pré-estabelecidos. Assim, cada estratégia apresenta seus prós e contras, e somente através de uma avaliação criteriosa, partindo de resultados de diferentes montagens, é possível selecionar uma que se adeque ao objetivo do estudo em questão. Foram testadas três diferentes montagens a princípio, com o conjunto de *reads* de *P. maximum*. Duas delas foram realizadas utilizando o montador CLC, alterando os valores de tamanho do k-mer (25 e 45 pb), e uma montagem através do montador Trinity (GRABHERR et al., 2011), no qual o tamanho de k-mer é fixado em 25 pb. Em teoria, montagens que utilizem

maiores valores de k-mer tendem a ser mais estridentes quanto a formação de contigs, resultando em um menor e mais confiável conjunto de contigs. Por outro lado, existe a perda de informação, muitas vezes com significado biológico. Alguns resultados e métricas destas três montagens são apresentados na Tabela 1:

Tabela 1: Métricas utilizadas para avaliação das montagens dos *reads* originados do sequenciamento do transcriptoma de *Panicum maximum*.

	CLC Genomics		Trinity
k-mer	25	45	25
Contigs	36.029	31.264	88.292
N50	944 bp	848 bp	1.272 bp
<i>reads</i> mapeados (%)*	69.81%	74.29%	85.96%

*mapeador Bowtie (LANGMEAD et al., 2009)

Nota-se que a montagem do programa Trinity obteve um número maior de contigs, o que é esperado devido as premissas do método, mas também apresentou as métricas de N50 no valor esperado para transcriptoma de plantas (entre 1-2 kb). A quantidade de *reads* mapeados corretamente de volta aos contigs formados também foi avaliada, processo realizado através do mapeador Bowtie (LANGMEAD et al., 2009). Espera-se que um maior número de *reads* seja representado em uma montagem bem sucedida. Neste parâmetro, novamente o montador Trinity apresentou vantagem em relação as montagens do CLC.

Assim, através da avaliação conjunta destes resultados, e outras métricas e testes realizados, decidimos continuar as análises referentes ao transcriptoma de folhas de *P. maximum* utilizando a montagem do programa Trinity (GRABHERR et al., 2011). O processo subsequente de anotação mostrou que grande parte dos contigs foi relacionada com sucesso a genes de outras espécies próximas, assim como análise via KEGG (KANEHISA e GOTO, 2000) que identificou diversas vias metabólicas conhecidas, muitas delas totalmente representadas pelos transcritos gerados (TOLEDO-SILVA et al., 2013). Os *reads* gerados a partir do trabalho do primeiro capítulo desta tese, possibilitam outras abordagens no futuro,

conforme as metodologias de montagem de *short reads* forem se aprimorando. Conforme citado por POLAND e RIFE (2012), dados pertinentes a análises de bioinformática são dinâmicos, significando que podem e devem ser revisitados ao longo do tempo para mais inferências, visto o desenvolvimento de novas metodologias e o avanço das informações contidas nos bancos de genes públicos. Esperamos que assim como o resultado das análises do Capítulo I, os *reads* brutos do sequenciamento possam contribuir para futuras análises e trabalhos relacionados a *Panicum maximum* e gramíneas tropicais.

Em relação ao Capítulo II, destaca-se o desenvolvimento dos novos marcadores microssatélites, que servirão como ferramentas moleculares no programa de melhoramento de *P. maximum*, podendo ser usados na identificação de híbridos, caracterização de cultivares, seleção de genitores para cruzamentos, avaliação do fluxo gênico e verificação da taxa de cruzamento da espécie. Todos os marcadores desenvolvidos neste trabalho, além da genotipagem na população de mapeamento, também foram genotipados em 25 acessos do banco de germoplasma da espécie, presente na Embrapa Gado de Corte (Campo Grande – MS). Foram selecionados os mesmos acessos utilizados no trabalho de SOUSA et al. (2011a). Estes dados serão analisados conforme o trabalho citado, na intenção de uma melhor caracterização acerca destes novos marcadores SSR desenvolvidos.

Em relação aos mapas de ligação obtidos no capítulo II, a construção dos mesmos a partir de locos microssatélites, proporcionou um melhor conhecimento sobre a estrutura do genoma autotetraploide de *P. maximum*. Os métodos estatísticos utilizados permitiram a análise de diversos locos polimórficos na população de mapeamento, sendo que parte deles puderam ser classificados e utilizados como marcadores multi-alélicos codominantes. Este feito é inédito para a espécie visto que o único mapa (EBINA et al., 2005) baseou-se apenas na metodologia de marcadores de dose única (WU et al., 1992). Nota-se que o mapeamento de QTLs provou ter enorme utilidade para o melhoramento de plantas e, geralmente, obtêm

sucesso na identificação de locos de características quantitativas (MORRELL, BUCKLER e ROSS-IBARRA, 2011). Contudo, para características quantitativas mais complexas, o mapa construído no Capítulo II deverá ser saturado com mais marcadores, garantindo assim maior probabilidade de sucesso na detecção de QTLs. Uma característica possível de ser mapeada com os mapas de ligação construídos neste trabalho é relativa ao modo de reprodução apomítico. Trabalhos com mapas menos saturados obtiveram sucesso na identificação do loco ligado a aposporia, como em *Panicum maximum* (EBINA et al., 2005) e *Brachiaria humidicola* (VIGNA, 2010). Isso deve-se ao fato de que, em ambos os casos, a aposporia foi mapeada em um único grupo de ligação, aumentando a probabilidade de sucesso no mapeamento da característica.

RESUMO DOS RESULTADOS

No presente trabalho, concluímos que os objetivos foram alcançados de maneira satisfatória, com os principais resultados sendo resumidos a seguir:

Capítulo I:

- Primeiro transcriptoma completo de folhas descrito para *P. maximum*;
- Obtenção de 38.192 unigenes, sendo cerca de 65% destes anotados com sucesso em diferentes bancos de dados;
- Maior parte dos genes de interesse encontrados, referentes às vias de metabolismo C₄ e síntese de lignocelulose;
- Grande número de marcadores moleculares putativos identificados: 5.035 SSRs e 346.236 SNPs.

Capítulo II:

- Desenvolvimento de 131 novos marcadores do tipo microssatélite para espécie *Panicum maximum*;
- Construção de mapa genético parcial utilizando marcadores codominantes, utilizando metodologia apropriada para autotetraploides;
- Mapa parcial do genitor S10 apresentou 10 grupos de ligação, cobrindo cerca de 672.1 cM do genoma, a uma densidade de 1 marcador/7.63 cM;
- Mapa parcial do genitor Mombaça apresentou 13 grupos de ligação, cobrindo cerca de 802.2 cM do genoma, a uma densidade de 1 marcador/8.35 cM.

CONCLUSÕES

De forma geral, foi possível conhecer uma parte do genoma e transcriptoma de *Panicum maximum*, pela observação das características dos marcadores microssatélites e do mapa de ligação construído, e pela obtenção do transcriptoma de folhas através de sequenciamento de nova geração. Os marcadores e o mapa forneceram simultaneamente informações previamente desconhecidas para a espécie, e o transcriptoma sequenciado permitiu descrever pela primeira vez os genes transcritos envolvidos no metabolismo de folhas de *P. maximum*, além da identificação de uma gama de novos marcadores moleculares.

Concluimos que os dados obtidos neste trabalho, considerando os dois capítulos desta tese, trazem as informações e ferramentas necessárias a realização de um grande avanço nos programas de melhoramentos genético, bem como ao conhecimento genético-molecular de *Panicum maximum*.

PERSPECTIVAS

Os resultados obtidos nesta tese de doutorado permitirão a continuação da pesquisa apresentada em ambos os capítulos:

Capítulo I: As gramíneas africanas apresentam vigor e crescimento notáveis quando em comparação com as forrageiras sul-americanas como as do gênero *Paspalum*. Acredita-se que as forrageiras africanas tenham sido selecionadas ao longo de sua história evolutiva, sendo pisoteadas e usadas como alimento por grandes mamíferos como elefantes, rinocerontes, entre outros (JANK, VALLE e RESENDE, 2011). Este processo teria favorecido populações com maior taxa de crescimento e resistência. As sequências obtidas neste capítulo serão comparadas com o transcriptoma de outra gramínea africana, *Brachiaria humidicola*, trabalho que está sendo desenvolvido por nosso grupo. Através da identificação de possíveis genes ortólogos entre as duas espécies, estes serão comparados com outros transcriptomas disponíveis de gramíneas originadas de outras localidades, como *Panicum virgatum* e *Paspalum notatum*. A interpretação das relações de ortologia entre as espécies africanas e não-africanas pode evidenciar possíveis modificações vantajosas para as forrageiras africanas, além de genes exclusivos. Uma análise preliminar, utilizando a metodologia de blast recíproco (*reciprocal best blast hit*) (MORENO-HAGELSIEB e LATIMER, 2008), mostrou que existem cerca de 15.000 genes ortólogos entre *P. maximum* e *B. humidicola*. Este conjunto de ortólogos foi então comparado aos transcriptomas disponíveis de gramíneas no banco de dados Phytozome (www.phytozome.net). A relação filogenética sugerida pelo banco de dados pode ser correlacionada com o número de ortólogos encontrados. Análises mais aprofundadas, utilizando outras metodologias e bancos de dados serão realizadas, visando levantar hipóteses sobre os mecanismos moleculares responsáveis pelo vigor das forrageiras africanas.

Capítulo II: Os mapas genéticos parciais de *Panicum maximum* obtidos neste capítulo servirão de base para a construção de mapas mais saturados. Para tal, serão desenvolvidos e

genotipados marcadores moleculares do tipo *single nucleotide polymorphism* (SNP) para a espécie. Duas abordagens serão utilizadas a princípio: (a) Desenvolvimento de marcadores SNPs a partir do transcriptoma de *Panicum maximum*, para genotipagem na plataforma Sequenom iPlex Mass Array (Sequenom Inc). Através da metodologia implementada no programa SuperMASSA (SERANG, MOLLINARI e GARCIA, 2012), será possível estimar a dosagem alélica exata de cada marcador genotipado com sucesso, permitindo usar todo o potencial deste tipo de marcador molecular para a construção de mapas de ligação. (b) aplicação da metodologia de genotipagem-por-sequenciamento (*genotyping-by-sequencing* ou GBS) (ELSHIRE et al., 2011), obtendo milhares de locos SNPs genotipados. Será utilizado o genoma *draft* de *Panicum virgatum* (disponível no banco Phytozome) para realizar as análises de bioinformática associadas. Os dados referentes a genotipagem de SNPs de ambas as metodologias serão integradas aos mapas de ligação gerados neste capítulo. Com a obtenção de mapas mais saturados, as análises para identificação de QTLs serão realizadas a partir de dados fenotípicos coletados pela Dra. Liana Jank na Embrapa Gado de Corte (Campo Grande – MS). Diversas características de interesse serão avaliadas no mapa de ligação, na tentativa de identificar locos relacionados ao desenvolvimento das características em questão.

LITERATURA CITADA

- AKIYAMA, Y. et al. Estimation of genome size and physical mapping of ribosomal DNA in diploid and tetraploid guineagrass (*Panicum maximum* Jacq.). **Grassland Science**, v. 54, n. 2, p. 89–97, 2008.
- ALVES, R. G. O. Prefácio. In: **Melhoramento de Forrageiras Tropicais**. Campo Grande: Embrapa Gado de Corte, p. 7, 2008.
- BADGER, M. R. e PRICE, G. D. The role of carbonic anhydrase in photosynthesis. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 45, p. 369–392, 1994.
- BALL, A. D. et al. A comparison of SNPs and microsatellites as linkage mapping markers: lessons from the zebra finch (*Taeniopygia guttata*). **BMC genomics**, v. 11, n. 1, p. 218, 2010.
- BARRATT, D. H. P.; DERBYSHIRE, P.; FINDLAY, K.; PIKE, M.; WELLNER, N.; et al. Normal growth of *Arabidopsis* requires cytosolic invertase but not sucrose synthase. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, p. 13124–13129, 2009.
- BASHAW, E. C.; HOVIN, A. W. e HOLT, E. C. Apomixis, its evolutionary significance and utilization in plant breeding. In: **Proceedings of the XI International Grassland Congress**, Queensland, Australia. Anais, [S.l.: s.n.], 1970.
- BATISTOTI, C.; LEMPP, B.; JANK, L.; MORAIS, M. D. G.; CUBAS, A. C.; et al. Correlations among anatomical, morphological, chemical and agronomic characteristics of leaf blades in *Panicum maximum* genotypes. **Animal Feed Science and Technology**, v. 171, p. 173–180, 2012.
- BENSON, D. A. et al. GenBank. **Nucleic acids research**, v. 41, n. Database issue, p. D36–42, 2013.
- BETHEL, C. M.; SCIARA, E. B.; ESTILL, J. C.; BOWERS, J. E.; HANNA, W. e PATERSON, A. H. A framework linkage map of bermudagrass (*Cynodon dactylon* transvaalensis) based on single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 112, n. 4, p. 727–737, 2006.
- BOERJAN, W.; RALPH, J. e BAUCHER, M. Lignin biosynthesis. **Annual Review of Plant Biology**, v. 54, p. 519–546, 2003.
- BOGDAN, A. V. **Tropical pasture and fodder plants**. [S.l.]: Longman., 1977.
- BOUTON, J. H. Molecular breeding of switchgrass for use as a biofuel crop. **Current Opinion in Genetics & Development**, v. 17, p. 553–558, 2007.
- BRADSHAW, J. E.; HACKETT, C. A.; PANDE, B.; WAUGH, R. e BRYAN, G. J. QTL mapping of yield, agronomic and quality traits in tetraploid potato (*Solanum tuberosum* subsp. *tuberosum*). **Theoretical and Applied Genetics**, v. 116, n. 2, p. 193–211, 2008.

