KARINE DA COSTA BERNARDINO

MAPEAMENTO DE QTLS, ASSOCIAÇÃO GENÔMICA AMPLA E SELEÇÃO GENÔMICA PARA ESTRESSES ABIÓTICOS EM SORGO

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

Orientador: Pedro Crescêncio Souza Carneiro

Coorientador: Jurandir Vieira de Magalhães

VIÇOSA - MINAS GERAIS
2019

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

T
B523m
2019

Bernardino, Karine da Costa, 1989-
Mapeamento de QTLs, associação genômica ampla e seleção genômica para estresses abióticos em sorgo / Karine da Costa Bernardino. – Viçosa, MG, 2019.
141 f. : il. (algumas color.) ; 29 cm.

Texto em português e inglês.
Orientador: Pedro Crescêncio Souza Carneiro.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Sorgo - Melhoramento genético. 2. Locos de características quantitativas. 3. Estudo de associação genômica ampla. 4. Genômica. 5. Stress (Fisiologia). I. Universidade Federal de Viçosa. Departamento de Biologia Geral. Programa de Pós-Graduação em Genética e Melhoramento. II. Título.

CDD 22. ed. 633.1742

KARINE DA COSTA BERNARDINO

MAPEAMENTO DE QTLS, ASSOCIAÇÃO GENÔMICA AMPLA E SELEÇÃO GENÔMICA PARA ESTRESSES ABIÓTICOS EM SORGO

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

APROVADA: 30 de agosto de 2019

Assentimento:

Karine da Costa Bernardino
Autor

Pedro Crescêncio Souza Carneiro
Orientador

A Deus e aos meus familiares,

dedico.

AGRADECIMENTOS

A Deus, por ser meu guia;

À minha família, por estar sempre ao meu lado;

Ao meu namorado, por me dar ânimo nos momentos difíceis;

A Embrapa Milho Sorgo, por possibilitar que esse trabalho fosse realizado;

Aos órgãos de fomento, uma vez que o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, da Fundação de Amparo à Pesquisa do Estado de Minas Gerais, (FAPEMIG) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq);

A todos os professores e colegas da Universidade Federal de Viçosa, pelos ensinamentos e companheirismo;

Ao Dr Jurandir por todos os ensinamentos teóricos e pela orientação que foi muito além de uma orientação profissional. Hoje vejo que muitos conselhos foram para vida e sou muito grata por isso;

A Drª Maria Marta, que foi ouvinte, conselheira, amiga e que hoje é minha referência de profissional. Um dia ainda vou saber pelo menos uns 10% do que ela sabe de estatística, genômica e R;

Ao Dr Schaffert, por acreditar que um dia chegaria onde estou hoje, mesmo quando ainda estava na graduação, por me apresentar a profissão de melhorista, e por sempre estar disposto a me ajudar no que fosse necessário;

Ao Prof. Pedro Crescêncio por me apresentar a biometria, conteúdo que adorei estudar durante o doutorado e que com certeza vai fazer parte da minha vida profissional;

A todos os meus amigos da bioinformática e do NBA: Bárbara Godinho, Camila, Isadora, Karla, Michele, Laiane, Natanael, Pedro César e Zé Henrique. Vocês foram fundamentais nessa conquista, podem ter certeza disso;

A minha psicóloga Caroline, por me fazer enxergar que a caminhada pode ser mais leve quando estamos abertos a mudanças e que muitas pedras nos mesmos colocamos em

nosso caminho;

Aos mestres de muay thai, Márcio e Mirelle, que durante o doutorado se tornaram meus amigos, fazendo que a minha caminhada fosse mais leve, sou muita grata pela torcida e por cada palavra de incentivo;

E a todos aqueles que não foram citados, mas que de alguma maneira contribuíram para essa conquista.

BIOGRAFIA

Karine da Costa Bernardino, filha de Marco Antônio da Costa Bernardino e de Cicele de Fátima Costa Bernardino, nasceu em 04 de novembro de 1989, em Sete Lagoas, no estado de Minas Gerais, Brasil.