- BRAZ, T. **Parâmetros genéticos, repetibilidade e diversidade genética em híbridos de *Panicum maximum***. Tese (Doutorado) - Universidade Federal de Viçosa, 2012.
- BYERS, R. L. et al. Development and mapping of SNP assays in allotetraploid cotton. **Theoretical and Applied Genetics**, v. 124, n. 7, p. 1201–1214, 2012.
- CAIXETA, E. et al. Tipos de marcadores moleculares. In: BORÉM, A.; CAIXETA, E. (Eds.). **Marcadores moleculares**. Viçosa - MG: Editora Folha de Viçosa, 2009. p. 11–94.
- CAMACHO, C.; COULOURIS, G.; AVAGYAN, V.; MA, N.; PAPADOPOULOS, J.; et al. BLAST+: architecture and applications. **BMC Bioinformatics**, v. 10, p. 421, 2009.
- CARROLL, A. e SOMERVILLE, C. Cellulosic biofuels. **Annual Review of Plant Biology**, v. 60, p. 165–182, 2009.
- CHANDRA, A. e TIWARI, K. K. Isolation and characterization of microsatellite markers from guineagrass (*Panicum maximum*) for genetic diversity estimate and cross-species amplification. **Plant Breeding**, v. 129, n. 1, p. 120–124, 2010.
- CHEN, L. et al. Isolation and Characterization of a Gene Expressed during Early Embryo Sac Development in Apomictic Guinea Grass (*Panicum maximum*). **Journal of plant physiology**, v. 154, n. 1, p. 55–62, 1999.
- CHEN, L. et al. Developmental expression of ASG-1 during gametogenesis in apomictic guinea grass (*Panicum maximum*). **Journal of plant physiology**, v. 162, n. 10, p. 1141–1148, 2005.
- CLIFTON-BROWN, J.; BREUER, J. e JONES, M. Carbon mitigation by the energy crop, *Miscanthus*. **Global Change Biology**, v. 13, n. 11, p. 2296–2307, 2007.
- COBBETT, C. e GOLDSBROUGH, P. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. **Annual Review of Plant Biology**, v. 53, p. 159–182, 2002.
- CONESA, A.; GÖTZ, S.; GARCÍA-GÓMEZ, J. M.; TEROL, J.; TALÓN, M.; et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. **Bioinformatics**, v. 21, p. 3674–3676, 2005.
- COSTA, R. e JANK, L. Avaliação da Biomassa da Forrageira *Panicum maximum* Como Alternativa à Produção de Etanol Celulósico. In: **5a Jornada Científica da Embrapa Gado de Corte**. Anais, Campo Grande - MS: [s.n.], 2009.
- CRESTE, S.; NETO, A. T. e FIGUEIRA, A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. **Plant Molecular Biology Reporter**, v. 19, n. 4, p. 299–306, 2001.
- DALL'AGNOL, M. e SCHIFINO-WITTMANN, M. T. Apomixia, genética e melhoramento de plantas. **Revista Brasileira de Agrociência**, v. 11, n. 2, p. 127–133, 2005.

DANECEK, P.; AUTON, A.; ABECASIS, G.; ALBERS, C. A.; BANKS, E.; et al. The variant call format and VCFtools. **Bioinformatics**, v. 27, p. 2156–2158, 2011.

DIAS-FILHO, M. B. **Degradação de pastagens: processos, causas e estratégias de recuperação**. [S.l.]: Embrapa Amazônia Oriental, 2007.

EBINA, M.; NAKAGAWA, H.; YAMAMOTO, T.; ARAYA, H.; TSURUTA, S. I.; TAKAHARA, M. e NAKAJIMA, K. Co-segregation of AFLP and RAPD markers to apospory in Guinea grass (*Panicum maximum* Jacq.). **Grassland Science**, v. 51, n. 1, p. 71–78, 2005.

EBINA, M. et al. Genetic relationship estimation in guineagrass (*Panicum maximum* Jacq.) assessed on the basis of simple sequence repeat markers. **Grassland Science**, v. 53, n. 3, p. 155–164, 2007.

EKBLOM, R. e GALINDO, J. Applications of next generation sequencing in molecular ecology of non-model organisms. **Heredity**, v. 107, n. 1, p. 1–15, 2011.

ELSHIRE, R. J.; GLAUBITZ, J. C.; SUN, Q.; POLAND, J. A.; KAWAMOTO, K.; BUCKLER, E. S. e MITCHELL, S. E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. **PloS one**, v. 6, n. 5, p. e19379, 2011.

ENDLER, A. e PERSSON, S. Cellulose synthases and synthesis in *Arabidopsis*. **Molecular Plant**, v. 4, p. 199–211, 2011.

EUCLIDES, V. P. B. et al. Brazilian scientific progress in pasture research during the first decade of XXI century. **Revista Brasileira de Zootecnia**, v. 39, p. 151–168, 2010.

FAOSTAT. **Live Animals Data**. Disponível em: <http://faostat.fao.org>. Acessado em 8 de Novembro de 2013. *Food and Agriculture Organization of United Nations*, Organização das Nações Unidas (ONU), 2011.

FERREIRA, M. E. e GRATTAPAGLIA, D. Introdução ao uso de marcadores moleculares em análise genética. **Embrapa**, Brasília, 1996.

GAR, O.; SARGENT, D. J.; TSAI, C. J.; PLEBAN, T.; SHALEV, G.; BYRNE, D. H. e ZAMIR, D. An autotetraploid linkage map of rose (*Rosa hybrida*) validated using the strawberry (*Fragaria vesca*) genome sequence. **PloS one**, v. 6, n. 5, p. e20463, 2011.

GARCIA, A. A. F.; KIDO, E. A.; MEZA, A. N.; SOUZA, H. M. B.; PINTO, L. R.; PASTINA, M. M.; LEITE, C. S.; DA SILVA, J. A. G.; ULIAN, E. C. e FIGUEIRA, A. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. **Theoretical and Applied Genetics**, v. 112, n. 2, p. 298–314, 2006.

GARRISON, E. M. G. Haplotype-based variant detection from short-read sequencing. **arXiv preprint**, arXiv:12073907: 9, 2012

GAUER, L. e CAVALLI-MOLINA, S. Apomixia: um método alternativo para a produção de sementes em plantas. **Pesquisa Agropecuária Brasileira**, v. 6, n. 1, p. 157–170, 2000.

- GOLDSTEIN, D. B. e SCHLOTTERER, C. **Microsatellites: evolution and applications**. Oxford University Press: Oxford, 1999.
- GRABHERR, M. G. et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology**, v. 29, n. 7, p. 644–52, 2011.
- GUO, M.; DAVIS, D. e BIRCHLER, J. A. Dosage effects on gene expression in a maize ploidy series. **Genetics**, v. 142, n. 4, p. 1349–1355, 1996.
- GUTTMAN, M. et al. Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. **Nature biotechnology**, v. 28, n. 5, p. 503–510, 2010.
- HACKETT, C. A.; BRADSHAW, J. E. e McNICOL, J. W. Interval mapping of quantitative trait loci in autotetraploid species. **Genetics**, v. 159, n. 4, p. 1819–32, 2001.
- HACKETT, C. A.; BRADSHAW, J. E.; MEYER, R. C.; McNICOL, J. W.; MILBOURNE, D. e WAUGH, R. Linkage analysis in tetraploid species: a simulation study. **Genetical research**, v. 71, n. 02, p. 143–153, 1998.
- HACKETT, C. A.; MILNE, I.; BRADSHAW, J. E. e LUO, Z. TetraploidMap for Windows: linkage map construction and QTL mapping in autotetraploid species. **The Journal of heredity**, v. 98, n. 7, p. 727–9, 2007.
- HACKETT, C. A.; BRADSHAW, J. E. e McNICOL, J. W. Interval mapping of quantitative trait loci in autotetraploid species. **Genetics**, v. 159, n. 4, p. 1819–1832, 2001.
- HACKETT, C. A. e LUO, Z. W. TetraploidMap: construction of a linkage map in autotetraploid species. **Journal of Heredity**, v. 94, n. 4, p. 358–359, 2003.
- HACKETT, C. A.; PANDE, B. e BRYAN, G. J. Constructing linkage maps in autotetraploid species using simulated annealing. **Theoretical and Applied Genetics**, v. 106, n. 6, p. 1107–1115, 2003.
- HANNA, W. W. Use of apomixis in cultivar development. **Advances in Agronomy**, v. 54, p. 333–350, 1995.
- HARRIS, M. a et al. The Gene Ontology (GO) database and informatics resource. **Nucleic acids research**, v. 32, n. Database issue, p. D258–61, 2004.
- HENRY, R. J. **Plant genotyping II: SNP technology**. [S.l.]: CABI, v. 2, 2008.
- HITCHCOCK, A. S. **Manual of the grasses of the United States: in two volumes**. [S.l.]: Courier Dover Publications, v. 2, 1971
- HORNETT, E. A. e WHEAT, C. W. Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species. **BMC genomics**, v. 13, n. 1, p. 361, 2012.

IBGE. **Censo Agropecuário**. Instituto Brasileiro de Geografia e Estatística, Ministério do Planejamento, Orçamento e Gestão. Governo Brasileiro, 2006.

IBGE. **Pesquisa pecuária municipal**. Disponível em www.sidra.ibge.gov.br. Acessado em 8 de Dezembro de 2013. Instituto Brasileiro de Geografia e Estatística, Ministério do Planejamento, Orçamento e Gestão. Governo Brasileiro, 2011.

JANK, L. et al. Avaliação do germoplasma de *Panicum maximum* introduzido da África: 1. Produção forrageira. **Revista da Sociedade Brasileira de Zootecnia**, v. 23, n. 3, 1994.

JANK, L. Melhoramento e seleção de variedades de *Panicum maximum*. **Simpósio sobre manejo da pastagem**, v. 12, p. 21–58, 1995.

JANK, L. et al. **Catálogo de caracterização e avaliação de germoplasma de *Panicum maximum*: descrição morfológica e comportamento agrônomo**. [S.l.]: EMBRAPA-CNPQC, 1997.

JANK, L.; VALLE, C. B.; RESENDE, R. M. S. Grass and forage plant improvement in the tropics and sub-tropics. In: **Grassland: a global resource**. Wageningen: Wageningen Academic Publishers. p. 69–81, 2005.

JANK, L.; RESENDE, R. M. S.; VALLE, C. B.; RESENDE, M. D. V.; CHIARI, L.; CANÇADO, L. J. e SIMONI, C. Melhoramento genético de *Panicum maximum* Jacq. In: **Melhoramento de forrageiras tropicais**. Campo Grande: Embrapa Gado de Corte, p. 55–87, 2008.

JANK, L. et al. Potencial de utilização do banco ativo de germoplasma da forrageira *Panicum maximum* Jacq. como fonte de energia. In: Congresso Brasileiro De Recursos Genéticos. **Anais**, Salvador - BA: [s.n.], 2010.

JANK, L.; VALLE, C. B. e RESENDE, R. Breeding tropical forages. **Crop Breeding and Applied Biotechnology**, v. 11, n. spe, p. 27–34, 2011.

JESSUP, R. W.; BURSON, B. L.; BUROW, G.; WANG, Y-W.; CHANG, C. LI, Z.; PATERSON, A. H. e HUSSEY, M. A. Segmental allotetraploidy and allelic interactions in buffelgrass (*Pennisetum ciliare* (L.) Link syn. *Cenchrus ciliaris* L.) as revealed by genome mapping. **Genome**, v. 46, n. 2, p. 304–313, 2003.

JULIER, B.; FLAJOULOT, S.; BARRE, P.; CARDINET, G.; SANTONI, S.; HUGUET, T. e HUYGHE, C. Construction of two genetic linkage maps in cultivated tetraploid alfalfa (*Medicago sativa*) using microsatellite and AFLP markers. **BMC plant biology**, v. 3, p. 9, 2003.

JUNG, H. G. e ALLEN, M. S. Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. **Journal of Animal Science**, v. 73, p. 2774–2790, 1995.

KALENDAR, R.; LEE, D. e SCHULMAN, A. H. FastPCR software for PCR primer and probe design and repeat search. **Focus on Bioinformatics/Mansour, A.(Ed)**, 2009.

- KANAI, R. e EDWARDS, G. E. The biochemistry of C4 photosynthesis. In: **C4 Plant Biology**. San Diego, Academic Press, v. 4., p. 49–87, 1999.
- KANEHISA, M. e GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. **Nucleic acids research**, v. 28, n. 1, p. 27–30, 2000.
- KISTNER, C. e MATAMOROS, M. RNA isolation using phase extraction and LiCl precipitation. In: *Lotus japonicus Handbook*. Dordrecht: Springer Netherlands. p. 123–124, 2005.
- KWOK, P. Methods for genotyping single nucleotide polymorphisms. **Annual review of genomics and human genetics**, n. 11, 2001.
- LABATE, J. A. e BALDO, A. M. Tomato SNP discovery by EST mining and resequencing. **Molecular Breeding**, v. 16, n. 4, p. 343–349, 2005.
- LABORDA, P. **Marcadores moleculares microssatélites na investigação do genoma de *Drosophila mediopunctata*: desenvolvimento e construção de mapa genético de ligação**. Universidade Estadual de Campinas - [S.l.]. 2011.
- LANGMEAD, B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. **Genome biology**, v. 10, n. 3, p. R25, doi:10.1186/gb-2009-10-3-r25, 2009.
- LEACH, L. J. et al. Multilocus tetrasomic linkage analysis using hidden Markov chain model. **Proceedings of the National Academy of Sciences**, v. 107, n. 9, p. 4270–4274, 2010.
- LEOPOLDINO, W. M. **Avaliação nutricional de pastagens consorciadas com leguminosas tropicais, dinâmica ruminal e produção de leite de vacas mestiças**. Dissertação (Mestrado), Universidade Federal de Minas Gerais, 2000.
- LI, X. P.; BJÖRKMAN, O.; SHIH, C.; GROSSMAN, A. R.; ROSENQUIST, M.; et al. A pigment-binding protein essential for regulation of photosynthetic light harvesting. **Nature**, v. 403, p. 391–395, 2000.
- LI, H. e DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, v. 25, p. 1754–1760, 2009.
- LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, p. 2078–2079, 2009.
- LI, B. e DEWEY, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. **BMC Bioinformatics**, v.12, p 323, 2011.
- LITT, M. e LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American journal of human genetics**, v. 44, n. 3, p. 397, 1989.
- LIU, M.; QIAO, G.; JIANG, J.; YANG, H.; XIE, L.; et al. Transcriptome sequencing and *de novo* analysis for Ma bamboo (*Dendrocalamus latiflorus* Munro) using the Illumina platform. **PLoS One**, v. 7, e46766, 2012.

- LIU, S.; LI, W.; WU, Y.; CHEN, C.; LEI, J. *De novo* transcriptome assembly in chili pepper (*Capsicum frutescens*) to identify genes involved in the biosynthesis of capsaicinoids. **PloS One**, v. 8, e48156, 2013.
- LULIN, H.; XIAO, Y.; PEI, S.; WEN, T.; SHANGQIN, H. The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. **PloS One**, v. 7, e38653, 2012.
- LUO, Z. W.; HACKETT, C. A.; BRADSHAW, J. E.; McNICOL, J. W. e MILBOURNE, D. Predicting parental genotypes and gene segregation for tetrasomic inheritance. **Theoretical and Applied Genetics**, v. 100, n. 7, p. 1067–1073, 2000.
- LUO, Z. W.; HACKETT, C. A.; BRADSHAW, J. E.; McNICOL, J. W. e MILBOURNE, D. Construction of a genetic linkage map in tetraploid species using molecular markers. **Genetics**, v. 157, n. 3, p. 1369–1385, 2001.
- MALIEPAARD, C.; JANSEN, J. e OOIJEN, J. W. V. Linkage analysis in a full-sib family of an outbreeding plant species: overview and consequences for applications. **Genetical Research**, v. 70, n. 3, p. 237–250, 1997.
- MATSUOKA, M.; FURBANK, R. T.; FUKAYAMA, H.; MIYAO, M. Molecular engineering of C4 photosynthesis. **Annual Review of Plant Physiology and Plant Molecular Biology**, v. 52, p. 297–314, 2001.
- MARDIS, E. R. The impact of next-generation sequencing technology on genetics. **Trends in genetics : TIG**, v. 24, n. 3, p. 133–41, doi:10.1016/j.tig.2007.12.007, 2008.
- MARGUERAT, S. e BÄHLER, J. RNA-seq: from technology to biology. **Cellular and molecular life sciences : CMLS**, v. 67, n. 4, p. 569–79, doi:10.1007/s00018-009-0180-6, 2010.
- MARTIN, J. A. e WANG, Z. Next-generation transcriptome assembly. **Nature reviews. Genetics**, v. 12, n. 10, p. 671–82, doi:10.1038/nrg3068, 2011.
- McKENDRY, P. Energy production from biomass (Part 1): Overview of biomass. **Bioresource technology**, v. 83, n. 1, p. 37–46, 2002.
- McLAUGHLIN, S. B. e ADAMS KSZOS, L. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. **Biomass and Bioenergy**, v. 28, n. 6, p. 515–535, 2005.
- MEYER, R. C.; MILBOURNE, D.; HACKETT, C. A.; BRADSHAW, J. E.; McNICHOL, J. W. e WAUGH, R. Linkage analysis in tetraploid potato and association of markers with quantitative resistance to late blight (*Phytophthora infestans*). **Molecular & general genetics**, v. 259, n. 2, p. 150–60, 1998.
- MIAN, M. A. R.; SAHA, M. C.; HOPKINS, A. A. e WANG, Z-Y. Use of tall fescue EST-SSR markers in phylogenetic analysis of cool-season forage grasses. **Genome**, v. 48, n. 4, p. 637–647, 2005.

MIR, G.; DOMÈNECH, J.; HUGUET, G.; GUO, W.-J.; GOLDSBROUGH, P.; et al. A plant type 2 metallothionein (MT) from cork tissue responds to oxidative stress. **Journal of Experimental Botany**, v. 55, p. 2483–2493, 2004.

MOLLINARI, M. **Comparação de algoritmos usados na construção de mapas genéticos**. Dissertação (Mestrado) - Universidade Estadual de São Paulo, 2007.

MOLLINARI, M. **Desenvolvimento de um modelo para construção de mapas genéticos em autopoliploides, com aplicações em cana-de-açúcar**. Universidade do Estado de São Paulo - [S.l.]. 2012.

MORENO-HAGELSIEB, G. e LATIMER, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. **Bioinformatics (Oxford, England)**, v. 24, n. 3, p. 319–24, 2008.

MORIYA, Y.; ITOH, M.; OKUDA, S.; YOSHIZAWA, A. C. e KANEHISA, M. KAAS: an automatic genome annotation and pathway reconstruction server. **Nucleic Acids Research**, v. 35, p. W182–W185, 2007.

MOROZOVA, O. e MARRA, M. a. Applications of next-generation sequencing technologies in functional genomics. **Genomics**, v. 92, n. 5, p. 255–64, 2008.

MORRELL, P. L.; BUCKLER, E. S. e ROSS-IBARRA, J. Crop genomics: advances and applications. **Nature reviews. Genetics**, v. 13, n. 2, p. 85–96, 2011.

MUIR, J. P. e JANK, L. Guineagrass. *In*: **Warm-season (C4) grasses**, p. 589–621, 2004.

MUTZ, K. et al. Transcriptome analysis using next-generation sequencing. **Current opinion in biotechnology**, v. 24, n. 1, p. 22–30, 2013.

NAKAJIMA, K.; KOMATSU, T.; MOCHIZUKI, N. e SUZUKI, S. Isolation of diploid and tetraploid sexual plants in guinea grass (*Panicum maximum* Jacq.). **Japanese Journal of Breeding**, v. 29, 1979.

NALLATHAMBI GUNASEELAN, V. Anaerobic digestion of biomass for methane production: A review. **Biomass and Bioenergy**, v. 13, n. 1-2, p. 83–114, 1997.

NESBITT, T. C. e TANKSLEY, S. D. Comparative sequencing in the genus *Lycopersicon*: implications for the evolution of fruit size in the domestication of cultivated tomatoes. **Genetics**, v. 162, n. 1, p. 365–379, 2002.

NORDBORG, M. e WEIGEL, D. Next-generation genetics in plants. **Nature**, v. 456, n. 7223, p. 720–723, doi:10.1038/nature07629, 2008.

OKADA, M.; LANZATELLA, C.; SAHA, M. C.; BOUTON, J.; WU, R. e TOBIAS, C. M. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. **Genetics**, v. 185, n. 3, p. 745–60, 2010.

OLIVEIRA, K. M.; PINTO, L. R.; MARCONI, T. G.; MARGARIDO, G. R. A.; PASTINA, M. M.; TEIXEIRA, L. H. M.; FIGUEIRA, A. V.; ULIAN, E. C.; GARCIA, A. A. F. e

- SOUZA, A. P. Functional integrated genetic linkage map based on EST-markers for a sugarcane (*Saccharum* spp.) commercial cross. **Molecular Breeding**, v. 20, n. 3, p. 189–208, 2007.
- PATEL, R. K. e JAIN, M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. **PloS One**, v. 7, e30619, 2012.
- PENG, L.; KAWAGOE, Y.; HOGAN, P. e DELMER, D. Sitosterol-beta-glucoside as primer for cellulose synthesis in plants. **Science**, v. 295, p. 147–150, 2002.
- POLAND, J. A. e RIFE, T. W. Genotyping-by-Sequencing for Plant Breeding and Genetics. **The Plant Genome Journal**, v. 5, n. 3, p. 92, 2012.
- PORCEDDU, A.; ALBERTINI, E.; BARCACCIA, G.; FALISTOCCO, E. e FALCINELLI, M. Linkage mapping in apomictic and sexual Kentucky bluegrass (*Poa pratensis* L.) genotypes using a two way pseudo-testcross strategy based on AFLP and SAMPL markers. **Theoretical and Applied Genetics**, v. 104, n. 2-3, p. 273–280, 2002.
- PUPKO, T. e GRAUR, D. Evolution of microsatellites in the yeast *Saccharomyces cerevisiae*: role of length and number of repeated units. **Journal of Molecular Evolution**, v. 48, p. 313–316, 1999.
- RESENDE, R. M. S.; RESENDE, M. D. V.; JANK, L.; VALLE, C. B.; CANÇADO, L. J. e CHIARI, L. Melhoramento genético de leguminosas forrageiras. In: **Melhoramento de Forrageiras Tropicais**. Campo Grande: Embrapa Gado de Corte. p. 117-159, 2008.
- RICKERT, A. M. et al. First-generation SNP/InDel markers tagging loci for pathogen resistance in the potato genome. **Plant Biotechnology Journal**, v. 1, n. 6, p. 399–410, 2003.
- RIPOL, M. I.; CHURCHILL, G. A.; DA SILVA, J. A. G. e SORRELLS, M. Statistical aspects of genetic mapping in autopolyploids. **Gene**, v. 235, n. 1, p. 31–41, 1999.
- ROUNSLLEY, S. et al. De Novo Next Generation Sequencing of Plant Genomes. **Rice**, v. 2, n. 1, p. 35–43, 2009.
- SAGE, R.F. The evolution of C4 photosynthesis. **New Phytologist**, v. 161, p. 341–370, 2004.
- SAHA, M. C.; MIAN, R.; ZWONITZER, J. C.; CHEKHOVSKIY, K. e HOPKINS, A. A. An SSR-and AFLP-based genetic linkage map of tall fescue (*Festuca arundinacea* Schreb.). **Theoretical and applied genetics**, v. 110, n. 2, p. 323–336, 2005.
- SANGER, F. e NICKLEN, S. DNA sequencing with chain-terminating. v. 74, n. 12, p. 5463–5467, 1977.
- SANTOS, A. **Anaerobic digestion of Organic Matter**. Universidade Técnica de Lisboa, 2011.
- SAVIDAN, Y. **Nature and heridity of apomixis in *Panicum maximum***, 1982.

- SAVIDAN, Y. Genetics and utilization of apomixis for the improvement of guineagrass (*Panicum maximum* Jacq.). In: **Proceedings of the 14th International Grasslands Congress**. Lexington, Westview Press. p. 182–184, 1983.
- SAVIDAN, Y. Apomixis: genetics and breeding. **Plant Breeding Reviews**, v. 18, p. 13–86, 2000.
- SAVIDAN, Y. H. et al. Breeding *Panicum maximum* in Brazil. 1. Genetic resources, modes of reproduction and breeding procedures. **Euphytica**, v. 41, n. 1-2, p. 107–112, 1989.
- SAVIDAN, Y. e PERNÈS, J. Diploid-tetraploid-dihaploid cycles and the evolution of *Panicum maximum* Jacq. **Evolution**, p. 596–600, 1982.
- SCHLIESKY, S. et al. RNA-Seq Assembly - Are We There Yet? **Frontiers in plant science**, v. 3, p. 220, 2012.
- SCHRICK, K.; FUJIOKA, S.; TAKATSUTO, S.; STIERHOF, Y.; STRANSKY, H.; et al. A link between sterol biosynthesis, the cell wall, and cellulose in *Arabidopsis*. **The Plant Journal**, v. 38, p. 227–243, 2004.
- SCHUELKE, M. An economic method for the fluorescent labeling of PCR fragments. **Nature biotechnology**, v. 18, n. 2, p. 233–4, 2000.
- SCHULZ, M. H. et al. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. **Bioinformatics**, v. 28, n. 8, p. 1086–92, 2012.
- SCOLES, G.; SIM, S-C.; YU, J-K.; JO, Y-K.; SORRELLS, M. E. e JUNG, G. Transferability of cereal EST-SSR markers to ryegrass. **Genome**, v. 52, n. 5, p. 431–437, 2009.
- SERANG, O.; MOLLINARI, M. e GARCIA, A. A. F. Efficient exact maximum a posteriori computation for bayesian SNP genotyping in polyploids. **PloS one**, v. 7, n. 2, p. e30906, 2012.
- SHARMA, M. K.; SHARMA, R.; CAO, P.; JENKINS, J.; BARTLEY, L. E.; et al. A genome-wide survey of switchgrass genome structure and organization. **PloS One**, v. 7, e33892, 2012.
- SILVA, J. A. DA e SORRELLS, M. E. Linkage analysis in polyploids using molecular markers. **Methods of genome analysis in plants**. CRC Press, Boca Raton, Florida, p. 211–228, 1996.
- SKERMAN, P. J. e RIVEROS, F. **Tropical Grasses**. Rome: FAO, 1990.
- SLATE, J. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. **Molecular ecology**, v. 14, n. 2, p. 363–79, 2005.
- SOKHANSANJ, S.; MANI, S.; TURHOLLOW, A.; KUMAR, A.; BRANSBY, D.; et al. Large-scale production, harvest and logistics of switchgrass (*Panicum virgatum* L.) - current technology and envisioning a mature technology. **Biofuels, Bioproducts and Biorefining**, v. 3, p. 124–141, 2009.

SOLOMON, S. et al. Irreversible climate change due to carbon dioxide emissions. **Proceedings of the national academy of sciences**, v. 106, n. 6, p. 1704–1709, 2009.

SOLTIS, D. E.; SOLTIS, P. S. e RIESEBERG, L. H. Molecular data and the dynamic nature of polyploidy. **Critical Reviews in Plant Sciences**, v. 12, n. 3, p. 243–273, 1993.

SOUSA, A. C. B.; JUNGSMANN, L.; CAMPOS, T.; SFORÇA, D. A.; BOAVENTURA, L. R.; SILVA, G. M. B.; ZUCCHI, M. I.; JANK, L. e SOUZA, A. P. Development of microsatellite markers in Guineagrass (*Panicum maximum* Jacq.) and their transferability to other tropical forage grass species. **Plant Breeding**, v. 130, n. 1, p. 104–108, 2011a.

SOUSA, A. C. B.; JANK, L.; CAMPOS, T.; SFORÇA, D. A.; ZUCCHI, M. I. e SOUZA, A. P. Molecular Diversity and Genetic Structure of Guineagrass (*Panicum maximum* Jacq.), a Tropical Pasture Grass. **Tropical Plant Biology**, v. 4, n. 3-4, p. 185–202, 2011b.