Em dezembro de 2011, formou-se em Ciências Biológicas – Licenciada, pelo Centro Universitário de Sete Lagoas (UNIFEMM), Minas Gerais. Durante a graduação, desenvolveu atividades de estágio na linha de pesquisa de Melhoramento vegetal, na Empresa Brasileira de Pesquisa Agropecuária – Embrapa Milho e Sorgo, em Sete Lagoas, Minas Gerais.

Em agosto de 2012, ingressou no curso de Pós Graduação em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, Minas Gerais, submetendo-se a defesa de dissertação em novembro de 2014.

Em agosto de 2015, ingressou no curso de Pós Graduação em Genética e Melhoramento da Universidade Federal de Viçosa, em Viçosa, Minas Gerais, submetendo-se a defesa de teste em agosto de 2019.

SUMÁRIO

RESUMO

BERNARDINO, Karine da Costa, D.Sc., Universidade Federal de Viçosa, agosto de 2019. **Mapeamento de QTLs, associação genômica ampla e seleção genômica para estresses abióticos em sorgo.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientador: Jurandir Vieira de Magalhães.

O cultivo de sorgo em solos ácidos, caracterizados por múltiplos estresses como deficiência de fósforo (P) e toxidez por alumínio (Al), relevou uma necessidade pelo desenvolvimento de cultivares adaptados, uma vez que a susceptibilidade a estas condições acarreta reduções na produção de grãos. Nesse contexto, este estudo almejou elucidar o controle genético da resposta a condições baixa disponibilidade de P e níveis tóxicos de Al em sorgo. Assim, buscou-se: i) identificar regiões genômicas, via mapeamento associativo (GWAS) e mapeamento de QTLs (*Quantitative trait loci)*, associadas a características relacionadas à eficiência no uso de P, produção de grãos e morfologia radicular em baixo P, bem como à tolerância ao Al, utilizando uma população RILs (396 linhagens) e uma população multiparental de acasalamento ao acaso (200 progênies em ciclo S2); ii) selecionar progênies S2 superiores de acordo com a presença de alelos favoráveis para SNPs que no GWAS foram associados a características de interesse; iii) realizar análises de seleção genômica sem e com genes de efeito maior. RILs e progênies S2 foram genotipadas via genotipagem por sequenciamento (GBS) e via Kasp para SNPs gene-específicos para os genes *SbPSTOL1* e *SbMATE* (relacionados à eficiência na aquisição de P e à tolerância ao Al em sorgo, respectivamente). Ambas as populações apresentaram variabilidade genética para todas as características avaliadas, com valores de herdabilidade entre 0,34 e 0,83. Constatou-se que a eficiência na aquisição de fósforo é o principal componente na eficiência do uso de P em sorgo. As co-localizações entre genes *SbPSTOL1* e QTLs para eficiência de P nas RILs, sugeriram que os genes *SbPSTOL1* são uma ferramenta promissora para desenvolvimento de cultivares superiores em condições de estresse de P. A população BRP13R foi considerada um recurso polivalente, útil para a descoberta de genes e para a seleção de progênies que acumulem alelos favoráveis em múltiplos locos, ocasionando uma adaptação mais ampla a diversas condições de estresse. Além disso, a população BRP13R é indicada para estudos de seleção genômica usando modelos que acomodem efeitos aditivos e dominantes, sendo a acurácia preditiva aprimorada pela adição de marcadores de efeito fixo em desequilíbrio de ligação com os genes principais.

Palavras-chave: Mapeamento de QTLs. GWAS. Seleção genômica. Sorgo. Estresse abiótico.

ABSTRACT

BERNARDINO, Karine da Costa, D.Sc., Universidade Federal de Viçosa, August, 2019. **QTL mapping, genome-wide association mapping, and genomic selection for abiotic stress tolerance in sorghum.** Adviser: Pedro Crescêncio Souza Carneiro. Co-adviser: Jurandir Vieira de Magalhães.

Sorghum cultivation on acidic soils is limited by phosphorus deficiency (P) and aluminum (Al) toxicity, which reduce grain yield. Hence, the development of sorghum cultivars adapted to acidic soils is needed. In the context, this study aimed at elucidating the genetic control of a P efficiency and Al tolerance, and to explore genomic selection as a tool to enhance sorghum adaptation to acidic soils. With this purpose we: i) identified via genome-wide association mapping (GWAS) and Quantitative Trait Loci (QTL) mapping genomic regions associated with phosphorus efficiency, focusing on grain yield under low P availability in the soil and root morphology traits, in addition to Al tolerance. Both a recombinant inbred line population (396 lines) and half-sib progeny derived from a multiparental population (BRP13R, with 200 S2 progeny) were used; ii) selected superior S2 progeny based on the presence of favorable alleles for SNPs associated with the target traits; and iii) performed genomic selection with and without major-effect SNPs as cofactors. Both populations were genotyped by genotyping-by-sequencing (GBS) and with tag markers specific to *SbPSTOL1* and *SbMATE*, which have been previously shown to influence P efficiency and Al tolerance, respectively. Phenotypic analyses revealed the presence of genetic variability for all characteristics evaluated in the two populations, the heritabilities varied from 0.34 to 0.83. Phosphorus acquisition efficiency was the main component of phosphorus use efficiency in sorghum. We observed co-localization between P efficiency QTLs *SbPSTOL1* genes in the RIL population, suggesting that *SbPSTOL1* genes can be used to speed up cultivar development targeting low-P soils. The random mating population, BRP13R, was found to be a multipurpose resource useful both for gene discovery and selection of progeny accumulating favorable alleles at multiple loci, which can lead to broader adaptation to multiple stress conditions. BRP13R was also found to be amenable for genomic selection approaches using models accommodating both additive and dominant effects, and prediction accuracies can be enhanced by adding as fixed effects marker in linkage disequilibrium with major genes.

Keywords: QTLs mapping. GWAS. Genomic selection. Sorghum. Abiotic stress.

INTRODUÇÃO GERAL

Estresses abióticos, em solos caracterizados por pH inferior a 5, limitam a produção agrícola mundial, com destaque para as culturas em que o produto final é o grão (Kochian et al. 2004). Aproximadamente 30% de toda área terrestre e 50% das terras mundiais, potencialmente agricultáveis, são constituídas por solos ácidos (von Uexküll and Mutert 1995), os quais apresentam níveis tóxicos de alumínio (Al), manganês (Mn) e ferro (Fe), e deficiência de elementos minerais essenciais, como o fósforo (P) (Kochian et al. 2004).

O P é um elemento essencial à vida por ser necessário na formação dos ácidos nucléicos e fosfolipídios, por atuar no metabolismo do carbono e na ativação de várias enzimas (Lambers et al. 2006), sendo um componente chave para a fotossíntese (Vance et al. 2003). Mais da metade das terras agricultáveis existentes no mundo apresentam solos caracterizados por baixa disponibilidade de P (Lynch 2011), devido a sua alta fixação aos óxidos de ferro e Al na fração argilosa desses solos (Shen et al. 2011). A baixa disponibilidade de P em solos tropicais aumenta os custos de produção devido à necessidade do uso de adubação fosfatada, a qual apresenta como matéria-prima recursos naturais não renováveis, as rochas fosfáticas. Isso dificulta a produção agrícola por pequenos produtores, e limita o patamar de produção em condições de alta tecnologia.

Em solos tropicais, como os solos do cerrado brasileiro, o cultivo de sorgo (*Sorghum bicolor*) é afetado pela ação conjunta de estresses causados por baixo pH, deficiência de P e toxidez por Al, o que representa uma barreira para a expansão da cultura no país (Schaffert et al. 2001). Em relação à toxidez por Al, a inibição do crescimento radicular é um dos sintomas que caracterizam a susceptibilidade a esse estresse, ocasionando redução na absorção de água e de nutrientes minerais e, consequentemente, perda de vigor da cultura (Singh et al. 2017). A combinação entre toxidez de Al e baixa disponibilidade de P em solos ácidos intensifica as

perdas em produtividade das culturas cultivadas em solos ácidos, como são as extensas áreas do cerrado brasileiro.