SOUZA, L. **Mapa genético-molecular para Hevea brasiliensis e mapeamento de QTL's para características de impotância econômica**. Universidade Estadual de Campinas, 2012.

STEIN, J.; PESSINO, S. C.; MARTÍNEZ, E. J.; RODRIGUEZ, M. P.; SIENA, L. A.; QUARIN, C. L. e ORTIZ, J. P. A. A genetic map of tetraploid *Paspalum notatum* Flüggé (bahiagrass) based on single-dose molecular markers. **Molecular Breeding**, v. 20, n. 2, p. 153–166, 2007.

STUDER, B.; KÖLLIKER, R.; MUYLLE, H.; ASP, T.; FREI, U. e ROLDÁN-RUIZ, I. EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). **BMC plant biology**, v. 10, p. 177, 2010.

SYVÄNEN, A.-C. Accessing genetic variation: genotyping single nucleotide polymorphisms. **Nature Reviews Genetics**, v. 2, n. 12, p. 930–942, 2001.

TATUSOV, R. L.; FEDOROVA, N. D.; JACKSON, J. D.; JACOBS, A. R.; KIRYUTIN, B.; et al. The COG database: an updated version includes eukaryotes. **BMC Bioinformatics**, v. 4, p. 41, 2003.

THIEL, T.; MICHALEK, W.; VARSHNEY, R. K. e GRANER, A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). **Theoretical and Applied Genetics**, v. 106, p. 411–422, 2003.

TIWARI, K. K. e CHANDRA, A. Use of degenerate primers in rapid generation of microsatellite markers in *Panicum maximum*. **Journal of Environmental Biology**, v. 31, p. 965–968, 2010.

TOLEDO-SILVA, G.; CARDOSO-SILVA, C. B.; JANK, L. e SOUZA, A. P. *De novo* transcriptome assembly for the tropical grass *Panicum maximum* Jacq. **PLoS one**, v. 8, n. 7, p. e70781, 2013.

TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature biotechnology**, v. 28, n. 5, p. 511–515, 2010.

- VALLE, C.B.; JANK, L. e RESENDE, R.M.S. O melhoramento de forrageiras tropicais no Brasil. **Revista Ceres**. v.56, n.4, p.460-472, 2009.
- VANHOLME, R.; DEMEDTS, B.; MORREEL, K.; RALPH, J. e BOERJAN, W. Lignin biosynthesis and structure. **Plant Physiology**, v. 153, p. 895–905, 2010.
- VARSHNEY, R. K.; GRANER, A. e SORRELLS, M. E. Genic microsatellite markers in plants: features and applications. **Trends in biotechnology**, v. 23, n. 1, p. 48–55, 2005.
- VASIL, V. et al. Comparative analysis of free DNA delivery and expression into protoplasts of *Panicum maximum* Jacq. (Guinea grass) by electroporation and polyethylene glycol. **Plant Cell Reports**, v. 7, n. 7, p. 499–503, 1988.
- VICTORIA, F. C.; DA MAIA, L. C. e DE OLIVEIRA, A. C. *In silico* comparative analysis of SSR markers in plants. **BMC Plant Biology**, v. 11, p. 15, 2011.
- VIGNA, B. B. Z. **Estudos genético-moleculares em *Brachiaria humidicola* (Rendle) Schweick. (Poaceae)**. Tese (Doutorado) - Universidade Estadual de Campinas, 2010.
- VIGNAL, A. et al. A review on SNP and other types of molecular markers and their use in animal genetics. **Genetics Selection Evolution**, v. 34, n. 3, p. 275–306, 2002.
- VOORRIPS, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. **Journal of Heredity**, v. 93, n. 1, p. 77–78, 2002.
- VOS, P. et al. AFLP: a new technique for DNA fingerprinting. **Nucleic acids research**, v. 23, n. 21, p. 4407–4414, 1995.
- WANG, J. et al. Genomewide nonadditive gene regulation in *Arabidopsis* allotetraploids. **Genetics**, v. 172, n. 1, p. 507–17, 2006.
- WANG, Z.; GERSTEIN, M. e SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature reviews. Genetics**, v. 10, n. 1, p. 57–63, 2009.
- WANG, Y.; ZENG, X.; IYER, N. J.; BRYANT, D. W.; MOCKLER, T. C.; et al. Exploring the switchgrass transcriptome using second-generation sequencing technology. **PloS One**, v. 7, e34225, 2012.
- WARMKE, H. E. Apomixis in *Panicum maximum*. **American Journal of Botany**, p. 5–11, 1954.
- WELSH, J. e McCLELLAND, M. Fingerprinting genomes using PCR with arbitrary primers. **Nucleic acids research**, v. 18, n. 24, p. 7213–7218, 1990.
- WU, K. K.; BURNQUIST, W.; SORRELLS, M. E.; TEW, T. L.; MOORE, P. H. e TANKSLEY, S. D. The detection and estimation of linkage in polyploids using single-dose restriction fragments. **Theoretical and Applied Genetics**, v. 83, n. 3, p. 294–300, 1992.

YAMADA-AKIYAMA, H. et al. Analysis of expressed sequence tags in apomictic guineagrass (*Panicum maximum*). **Journal of plant physiology**, v. 166, n. 7, p. 750–61, doi:10.1016/j.jplph.2008.10.001, 2009.

YE, J.; FANG, L.; ZHENG, H.; ZHANG, Y.; CHEN, J.; et al. WEGO: a web tool for plotting GO annotations. **Nucleic Acids Research**, v. 34, p. W293–W297, 2006.

YOU, F. M.; HUO, N.; GU, Y. Q.; LUO, M-C.; MA, Y.; HANE, D.; LAZO, G. R.; DVORAK, J. e ANDERSON, O. D. BatchPrimer3: a high throughput web application for PCR and sequencing primer design. **BMC bioinformatics**, v. 9, p. 253, 2008.

ZIETKIEWICZ, E.; RAFALSKI, A. e LABUDA, D. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. **Genomics**, v. 20, n. 2, p. 176–183, 1994.

**Using genetic diversity information to establish
core collections of *Stylosanthes capitata* and
*Stylosanthes macrocephala***



Using genetic diversity information to establish core collections of *Stylosanthes capitata* and *Stylosanthes macrocephala*

Melissa Oliveira Santos-Garcia^{1#}, Guilherme de Toledo-Silva^{1#}, Rodrigo Possidonio Sasaki¹, Thais Helena Ferreira¹, Rosângela Maria Simeão Resende², Lucimara Chiari², Cláudio Takao Karia³, Marcelo Ayres Carvalho³, Fábio Gelape Faleiro³, Maria Imaculada Zucchi⁴ and Anete Pereira de Souza^{1,5}

¹Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Campinas, SP, Brazil.

²Empresa Brasileira de Pesquisa Agropecuária (Embrapa) - Gado de Corte. Campo Grande, MS, Brazil.

³Empresa Brasileira de Pesquisa Agropecuária (Embrapa) - Cerrados. Planaltina, DF, Brazil.

⁴Pólo Centro Sul -Agência Paulista de Tecnologia dos Agronegócios Piracicaba, SP, Brazil.

⁵Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil.

Abstract

Stylosanthes species are important forage legumes in tropical and subtropical areas. *S. macrocephala* and *S. capitata* germplasm collections that consist of 134 and 192 accessions, respectively, are maintained at the Brazilian Agricultural Research Corporation Cerrados (Embrapa-Cerrados). Polymorphic microsatellite markers were used to assess genetic diversity and population structure with the aim to assemble a core collection. The mean values of H_o and H_e for *S. macrocephala* were 0.08 and 0.36, respectively, whereas the means for *S. capitata* were 0.48 and 0.50, respectively. Roger's genetic distance varied from 0 to 0.83 for *S. macrocephala* and from 0 to 0.85 for *S. capitata*. Analysis with STRUCTURE software distinguished five groups among the *S. macrocephala* accessions and four groups among those of *S. capitata*. Nei's genetic diversity was 27% in *S. macrocephala* and 11% in *S. capitata*. Core collections were assembled for both species. For *S. macrocephala*, all of the allelic diversity was represented by 23 accessions, whereas only 13 accessions were necessary to represent all allelic diversity for *S. capitata*. The data presented herein evidence the population structure present in the Embrapa-Cerrados germplasm collections of *S. macrocephala* and *S. capitata*, which may be useful for breeding programs and germplasm conservation.

Keywords: *Stylosanthes*, tropical forage, microsatellites, genetic diversity, core collection.

Received: May 26, 2012; Accepted: August 30, 2012.

Introduction

The genus *Stylosanthes* Sw. (Fabaceae) consists of approximately 48 species distributed throughout the tropical regions of the Americas, Africa and Asia (Costa and Ferreira, 1984; Mannetje, 1984; Kumar and Sane, 2003). Brazil is considered the major center of *Stylosanthes* diversity comprising 45% of all the species within this genus (Ferreira and Costa, 1979; Stace and Cameron, 1984). The central region of Brazil is recognized as having the highest phenotypic variation and endemism for this genus (Costa N, 2006, PhD thesis, Universidade Técnica de Lisboa, Lisboa, Lisbon, Portugal).

Some *Stylosanthes* species are used as pasture legumes and thus have economic importance in tropical and subtropical regions (Edye and Cameron, 1984). Some of these species can also be used for soil improvement through nitrogen fixation, regeneration of degraded wastelands, and for promoting water and soil conservation (Chakraborty, 2004).

Stylosanthes macrocephala M.B. Ferr. et Sousa Costa belongs to the section *Stylosanthes* (Mannetje, 1984). It is a diploid species with $2n = 20$. This species occurs on the sandy soils of the Brazilian Cerrado and Caatinga (Costa N, 2006, PhD thesis, Universidade Técnica de Lisboa, Lisbon, Portugal), and several of its ecotypes are tolerant to anthracnose (*Colletotrichum gloeosporioides*), which is the most important disease that affects this genus (Costa and Ferreira, 1984). The perennial subshrub *S. capitata* Vog. ($2n = 40$) occurs in Brazil and Venezuela. It has both prostrate and erect forms. The plant produces a large amount of

Send correspondence to Anete Pereira de Souza. Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Caixa Postal 6010, 13083-970 Campinas, SP, Brazil. E-mail: anete@unicamp.br.

[#]These authors contributed equally to this work.

seeds and dry matter, and its inflorescences have a high nutrition value (Williams *et al.*, 1984; Costa N, 2006, PhD thesis, Universidade Técnica de Lisboa, Lisbon, Portugal).

Major collections of important crop plants are held in gene banks around the world. These collections serve as repositories of the biodiversity available for each species and thus are a valuable resource for genes useful to plant breeders. The efficient maintenance and use of germplasm are commonly restricted due to the lack of genetic information and/or by the large numbers of accessions in these collections (Virk *et al.*, 1995). Molecular markers, along with morpho-agronomic data and ecological descriptions of sampling sites have proven to be relevant for evaluating germplasm (Westman and Kresovich, 1997; Zong *et al.*, 2009). The use of molecular markers can also help to select material for establishing a core collection, *i.e.*, a group of accessions from an existing germplasm collection that is chosen to represent the genetic spectrum of the entire collection (Hao *et al.*, 2006). Microsatellites or simple sequence repeats (SSRs) have proven to be among the most suitable markers for such purposes (Huang *et al.*, 2002; Hao *et al.*, 2006; Landjeva *et al.*, 2006; Wang *et al.*, 2006; Ebana *et al.*, 2008; Blair *et al.*, 2009; Cipriani *et al.*, 2010).

In this study, we evaluated the genetic diversity and population structure in accessions of the Embrapa-Cer-

rados germplasm collections of *S. macrocephala* and *S. capitata* using polymorphic SSRs. Based on this diversity information, we determined the minimum sample size acceptable for a core collection of each species.

Materials and Methods

DNA extraction and PCR

A total of 326 accessions from the Embrapa-Cerrados germplasm collections were used in this study: 134 accessions of *S. macrocephala* and 192 of *S. capitata* (Tables 1 and 2). The SSR markers developed by Santos *et al.* (2009a) (13 SSR *S. macrocephala* loci) and Santos *et al.* (2009b) (15 SSR *S. capitata* loci) were used to assess the genetic diversity of these accessions.

Total DNA was extracted from leaves of three plants from each accession according to the cetyltrimethylammonium bromide method described by Faleiro *et al.* (2003). PCR amplifications were performed using a PTC-200 (MJ Research) thermocycler in a 20- μ L final reaction volume consisting of 1X PCR buffer, 1.5 mM MgCl₂, 0.25 mM of each dNTP (Invitrogen), 0.8 μ M of each primer, 1U *Taq* DNA polymerase (Invitrogen) and 20 ng genomic DNA. The amplification protocol consisted of an initial denaturation step at 94 °C for 1 min, followed by

Table 1 - List of 134 accessions of *S. macrocephala* from the Embrapa-Cerrados germplasm collection that were analyzed for 13 microsatellite markers. The sample codes, the respective accession numbers and BRA or CIAT numbers in the germplasm collection of Embrapa-Cerrados (CPAC) and the place of origin are shown.

Code	CPAC	BRA/CIAT	Place of origin	Code	CPAC	BRA/CIAT	Place of origin
1	139	BRA-003697	Distrito Federal	40	1341	BRA-023264	Bahia
2	1030	CIAT 1942	Unknow	41	1345	BRA-022583	Bahia
3	1031	BRA-007773	Goiás	42	1346	BRA-023523	Bahia
4	1032	BRA-007820	Goiás	43	1347	BRA-023191	Bahia
5	1033	BRA-009032	Bahia	44	1367	BRA-022284	Goiás
6	1036	BRA-003008	Bahia	45	1370	BRA-023019	Bahia
7	1037	BRA-008052	Bahia	46	1373	BRA-022781	Bahia
8	1039	CIAT 2079	Bahia	47	1376	BRA-022985	Bahia
9	1040	BRA-008184	Bahia	48	1378	BRA-023329	Bahia
10	1043	BRA-008958	Bahia	49	1382	BRA-022586	Bahia
11	1190	CIAT 2270	Minas Gerais	50	1383	BRA-022616	Bahia
12	1191	CIAT 2271	Minas Gerais	51	1636	CIAT 1413	Unknow
13	1192	BRA-012297	Minas Gerais	52	1639	BRA-012866	Bahia
14	1193	CIAT 2273	Minas Gerais	53	1640	BRA-012947	Bahia
15	1194	CIAT 2274	Minas Gerais	54	2227	BRA-013030	Bahia
16	1196	CIAT 2276	Minas Gerais	55	2229	BRA-007226	Unknow
17	1197	CIAT 2277	Minas Gerais	56	2230	BRA-028487	Goiás
18	1198	CIAT 2278	Minas Gerais	57	2231	BRA-28495	Bahia
19	1200	CIAT 2280	Minas Gerais	58	2239	BRA-022828	Bahia
20	1201	CIAT 2281	Minas Gerais	59	2254	BRA-032883	Minas Gerais
21	1202	CIAT 2282	Minas Gerais	60	2255	BRA-032891	Minas Gerais
22	1204	CIAT 2284	Minas Gerais	61	2256	BRA-032905	Minas Gerais

Table 1 (cont.)