Programas de melhoramento de sorgo para cultivo em solos ácidos buscam o desenvolvimento de cultivares que sejam produtivos mesmo quando submetidos a múltiplos estresses de natureza abiótica. No entanto, as técnicas de melhoramento tradicional demandam muito tempo para a obtenção de materiais que incorporem, simultaneamente, essas características, uma vez que os mecanismos de reposta aos estresses abióticos são governados por genes altamente influenciados pelo meio, levando a ganhos genéticos reduzidos por ciclo de seleção. Sendo assim, a incorporação de técnicas de melhoramento molecular, como mapeamento de QTLs, mapeamento associativo e seleção genômica, pode ser uma estratégia para minimizar o tempo necessário para obtenção de ganhos genéticos em programas de melhoramento para condições de estresse abiótico.

REFERÊNCIAS BIBLIOGRÁFICAS

Kochian L V., Hoekenga OA, Piñeros MA (2004) How Do Crop Plants Tolerate Acid Soils? Mechanisms of Aluminum Tolerance and Phosphorous Efficiency. Annu Rev Plant Biol 55:459–493 . doi: 10.1146/annurev.arplant.55.031903.141655

Lambers H, Shane MW, Cramer MD, et al (2006) Root structure and functioning for efficient acquisition of phosphorus: Matching morphological and physiological traits. Ann Bot 98:693–713 . doi: 10.1093/aob/mcl114

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041–1049 . doi: 10.1104/pp.111.175414

Schaffert RE, Alves VMC, Pitta GVE, et al (2001) Genetic variability in sorghum for P efficiency and responsiveness. In: Horst WJ, Schenk MK, Bürkert A, et al. (eds) Plant Nutrition. Developments in Plant and Soil Sciences. Springer, Dordrecht, pp 72–73

Shen J, Yuan L, Zhang J, et al (2011) Phosphorus Dynamics: From Soil to Plant. Plant Physiol 156:997–1005 . doi: 10.1104/pp.111.175232

Singh S, Tripathi DK, Singh S, et al (2017) Toxicity of aluminium on various levels of plant cells and organism: A review. Environ Exp Bot 137:177–193 . doi: 10.1016/j.envexpbot.2017.01.005

Vance CP, Uhde-Stone C, Allan DL (2003) Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource. New Phytol 157:423–447 . doi: 10.1046/j.1469-8137.2003.00695.x

von Uexküll HR, Mutert E (1995) Global Extent, Development and Economic-Impact of Acid Soils. Plant Soil 171:1–15 . doi: 10.1007/BF00009558

CAPÍTULO 1

[1]THE GENETIC ARCHITECTURE OF PHOSPHORUS EFFICIENCY IN SORGHUM INVOLVES PLEIOTROPIC QTL FOR ROOT MORPHOLOGY AND GRAIN YIELD UNDER LOW PHOSPHORUS AVAILABILITY IN THE SOIL

[1]Artigo publicado na Revista BMC Plant Biology (https://doi.org/10.1186/s12870-019-1689-y).

RESUMO

BERNARDINO, Karine da Costa, D.Sc., Universidade Federal de Viçosa, agosto de 2019. **A arquitetura genética para eficiência na aquisição de fósforo em sorgo envolve QTL pleiotrópico para morfologia radicular e produção de grãos em solo com baixa disponibilidade de fósforo.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientador: Jurandir Vieira de Magalhães.