Code	CPAC	BRA/CIAT	Place of origin	Code	CPAC	BRA/CIAT	Place of origin
23	1205	CIAT 2285	Minas Gerais	62	2257	BRA-032913	Minas Gerais
24	1206	CIAT 2286	Minas Gerais	63	2258	BRA-032921	Minas Gerais
25	1303	-	Unknow	64	2259	BRA-032930	Minas Gerais
26	1304	-	Unknow	65	2260	BRA-032948	Minas Gerais
27	1305	-	Unknow	66	2261	BRA-032956	Minas Gerais
28	1306	-	Unknow	67	2262	BRA-032364	Minas Gerais
29	1307	BRA-017124	Distrito Federal	68	2263	BRA-032972	Minas Gerais
30	1308	BRA-017639	Goiás	69	2265	BRA-032999	Minas Gerais
31	1309	BRA-017281	Distrito Federal	70	2266	BRA-033006	Minas Gerais
32	1310	BRA-017663	Distrito Federal	71	2267	BRA-033014	Minas Gerais
33	1311	BRA-017442	Goiás	72	2710	BRA-028673	Goiás
34	1332	BRA-023202	Bahia	73	2711	BRA-028720	Bahia
35	1333	BRA-023124	Bahia	74	2712	BRA-028789	Piauí
36	1335	BRA-022837	Bahia	75	2713	BRA-028878	Bahia
37	1337	BRA-023345	Bahia	76	2714	BRA-028886	Bahia
38	1339	BRA-023493	Bahia	77	2715	BRA-028908	Bahia
39	1340	BRA-022829	Bahia	78	2716	BRA-028967	Bahia
79	2719	BRA-029025	Bahia	107	2709	-	Unknow
80	2720	BRA-029041	Bahia	108	2264	BRA-032981	Minas Gerais
81	2777	BRA-006245	Distrito Federal	109	2717	BRA-028771	Bahia
82	2778	BRA-011088	Minas Gerais	110	2790	BRA-008826	Bahia
83	2782	BRA-009008	Bahia	111	2792	BRA-008257	Bahia
84	2783	BRA-008061	Bahia	112	4135	BRA-036901	Unknow
85	1035	BRA-008010	Bahia	113	4137	BRA-036889	Unknow
86	2795	BRA-034142	Minas Gerais	114	1041	BRA-008222	Bahia
87	4136	BRA-036871	Unknow	115	1189	CIAT 2231	Rio de Janeiro
88	4138	BRA-036927	Unknow	116	1045	BRA-008362	Bahia
89	4139	BRA-036919	Unknow	117	1199	CIAT 2279	Minas Gerais
90	4140	BRA-036935	Unknow	118	1187	BRA-008168	Bahia
91	4166	BRA-036862	Unknow	119	1302	-	Minas Gerais
92	4167	BRA-036820	Unknow	120	1380	BRA-022721	Bahia
93	4168	BRA-036838	Unknow	121	1641	BRA-015261	Goiás
94	4200	BRA-036897	Unknow	122	1642	BRA-015253	Goiás
95	4271	BRA-037541	Unknow	123	1646	BRA-050173	Minas Gerais
96	4378	BRA-036854	Unknow	124	2232	BRA-028509	Bahia
97	5184	BRA-042731	Minas Gerais	125	2235	BRA-022516	Bahia
98	5296	-	Unknow	126	2268	BRA-033022	Minas Gerais
99	1362	BRA-23361	Bahia	127	2721	BRA-029076	Bahia
100	208		Goiás	128	2779	BRA-007773	Goiás
101	1336	BRA-024350	Bahia	129	2780	BRA-007820	Goiás
102	1363	BRA-022641	Bahia	130	2784	BRA-011126	Bahia
103	1361	BRA-022411	Bahia	131	2789	BRA-006301	Unknow
104	1358	BRA-022501	Bahia	132	2794	BRA-034215	Minas Gerais
105	2251	BRA-023566	Minas Gerais	133	4377	BRA-036846	Unknow
106	2252	BRA-024252	Bahia	134	4971	BRA-041441	Goiás

*The geographic sites indicate the Brazilian states in which the plants were collected.

Table 2 - List of 192 accessions of *S. capitata* from the Embrapa-Cerrados germplasm collection that were analyzed for 13 microsatellite markers. The sample codes, the respective accession numbers and BRA or CIAT numbers in the germplasm collection of Embrapa-Cerrados (CPAC) and the place of origin are shown.

Code	CPAC	BRA/CIAT	Place of origin	Code	CPAC	BRA/CIAT	Place of origin
1	515	BRA-006751	Mato Grosso do Sul	40	1160	BRA-001830	Maranhão
2	650	CIAT 1405	Mato Grosso do Sul	41	1161	BRA-001856	Maranhão
3	704	BRA-007251	Minas Gerais	42	1162	BRA-001899	Piauí
4	705	CIAT 1078	Bahia	43	1163	BRA-001902	Piauí
5	706	BRA-005886	Bahia	44	1164	BRA-001911	Piauí
6	707	BRA-001791	Maranhão	45	1165	BRA-000400	Piauí
7	901	CIAT 2249	Minas Gerais	46	1166	BRA-001929	Piauí
8	906	CIAT 1419	Goiás	47	1167	BRA-001937	Piauí
9	908	CIAT 1440	Ceará	48	1168	CIAT 2220	Bahia
10	909	CIAT 1441	São Paulo	49	1169	BRA-006190	Goiás
11	913	BRA-006742	Mato Grosso	50	1170	BRA-009181	Pernambuco
12	915	CIAT 1892	Venezuela	51	1171	BRA-007544	Ceará
13	916	CIAT 1899	Venezuela	52	1172	BRA-007522	Ceará
14	918	CIAT 1924	Venezuela	53	1174	BRA-007595	Ceará
15	922	BRA-007625	Distrito Federal	54	1177	CIAT 2259	Minas Gerais
16	924	BRA-007749	Goiás	55	1178	CIAT 2260	Minas Gerais
17	925	BRA-007871	Goiás	56	1182	CIAT 2265	Minas Gerais
18	926	BRA-007846	Goiás	57	1183	CIAT 2266	Minas Gerais
19	928	BRA-007803	Goiás	58	1185	CIAT 2268	Minas Gerais
20	929	BRA-007838	Goiás	59	1186	CIAT 2269	Minas Gerais
21	931	BRA-009059	Bahia	60	1278	BRA-017787	Minas Gerais
22	934	BRA-007960	Bahia	61	1279	BRA-017795	Minas Gerais
23	935	BRA-007994	Bahia	62	1281	-	Minas Gerais
24	936	BRA-008001	Bahia	63	1282	BRA-017881	Goiás
25	938	BRA-008087	Bahia	64	1283	BRA-017043	Goiás
26	939	BRA-008176	Bahia	65	1284	BRA-017094	Goiás
27	940	BRA-008231	Bahia	66	1285	BRA-016659	Goiás
28	943	BRA-008907	Bahia	67	1286	BRA-016675	Goiás
29	944	BRA-008869	Bahia	68	1287	BRA-016713	Goiás
30	945	BRA-008818	Bahia	69	1288	BRA-016519	Goiás
31	947	BRA-008401	Bahia	70	1289	BRA-016586	Goiás
32	949	BRA-008532	Bahia	71	1290	BRA-016403	Bahia
33	950	BRA-008583	Bahia	72	1291	BRA-016186	Goiás
34	951	BRA-008621	Bahia	73	1292	BRA-016268	Goiás
35	952	BRA-008681	Bahia	74	1293	BRA-016144	Goiás
36	953	BRA-008761	Bahia	75	1294	BRA-015962	Goiás
37	956	CIAT 2218	Bahia	76	1295	BRA-022136	Goiás
38	957	CIAT 2219	Bahia	77	1296	BRA-016039	Goiás
39	959	CIAT 2228	São Paulo	78	1297	BRA-016098	Goiás
79	1298	BRA-017507	Goiás	121	2226	BRA-032859	Minas Gerais
80	1299	BRA-017566	Goiás	122	2681	BRA-029084	Bahia
81	1300	BRA-017396	Goiás	123	2682	BRA-029068	Bahia
82	1328	BRA-013517	Goiás	124	2685	BRA-028860	Bahia
83	1350	BRA-011749	Maranhão	125	2686	BRA-028851	Piauí
84	1357	BRA-13371	Maranhão	126	2687	BRA-028843	Piauí
85	1384	BRA-022314	Maranhão	127	2689	BRA-028827	Piauí
86	1386	BRA-023485	Goiás	128	2691	BRA-028738	Bahia
87	1387	BRA-024317	Bahia	129	2692	BRA-028746	Bahia

Table 2 (cont.)

Code	CPAC	BRA/CIAT	Place of origin	Code	CPAC	BRA/CIAT	Place of origin
88	1388	BRA-023299	Bahia	130	2694	BRA-028681	Goiás
89	1389	BRA-22446	Bahia	131	2695	BRA-028762	Piauí
90	1392	BRA-022772	Bahia	132	2696	BRA-028657	Goiás
91	1394	BRA-024261	Bahia	133	2835	BRA-014346	Ceará
92	1395	BRA-022373	Bahia	134	2837	BRA-014362	Ceará
93	1588	BRA-012874	Bahia	135	2839	BRA-014443	Piauí
94	1590	BRA-012955	Bahia	136	2840	BRA-035190	Piauí
95	1591	BRA-003671	Bolivia	137	2841	BRA-014397	Piauí
96	1592	BRA-013021	Sergipe	138	2842	BRA-001848	Maranhão
97	1594	BRA-013935	Sergipe	139	2844	BRA-031160	Mato Grosso
98	1596	BRA-014036	Bahia	140	4123	BRA-036137	Ceará
99	1597	BRA-014117	Bahia	141	2821	BRA-035173	Venezuela
100	1600	BRA-015202	Goiás	142	2822	BRA-012840	Bahia
101	1598	BRA-015229	Goiás	143	2823	BRA-001881	Maranhão
102	1601	BRA-015199	Goiás	144	2826	BRA-014401	Piauí
103	1608	CIAT 2829	Venezuela	145	2828	BRA-035181	Mato Grosso
104	1609	BRA-011720	Maranhão	146	2829	BRA-014532	Maranhão
105	1611	BRA-050173	Minas Gerais	147	2830	BRA-014508	Maranhão
106	1612	BRA-033219	Distrito Federal	148	2831	BRA-014281	Ceará
107	1616	BRA-028177	Goiás	149	2833	BRA-035157	Distrito Federal
108	1617	CIAT 2946	Goiás	150	2834	BRA-000850	Piauí
109	1618	BRA-040738	Colômbia	151	2798	BRA-014311	Ceará
110	2207	BRA-027961	Bahia	152	2807	BRA-005924	Minas Gerais
111	2208	BRA-028002	Goiás	153	2809	BRA-005908	Venezuela
112	2209	BRA-028053	Venezuela	154	2811	BRA-021491	Distrito Federal
113	2211	BRA-028185	Minas Gerais	155	2817	BRA-005975	Venezuela
114	2212	BRA-014265	Ceará	156	2813	BRA-001864	Maranhão
115	2213	BRA-014320	Ceará	157	2819	BRA-013455	Paraíba
116	2214	BRA-014427	Piauí	158	4125	BRA-036081	Ceará
117	2215	BRA-007579	Ceará	159	4129	BRA-036153	Piauí
118	2216	BRA-008240	Bahia	160	4130	BRA-036161	Maranhão
119	2217	BRA-028258	Bahia	161	4131	BRA-036188	Maranhão
120	2224	BRA-032832	Minas Gerais	162	4155	BRA-035971	Bahia
163	4156	BRA-035963	Bahia	178	4354	BRA-036129	Ceará
164	4158	BRA-035955	Goiás	179	4355	BRA-036145	Ceará
165	4159	BRA-036048	Bahia	180	4356	BRA-036170	Maranhão
166	4341	BRA-035939	Distrito Federal	181	4357	BRA-036196	Maranhão
167	4343	BRA-035980	Bahia	182	4359	BRA-036218	Goiás
168	4344	BRA-035998	Bahia	183	4360	BRA-036226	Goiás
169	4345	BRA-036005	Bahia	184	4362	BRA-037583	Goiás
170	4346	BRA-036013	Bahia	185	4363	BRA-037605	Goiás
171	4347	BRA-036021	Bahia	186	4364	BRA-037770	São Paulo
172	4348	BRA-036030	Bahia	187	4973	BRA-041467	Minas Gerais
173	4349	BRA-036056	Ceará	188	4974	BRA-041475	Minas Gerais
174	4350	BRA-036064	Ceará	189	4977	BRA-041505	Minas Gerais
175	4351	BRA-036072	Ceará	190	4981	BRA-041543	Bahia
176	4352	BRA-036099	Ceará	191	4982	BRA-041556	Bahia
177	4353	BRA-036111	Ceará	192	4984	BRA-041572	Bahia

*The geographic sites indicate the country or Brazilian states in which the plants were collected.

30 cycles of 94 °C for 1 min, 60 °C for 1 min and at 72 °C for 1 min, with a final extension step at 72 °C for 5 min. PCR-amplified DNA fragments were separated by electrophoresis on 6% denaturing polyacrylamide gels at 75 W for approximately 2 h and then stained with silver nitrate according to Creste *et al.* (2001). Allele scoring was done by comparison to a 10-bp DNA ladder (10-330 bp range) (Invitrogen).

Data analysis

Allele frequencies, observed and expected heterozygosities (H_O and H_E) and Roger's genetic distance modified by Wright (1978) were calculated using the Tools for Population Genetic Analysis (TFPGA) software (Miller, 1997). Population structure was inferred using STRUCTURE 2.0 software (Pritchard *et al.*, 2000), and the accessions were assigned to groups based on their genotypes. STRUCTURE uses model-based clustering in which a Bayesian approach identifies clusters based on their fit to Hardy-Weinberg and linkage equilibria.

The optimum number of populations (K) was selected after ten independent runs with a burn-in period of 300,000 and 400,000 replications using a model that does not allow for admixture or correlated allele frequencies. The procedure described by Evanno *et al.* (2005) was used to estimate the most probable number of distinct genetic groups (K) in each germplasm collection. Nei's G_{ST} among the groups defined by the STRUCTURE analysis was calculated using the software FSTAT (Goudet, 2001).

Genetic relationships among the accessions based on the genotypic data and Roger's genetic distance were estimated using a Neighbor-Joining method in DARwin 5.0 software (Perrier and Jacquemoud-Collet, 2006).

Finally, by using the software COREFINDER (Cipriani *et al.*, 2010) we assembled a core collection that should represent 100% of the genetic diversity present within the entire collection.

Results

We used SSR markers developed for *S. macrocephala* and *S. capitata* to genotype all of the accessions in germplasm collections of both species. In *S. macrocephala*, 61 alleles were identified at 13 microsatellite loci, and 51 alleles were identified at 15 loci in *S. capitata*. In *S. macrocephala* the range was 2 to 11 alleles per locus (4.7 average) (Table 3), with H_E values ranging from 0.02 to 0.85 (0.36 on average) and H_O values varying from 0.01 to 0.17 (0.08 on average), thus representing a low level of genetic diversity. With regard to the *S. capitata* descriptive data, the numbers of alleles ranged from 2 to 9 for all of the loci analyzed (3.4 on average) (Table 4); the H_E values ranged from 0.27 to 0.74 (0.50 on average), and the H_O values ranged from 0.04 to 0.87 (0.48 on average). Roger's genetic distance values among the *S. macrocephala* accessions ranged

Table 3 - The 13 microsatellite loci used for the analysis of the Embrapa germplasm collection of *S. macrocephala*. The number of alleles (N) and observed and expected heterozygosities (H_O and H_E , respectively) are indicated for each locus.

Locus	N	H_O	H_E
SM02 A5	2	0.01	0.07
SM02 A10	3	0.01	0.02
SM01 D3	2	0.02	0.04
SM02 A2	11	0.19	0.85
SM01 B11	7	0.05	0.70
SM02 C9	3	0.04	0.07
SM02 G2	2	0.07	0.26
SM02 G5	8	0.14	0.80
SM01 B5	2	0.02	0.04
SM01 B6	3	0.07	0.36
SM02 A8	3	0.16	0.54
SM02 A9	4	0.08	0.14
SM02 G3B	11	0.17	0.75
Average	4.7	0.08	0.36

Table 4 - The 15 microsatellite loci used for the analysis of the Embrapa germplasm collection of *S. capitata*. The number of alleles (N) and observed and expected heterozygosities (H_O and H_E , respectively) are indicated for each locus.

Locus	N	H_O	H_E
SC 01 TF6A	3	0.52	0.40
SC 01 C7B	5	0.87	0.59
SC 01 E10A	3	0.68	0.45
SC 01 E4	3	0.52	0.40
SC 01 E11	3	0.64	0.44
SC 01 A5	8	0.40	0.66
SC 01 E10B	3	0.72	0.46
SC 01 TG9	3	0.48	0.44
SC 01 B3	3	0.41	0.41
SC 01 TF11A	3	0.45	0.46
SC 01 TG12A	5	0.39	0.61
SC 01 A2A	3	0.28	0.27
SC 02 E12	9	0.41	0.71
SC 01 H5	5	0.04	0.54
SC 01 H6A	6	0.45	0.71
Average	4.3	0.48	0.50

from 0 to 0.83, with an average of 0.54, whereas these values ranged from 0 to 0.85 (0.50 average) for *S. capitata*.

The method of Evanno *et al.* (2005) was used to define the maximal ΔK , which was at $K = 5$ in the *S. macrocephala* germplasm collection, based on the STRUCTURE analysis (Figure 1). Cluster analysis revealed that 75 of the accessions (57%) were assigned to a single group with

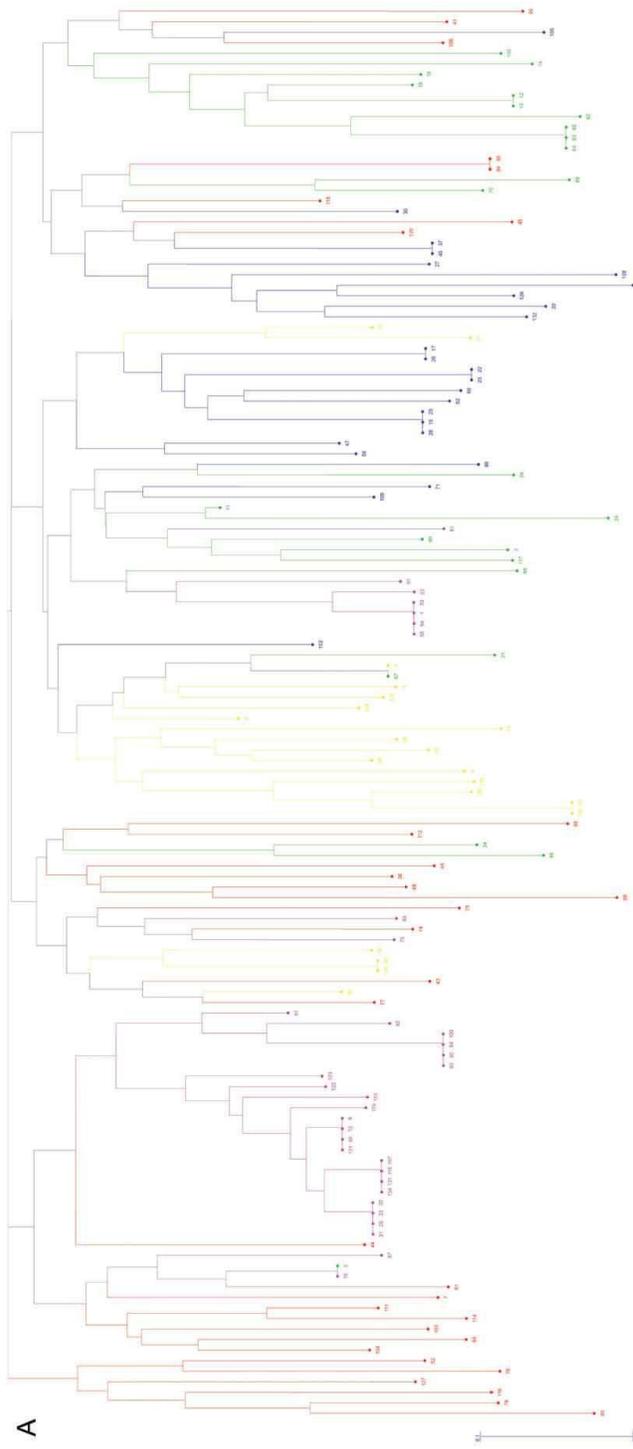


Figure 1 - Genetic diversity among *S. macrocephala* accessions. (A) As constructed from the Roger's dissimilarity matrix using the NJ method. (B) Bar plot representation of the percentage of the gene pool in each *S. macrocephala* accession.

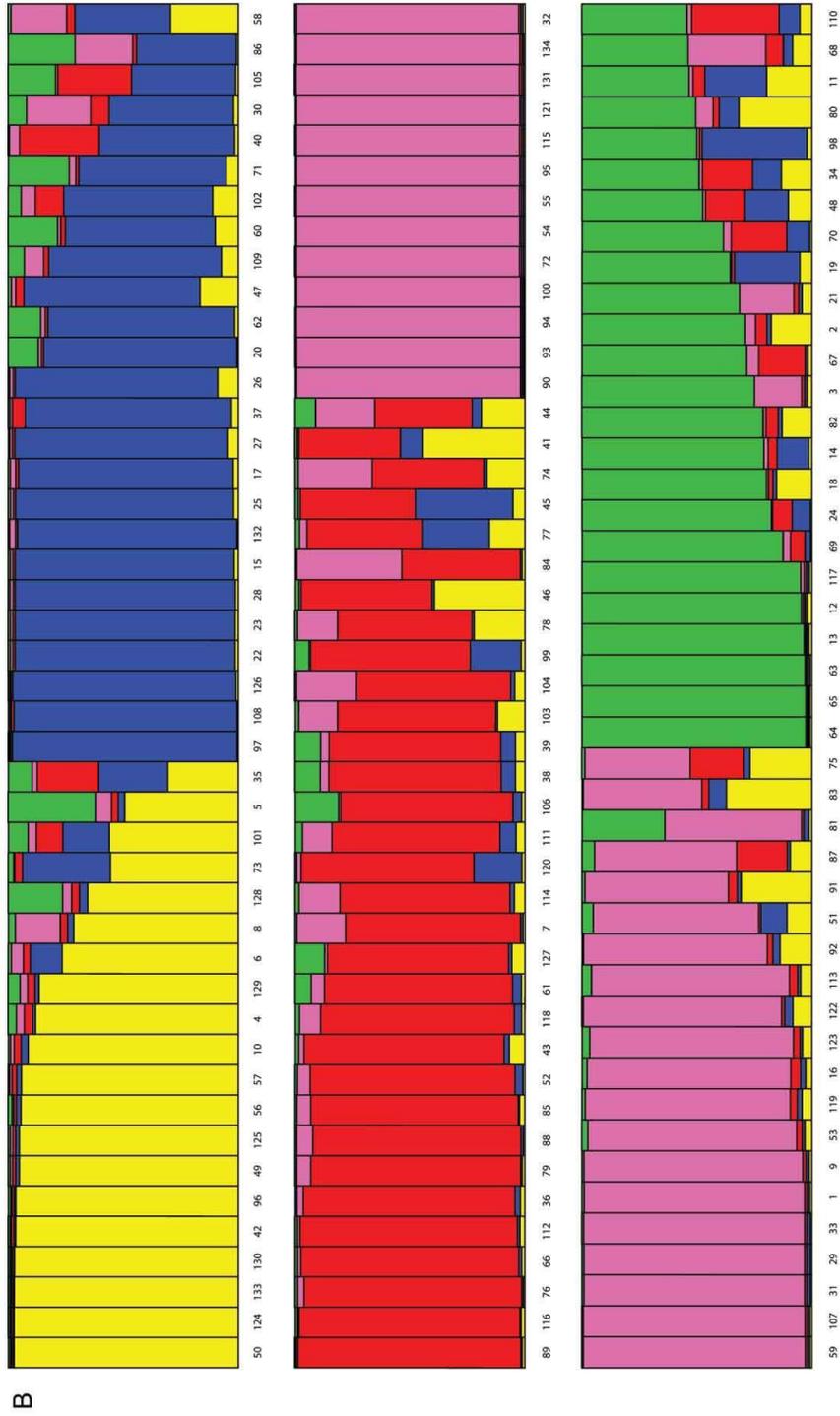


Figure 1 (cont.) - Genetic diversity among *S. macrocephala* accessions. (A) As constructed from the Roger's dissimilarity matrix using the NJ method. (B) Bar plot representation of the percentage of the gene pool in each *S. macrocephala* accession.

more than 80% probability, whereas the other 59 accessions represented a mixture of different groups. Group D comprised the largest number of non-mixed accessions, with 79% of the individuals in this cluster showing more than 80% probability of membership. In contrast, most accessions in groups C and E had less than 80% probability of membership (59% and 62%, respectively). The descriptive data calculated for the individual clusters revealed that H_O ranged from 0.03 in group D to 0.14 in group C, and that H_E values varied from 0.14 in group D to 0.38 in group C.

The STRUCTURE procedure clustered the *S. capitata* germplasm accessions into four groups (Figure 2), wherein 131 accessions (68%) were assigned to a single group with more than 80% probability of membership, and the remaining 61 accessions were so to different groups. Group D contained the largest number of accessions assigned with more than 80% membership probability (97%), whereas group A contained the highest percentage of mixed accessions (61%). H_O values ranged from 0.40 in group A to 0.56 in group C, and H_E values varied from 0.40 in group A to 0.49 in groups C and D. The Nei's genetic diversity among the groups (G_{ST}) was calculated to infer the proportion of genetic diversity due to differences among the groups clustered by STRUCTURE in both species. G_{ST} values were 27% and 11% for *S. macrocephala* and *S. capitata*, respectively.

We used DARwin software to arrive at a Neighbor-Joining (NJ) tree derived from the Roger's genetic distance results (Figures 1 and 2). In this analysis, the clusters formed by STRUCTURE with high levels of mixed accessions (less than 80% probability) became dispersed along the NJ tree.

We assembled representative core collections for both species (Figure 3), aiming to obtain 100% of the genetic diversity observed in this study. This goal was accomplished with 23 accessions of *S. macrocephala* and 13 accessions of *S. capitata*. The alleles identified in this study were fully represented in these core collections.

Discussion

The SSR markers analyzed in this work were suitable for evaluating the genetic information in the accessions of *S. macrocephala* and *S. capitata*. Santos *et al.* (2009c) observed the same range of alleles per locus (2 to 11) in 20 accessions of this same *S. macrocephala* germplasm collection, but with a smaller average of four alleles per locus. In *S. capitata*, another study observed a range of alleles per locus that varied from 2 to 7 alleles per locus and averaged 3.3 in 20 accessions of the same germplasm collection analyzed using eight microsatellites (Santos *et al.*, 2009b). In *S. guianensis* (Aubl.) Sw., the analysis of 20 loci in 20 accessions revealed allele numbers between two and seven, with an average of four (Santos *et al.*, 2009c). However, when the number of *S. guianensis* accessions was increased to 150, the number of alleles per locus was equal to the vari-

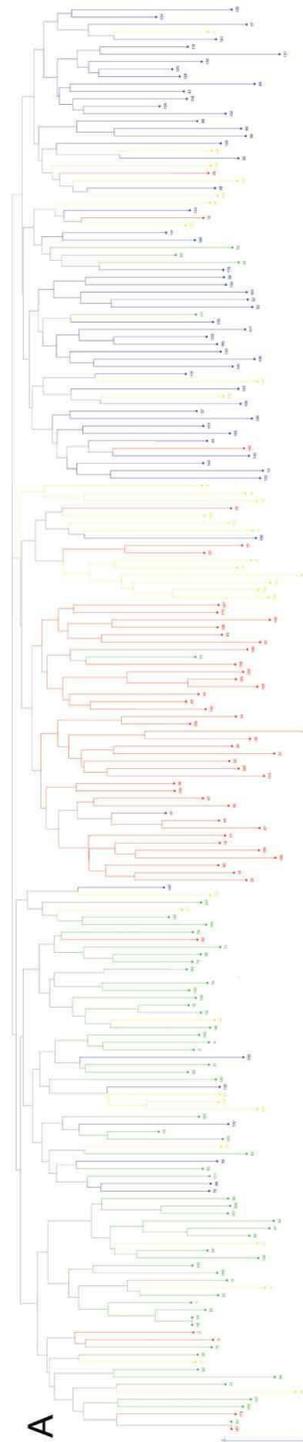


Figure 2 - Genetic diversity among *S. capitata* accessions. (A) As constructed from the Roger's dissimilarity matrix using the NJ method. (B) Bar plot representation of the percentage of the gene pool in each accession of *S. capitata*.

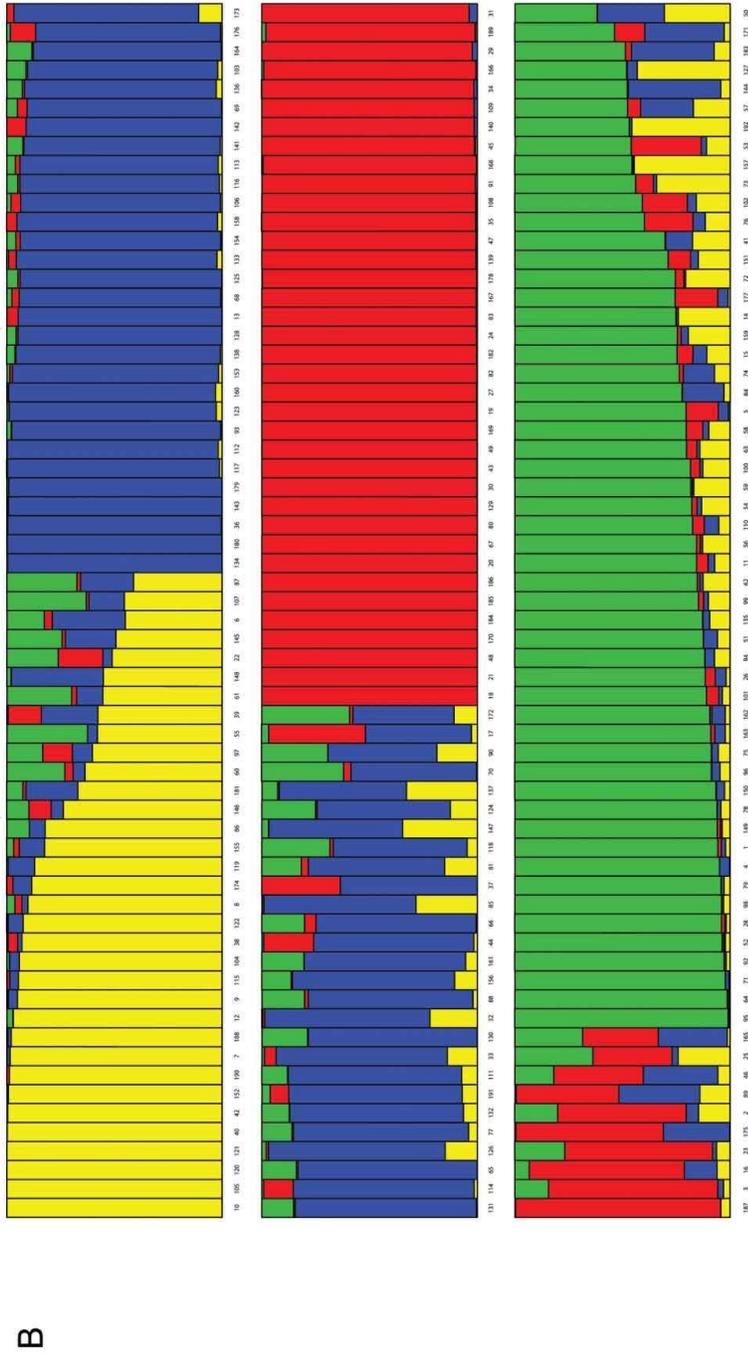


Figure 2 (cont.) - Genetic diversity among *S. capitata* accessions. (A) As constructed from the Roger's dissimilarity matrix using the NJ method. (B) Bar plot representation of the percentage of the gene pool in each accession of *S. capitata*.

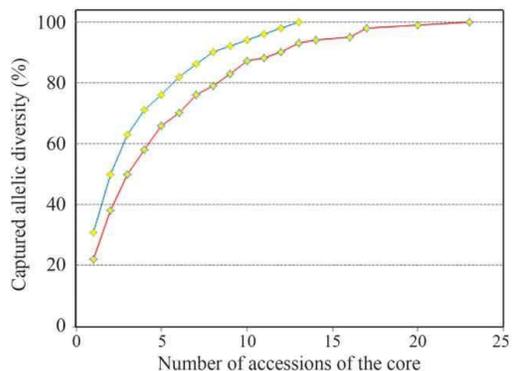


Figure 3 - Genetic diversity as a function of the number of accessions included in the *S. macrocephala* (red) and *S. capitata* (blue) core collections.

ation (2 to 11) and average (4.7) observed here for *S. macrocephala* (Santos-Garcia MO, 2009, PhD thesis, Universidade Estadual de Campinas, Campinas, Brazil). The allele sizes of *S. macrocephala* were consistent with the expected sizes reported in Santos *et al.* (2009a,c), with the exception of a few differences that occurred when higher numbers of alleles were observed for the same loci. The *S. capitata* accessions exhibited high levels of heterozygosity. Vander Stappen *et al.* (2002) showed that allotetraploid *Stylosanthes* species have high levels of fixed heterozygosity, which may explain the observed heterozygosity rates identified in the germplasm collection described for this study. As we used bulk samples, the observed heterozygosity could be explained by outcrossing and the inclusion of heterozygous individuals, or by heterogeneity in the GenBank accessions (Zhang *et al.*, 2008).

The genetic distances denoted in this study were higher than those previously reported for other species of the genus *Stylosanthes*. One possible explanation is that a larger number of accessions were analyzed here than in other studies. Furthermore, the types of molecular markers used in the previous studies were generally less polymorphic than our SSR markers. Barros *et al.* (2005) studied a subset of 86 accessions from the same *S. macrocephala* germplasm collection studied here using 15 RAPD primers and reported genetic distances ranging from 0.02 to 0.42. Hence, the microsatellite markers used herein revealed more genetic variation than the RAPD markers, similar to what has been shown in studies on other species (Powell *et al.*, 1996; Sun *et al.*, 1999; Laborda *et al.*, 2005). When evaluated using RAPD markers, the genetic dissimilarity in *S. scabra* J. Vogel was 0.06 among the accessions from Brazil, Colombia and Venezuela, and for *S. guianensis*, it averaged 0.26 among 31 accessions (Kazan *et al.*, 1993). The genetic distances among 42 *S. guianensis* accessions varied from 0.05 to 0.69 when measured using AFLP analysis (Chang-Shun *et al.*, 2004), and a recent analysis of 150

S. guianensis accessions using 20 microsatellite markers also resulted in high genetic distance values (Santos-Garcia *et al.*, 2012).

The population structure in the accessions of *S. macrocephala* and *S. capitata* was examined using STRUCTURE 2.0, which uses a Bayesian clustering approach to probabilistically assign individuals to populations based on their genotypes. The analysis of population structure using the model-based approach of Pritchard *et al.* (2000) provided support for the existence of genetic structure in these germplasm collections. Accordingly, five groups were formed among the *S. macrocephala* accessions, and four groups were formed among the *S. capitata* accessions.

The observed and expected heterozygosities were calculated considering the clusters as independent populations. Within the *S. macrocephala* groups we found that group C had the highest level of genetic diversity, whereas group D was the most homogeneous, with a low rate of heterozygosity. For *S. capitata*, the results showed no differences among groups. Such homogeneity was not unexpected because most of the accessions of the *S. capitata* collection were sampled in two locations only. When calculating the Nei's G_{ST} value among the groups formed by the STRUCTURE analysis approach, the *S. macrocephala* values were similar to other studies on species belonging to the Fabaceae family (Hamrick and Godt, 1996). In the *S. capitata* groups, the G_{ST} values were lower than those found for other *Stylosanthes* species. AFLP studies estimated a 30% variation between *S. humilis* accessions from Mexico and South America (Vander Stappen *et al.*, 2000), and another analysis on *S. humilis* H. B. K., based on AFLP, estimated 59% variation among groups. In contrast, the estimated variation among groups of *S. viscosa* (L.) Sw. was 66%, which is a higher degree of genetic difference than that observed for either of the species in our study (Sawkins *et al.*, 2001).

The sampling locations of the accessions of the *S. macrocephala* germplasm collection are listed in Table 1. The samples were collected in the Brazilian States of Bahia, Goiás, Minas Gerais, Piauí and the Distrito Federal, though information regarding the exact site of collection is lacking for several accessions.

Group A (Figure 1) consisted of accessions from Bahia and Goiás, and groups B and E included accessions from Bahia and Minas Gerais. Group C consisted mostly of accessions from Bahia, whereas group D included accessions from Bahia, Goiás, and the Distrito Federal. Barros *et al.* (2005) described 10 groups of *S. macrocephala* inferred from RAPD markers; 75% of all of the accessions were clustered into only one group, whereas seven of the remaining groups contained no more than two accessions. This clustering of 75% of the accessions into the same group limited the analysis of the genetic diversity and population structure in the *S. macrocephala* collection. Furthermore,

the grouping created difficulties for comparing the RAPD-derived clusters with those inferred from microsatellites. In this work, the Bayesian approach made it possible to identify patterns of genetic variation among five *S. macrocephala* clusters and clarified the relationships among accessions within the same RAPD cluster previously described by Barros *et al.* (2005). Our results showed that the accessions collected in Bahia were distributed throughout all five of the groups obtained with STRUCTURE and that the group consisting mostly of accessions collected in this state exhibited the highest levels of genetic diversity. Based on these results, we hypothesize that the state of Bahia might be the location of the origin of *S. macrocephala*. However, data from natural populations are necessary to confirm this hypothesis.

The sampling locations of the accessions of the *S. capitata* germplasm collection are listed in Table 2. The plants were collected in several Brazilian states, along with the Distrito Federal, and samples were also obtained from Colombia and Venezuela. The Colombia accession (CPAC 1618) is a mixture of several Brazilian accessions developed by Instituto Colombiano Agropecuario (ICA) as “Capita” variety and is considered a reference to *S. capitata*. The Capita variety was used as standard to check the phenotypic characterization of the *S. capitata* germplasm. Notwithstanding, most of the accessions were collected in Goiás and Bahia (54 and 39, respectively), representing 49% of the total collection. Groups A, B and C contained higher numbers of Bahia and Goiás accessions, whereas group B contained more samples from Bahia than from Goiás. Group D also contained several Bahia and Goiás accessions, but the majority of the accessions were from Minas Gerais. The only accession from Colombia was allocated to group B. The eight accessions from Venezuela were distributed among groups A, B, C and D, with five accessions from Venezuela clustering in group C, whereas each of the other groups contained only one accession each from this country. Group A comprised a great heterogeneity of localities, with accessions collected from all of the Brazilian states and South American countries, except for São Paulo and Colombia. Groups B and C contained the majority of the accessions from the northeastern states of Brazil and Goiás (central western region), whereas group D had more accessions from the southeastern states.

Due to sampling issues, many of the Brazilian states were poorly represented, and the genetic groups defined by STRUCTURE could not be correlated with geographic regions. Thus, for a more complete study of the genetic diversity of *S. capitata* in Brazil, new samples must be acquired, especially so from natural populations.

Using DARwin software, we constructed an NJ tree based on the Roger’s genetic distances for *S. macrocephala* (Figure 1) and *S. capitata* (Figure 2). For *S. macrocephala*, groups B and D, which contained the highest number of accessions assigned with more than 80% probability in the

STRUCTURE analysis, mostly remained clustered together in the tree. In contrast, other groups with more mixed individuals were randomly distributed along the NJ tree. Similar results were obtained for *S. capitata*, in which group A, with more mixed accessions, was also dispersed over the NJ tree. For the remaining groups, the majority of accessions clustered together in the NJ tree.

When directly compared, the results of the STRUCTURE and the NJ tree analyses revealed certain differences related to the number of groups and their genetic structure, but such differences are expected because these methods are based on distinct assumptions (Wang *et al.*, 2009). Model-based approaches, such as STRUCTURE, are more efficient than distance-based methods in discriminating genetic groups, as cluster identification is not affected by the genetic distance or graphical representation chosen (Pritchard *et al.*, 2000). Nevertheless, a combined analysis using different approaches may provide a better definition of the genetic diversity and structure in both of the *Stylosanthes* collections. Genetic diversity is the basis for genetic improvement, and consequently, knowledge about germplasm diversity has a significant impact on plant breeding (Huang *et al.*, 2002).

Costa and Schultze-Kraft (1993) performed a clustering analysis for *S. capitata* based on geographical regions and morpho-agronomic characteristics. As we used SSR markers obtained from genomic DNA, it is not possible to infer an association between the genetic markers and the phenotypic characters of the accessions. The groups obtained through molecular marker analysis are thus different from the ones obtained by Costa and Schultze-Kraft (1993), and both should be of importance to *Stylosanthes* breeders. In classical plant breeding programs, selection is done based on phenotypic evaluation, and improved progenies are obtained through crossing individuals of superior phenotypes and which, in general, are also genetically distant. Studies using molecular markers are complementary to phenotypic evaluation (Costa and Schultze-Kraft, 1993), and both are fundamental to genetic breeding programs.

Core collections were herein assembled for both *Stylosanthes* species, aiming to represent the entire genetic diversity identified in this study. The COREFINDER analysis showed that for *S. macrocephala*, 100% of the alleles found in this study could be represented by a core collection of 23 accessions. For *S. capitata*, only 13 accessions were necessary to represent 100% of the observed genetic diversity. Thus, we found that only a relatively small number of accessions were indeed necessary to represent the molecular diversity revealed in this study.

Certain factors may have contributed to the low number of accessions in the core collections suggested here. First, in terms of numbers of individuals collected in each region, the germplasm collection does not equally represent all of the distribution regions. As stated before, the germplasm collection includes some regions, such as the

state of Goiás, with 54 different accessions, while others have only few representatives. We think this unequal representation may to some extent compromise the genetic diversity present in the collection and is likely reflected in the reduced number of individuals necessary to fully represent allelic diversity. In addition, *S. capitata* is an allotetraploid species that exhibits high levels of heterozygosity, which may contribute to reducing the size of the core collection (Cipriani *et al.*, 2010). Sampling proportion and representation of base collection variation are the two most important characteristics to be observed when establishing a core collection (Hao *et al.*, 2006). Brown *et al.* (1987) suggested that the number of accessions in the core should account for 5 - 10% of the base collection, representing at least 70% of its genetic diversity. Van Hintum (1999) recommended that the sampling proportion should vary between 5% and 20% of the base collection, depending on the main objective. Both of the core collections proposed here represent 100% of the molecular diversity found in this study, with the number of accessions accounting for 17% and 7% of the base collection for *S. macrocephala* and *S. capitata*, respectively.

Our results demonstrate the great potential of using molecular data to construct a core collection and thus improve the management and utilization of the *Stylosanthes* germplasm collection of Embrapa-Cerrados. Nevertheless, because we used a relatively small number of genomic markers for the genetic analysis, the data presented here should not be used alone when deciding on which accessions from the germplasm collection should be discarded or maintained. Additional molecular markers, including more SSRs and single nucleotide polymorphisms (SNPs), should be used to provide better coverage of the genome. This information should be coupled with phenotypic data for traits of interest, such as phenology and disease resistance traits, to make a final decision on the accessions to be maintained. To initiate this effort, more genotyping and phenotyping should be initiated with the core collection proposed here and expanded to other accessions as necessary. In addition, the core collection can also be used in the selection of parents for future crosses, based both on genetic distance and phenotypic traits of the accessions.

Another issue that requires consideration is the genetic purity of the accessions used in this work. It was previously shown by our group that *S. capitata* and *S. guianensis* can cross-pollinate (Santos-Garcia *et al.*, 2010), but breeders have not accounted for cross-pollination during *Stylosanthes* seed multiplication. Here, we demonstrated a high level of heterozygosity in *S. capitata* in some undefined genetic groups obtained with STRUCTURE and the Neighbor-Joining based tree. These results might have been influenced by contaminations of the different accessions by seed multiplication plots established close to each other in the field.

In this work, we used polymorphic microsatellite markers to evaluate the genetic diversity of two *Stylosanthes* germplasm collections, and the results revealed a population structure among the accessions of both species. Our work indicates that even a small number of microsatellite markers is informative for genetic diversity studies in *Stylosanthes* species, providing a rapid and low-cost procedure for screening *Stylosanthes* germplasm collections. The results for *S. macrocephala* suggest some correlation between the region of collection and distribution among the groups based on the SSR markers. The same conclusion could not be reached for *S. capitata* because the collection does not equally represent the regions of distribution of this species in terms of quantity of accessions from each region, thereby indicating a need to improve sampling for this collection. The data from this study will certainly provide valuable information to geneticists and breeders for future improvement and conservation of *Stylosanthes* species.

Acknowledgments

The authors would like to thank Alexandre Franco Garcia for his assistance with the graphical representations and Fundação de Amparo à Pesquisa do Estado de São Paulo for financial support during the development of this project (Grant # 2005/51010-0). Both MO Santos-Garcia and G Toledo-Silva received scholarships from FAPESP (2005/52211-9, 2010/50301-9). AP Souza is awarded a research fellowship from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

- Barros AM, Faleiro FG, Karia CT, Shiratsuchi LS, Andrade RP and Lopes GK (2005) Variabilidade genética e ecológica de *Stylosanthes macrocephala* determinadas por RAPD e SIG. *Pesq Agropec Bras* 40:899-909.
- Blair MW, Díaz LM, Buendía HF and Duque MC (2009) Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor Appl Genet* 119:955-972.
- Brown AHD, Grace JP and Speer SS (1987) Designation of a "core" collection of perennial Glycine. *Soybean Genet Newslett* 14:59-70.
- Chang-Shun J, Hu-Sen J, Xin-Rong M, Dong-Mei Z and Yi-Zheng Z (2004) AFLP analysis of genetic variability among *Stylosanthes guianensis* accessions resistant and susceptible to stylo anthracnose. *Acta Bot Sinica* 46:480-488.
- Chakraborty S (2004) High-Yielding Anthracnose-Resistant *Stylosanthes* for Agricultural Systems. ACIAR Monograph No. 111, Canberra, 268 pp.
- Cipriani G, Spadotto A, Jurman I, Gaspero G, Crespan M, Meneghetti S, Frare E, Vignani R, Cresti M, Morgante M, *et al.* (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages and reveals a large admixture amongst varieties of different geographic origin. *Theor Appl Genet* 121:1569-1585.

- Costa NM and Ferreira MB (1984) Some Brazilian species of *Stylosanthes*. In: Stace HM and Edye L (eds) *The Biology and Agronomy of Stylosanthes*. Academic Press, Sydney, pp 23-48.
- Costa NM and Schultze-Kraft R (1993) Biogeografia de *Stylosanthes capitata* Vog. Y *Stylosanthes guianensis* Sw. var. *pauciflora*. *Pasturas Tropicales* 15:10-15.
- Creste S, Tulmann Neto A and Figueira A (2001) Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19:299-306.
- Ebana K, Kojima Y, Fukuoka S, Nagamine T and Kawase M (2008) Development of mini core collection of Japanese rice landrace. *Breeding Sci* 58:281-291.
- Edye L and Cameron D (1984) Prospects for *Stylosanthes* improvement and utilization. In: Stace HM and Edye L (eds) *The Biology and Agronomy of Stylosanthes*. Academic Press, Sydney, pp 571-588.
- Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol* 14:2611-20.
- Faleiro FG, Faleiro ASG, Cordeiro MCR and Karia TK (2003) DNA extraction methodology from leaves of native central Brazilian region species to operate molecular analysis. *Bol Téc Embrapa Cerrados* 92:1-5.
- Ferreira M and Costa N (1979) *Stylosanthes* Sw. Genus in Brazil. EPAMIG, Belo Horizonte, 107 pp.
- Hamrick J and Godt M (1996) Effects of life history traits on genetic diversity in plant species. *Phil Trans R Soc B* 351:1291-1298.
- Hao CY, Zhang XY, Wang LF, Dong YS, Shang XW and Jia JZ (2006) Genetic diversity and core collection evaluations in common wheat germplasm from the Northwestern Spring Wheat Region in China. *Mol Breed* 17:69-77.
- Huang XQ, Börner A, Röder MS and Ganai MW (2002) Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. *Theor Appl Genet* 105:699-707.
- Kazan K, Manners JM and Cameron DF (1993) Genetic relationships and variation in the *Stylosanthes guianensis* species complex assessed by random amplified polymorphic DNA. *Genome* 36:43-49.
- Kumar S and Sane PV (2003) *Legumes of South Asia*. Royal Botanic Gardens, Kew, 536 pp.
- Laborda PR, Oliveira KM, Garcia AA, Paterniani ME and Souza AP (2005) Tropical maize germplasm: What can we say about its genetic diversity in the light of molecular markers? *Theor Appl Genet* 111:1288-1299.
- Landjeva S, Korzun V and Ganeva G (2006) Evaluation of genetic diversity among Bulgarian Winter Wheat (*Triticum aestivum* L.) varieties during the period 1925-2003 using microsatellites. *Genet Resour Crop Evol* 53:1605-1614.
- Mannetje L't (1984) Considerations on the taxonomy of the genus *Stylosanthes*. In: Stace HM and Edye L (eds) *The Biology and Agronomy of Stylosanthes*. Academic Press, Sydney, pp 1-22.
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S and Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:1572-9788.
- Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945-959.
- Santos MO, Sasaki RP, Ferreira THS, Resende RMS, Chiari L, Karia CT, Faleiro FG, Jungmann L, Zucchi MI and Souza AP (2009a) Polymorphic microsatellite loci for *Stylosanthes macrocephala* Ferr. et Costa, a tropical forage legume. *Conserv Genet Resour* 1:481-485.
- Santos MO, Sasaki RP, Chiari L, Resende RMS and Souza AP (2009b) Isolation and characterization of microsatellite loci in tropical forage *Stylosanthes capitata* Vogel. *Mol Ecol Resour* 9:192-194.
- Santos MO, Karia CT, Resende RM, Chiari L, Jungmann L, Zucchi MI and Souza AP (2009c) Isolation and characterization of microsatellite loci in the tropical forage legume *Stylosanthes guianensis* (Aubl.) Sw. *Conserv Genet Resour* 1:43-46.
- Santos-Garcia MO, Resende RMS, Chiari L, Zucchi MI and Souza AP (2010) Mating systems in tropical forages: *Stylosanthes capitata* Vog. and *Stylosanthes guianensis* (Aubl.) Sw. *Euphytica* 178:185-193.
- Santos-Garcia MO, Karia CT, Resende RMS, Chiari L, Vieira MLC, Zucchi MI and Souza AP (2012) Identification of *Stylosanthes guianensis* varieties using molecular genetic analysis. *AoB Plants* 2012:pls001.
- Sawkins MC, Maass BL, Pengelly C, Newbury HJ, Ford-Lloyd BV, Maxted N and Smith R (2001) Geographical patterns of genetic variation in two species of *Stylosanthes* Sw. using amplified fragment length polymorphism. *Mol Ecol* 11:1947-1958.
- Stace HM and Cameron D (1984) Cytogenetics and the evolution of *Stylosanthes*. In: Stace HM and Edye L (eds) *The Biology and Agronomy of Stylosanthes*. Academic Press, Sydney, pp 49-72.
- Sun G, Díaz O, Salomon B and Von Bothmer R (1999) Genetic diversity in *Elymus caninus* as revealed by isozyme, RAPD, and microsatellite markers. *Genome* 42:420-431.
- Van Hintum TJJ (1999) The general methodology for creating a core collection. In: Johnson RC and Hodgkin T (eds) *Core Collections for Today and Tomorrow*. International Plant Genetic Resources Institute, Rome, pp 10-17.
- Vander Stappen J, Weltjens I, Gama Lopez S and Volckaert G (2000) Genetic diversity in Mexican *Stylosanthes humilis* as revealed by AFLP, compared to the variability of *S. humilis* accessions of South American origin. *Euphytica* 113:145-154.
- Vander Stappen J, Gama Lopez S, Davila P and Volckaert G (2002) Molecular evidence for the hybrid origin of a new endemic species of *Stylosanthes* Sw. (Fabaceae) from the Mexican Yucatán Peninsula. *Bot J Linn Soc* 140:1-13.
- Virk PS, Ford-Lloyd BV, Jackson MT and Newbury JN (1995) Use of RAPD for the study of diversity within plant germplasm collections. *Heredity* 74:170-179.
- Wang L, Guan R, Zhangxiong L, Chang R and Qiu L (2006) Genetic diversity of Chinese cultivated soybean revealed by SSR markers. *Crop Sci* 46:1032-1038.
- Wang ML, Zhu C, Barkley NA, Chen Z, Erpelding JE, Murray SC, Tuinstra MR, Tesso T, Pederson GA and Yu J (2009) Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theor Appl Genet* 120:13-23.

- Westman AL and Kresovich S (1997) Use of molecular marker techniques for description of plant genetic variation. In: Calow JA, Ford-Lloyd BV and Newbury JH (eds) *Biotechnology and Plant Genetic Resources: Conservation and Use*. CAB International, Wallingford, pp 9-48.
- Williams RJ, Reid R, Schultze-Kraft R, Costa NM and Thomas BD (1984) Natural distribution of *Stylosanthes*. In: Stace HM and Edye L (eds) *The Biology and Agronomy of Stylosanthes*. Academic Press, Sydney, pp 73-101.
- Wright S (1978) *Evolution and the Genetics of Populations*, volume 4: Variability within and Among Natural Populations. University of Chicago Press, Chicago, 590 pp.
- Zhang X, Blair MW and Wang S (2008) Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeat markers. *Theoretical and Applied Genetics* 117:629-40.
- Zong X, Redden RJ, Liu Q, Wang S, Guan J, Liu J, Xu Y, Liu X, Gu J, Yan L, *et al.* (2009) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. *Theor Appl Genet* 118:193-204.

Internet Resources

- Goudet J (2001) FSTAT, ver. 1.2: a computer program to calculate F-statistics. <http://www2.unil.ch/popgen/softwares/fstat.htm> (Mar 15, 2011).
- Miller MP (1997) Tools for Population Genetic Analyses (TFPGA), ver. 1.3. A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author. http://www.marksgeneticsoftware.net/_vti_bin/shtml.exe/tfpga.htm (Febr 22, 2011).
- Perrier X and Jacquemoud-Collet JP (2006) DARWIN software. <http://www.darwin.cirad.fr/darwin> (Mar 15, 2011).

Associate Editor: Everaldo Gonçalves de Barros

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.