A fixação de fósforo (P) em óxidos de alumínio (Al) e ferro (Fe) em argilas do solo restringe a disponibilidade de P para culturas cultivadas em solos tropicais altamente intemperizados, comuns em países em desenvolvimento, tornando a deficiência de P um grande obstáculo para a segurança alimentar global. A partir do mapeamento de QTL (*Quantitative Trait Locus*) estudou-se a arquitetura genética da eficiência de P e explorou-se a importância dos caracteres radiculares na produtividade de grãos de sorgo em solo tropical com baixo P. A eficiência de aquisição de P foi o componente mais importante para eficiência de uso, sendo ambas características altamente correlacionadas com a produtividade de grãos em condições de baixa disponibilidade desse macronutriente. A área da superfície radicular foi positivamente associada com rendimento de grãos. O genitor SC283 contribuiu com 58% de todos os alelos favoráveis detectados pelo mapeamento uni-característico. Com o mapeamento para múltiplas características foram detectados 14 QTLs para rendimento de grãos e/ou morfologia radicular. QTLs fortemente ligados ou pleiotrópicos para área de superfície de raízes finas (1-2 mm de diâmetro) e produção de grãos foram detectados nas posições 1-7 mega pares de bases (Mb) e 71 Mb no cromossomo 3, respectivamente, e um QTL para diâmetro de raiz/produção de grãos foi detectado a 3 Mb no cromossomo 7. Todos estes QTLs estavam próximos a genes de sorgo homólogos ao gene *OsPSTOL1*, gene de arroz que codifica uma proteína quinase serina/treonina relacionada à eficiência de aquisição de P em condição de estresse de P. Os genes *SbPSTOL1* no cromossomo 3, *Sb03g006765* a 7 Mb e *Sb03g031690* a 60 Mb foram mais expressos no SC283, que doou os alelos favoráveis em todos os QTLs encontrados nas proximidades dos genes *SbPSTOL1*. O gene de tolerância ao Al, *SbMATE*, também pode ter influencia num QTL para rendimento de grãos na porção terminal do cromossomo 3. Outro gene semelhante ao *PSTOL1*, *Sb07g02840*, parece aumentar o rendimento de grãos através de pequenos aumentos no diâmetro radicular. Análises de co-localização sugeriram um papel para outros genes, como os homólogos de sorgo para a enzima ubiquitina-conjugada (E2) de *Arabidopsis* e para o fosfato 2 (PHO2), os quais podem estar relacionados ao aumento na produção de grãos, nestes casos proporcionado, pelo alelo doado pela linhagem elite BR007. Ressalta-se que fatores genéticos que possibilitam maior área de superfície radicular e suaves

aumentos no diâmetro da raiz fina podem favorecer a absorção de P, aumentando assim o rendimento de grãos em baixa disponibilidade de P no solo. Além disso, marcadores moleculares para genes *SbPSTOL1* e para QTLs que aumentam o rendimento de grãos viabilizam a definição de estratégias de melhoramento genético visando o desenvolvimento de cultivares de sorgo adaptadas a solos com baixo teor de P.

Palavras-chave: Deficiência de fósforo. Estresse de fósforo. Solos ácidos. Arquitetura do sistema radicular.

ABSTRACT

BERNARDINO, Karine da Costa, D.Sc., Universidade Federal de Viçosa, August, 2019. **The genetic architecture of phosphorus efficiency in sorghum involves pleiotropic QTL for root morphology and grain yield under low phosphorus availability in the soil**. Adviser: Pedro Crescêncio Souza Carneiro. Co-adviser: Jurandir Vieira de Magalhães.

Phosphorus (P) fixation on aluminum (Al) and iron (Fe) oxides in soil clays restricts P availability for crops cultivated on highly weathered tropical soils, which are common in developing countries. Hence, P deficiency becomes a major obstacle for global food security. We used multi-trait quantitative trait locus (QTL) mapping to study the genetic architecture of P efficiency and to explore the importance of root traits on sorghum grain yield on a tropical low-P soil. P acquisition efficiency was the most important component of P efficiency, and both traits were highly correlated with grain yield under low P availability. Root surface area was positively associated with grain yield. The guinea parent, SC283, contributed 58% of all favorable alleles detected by single-trait mapping. Multi-trait mapping detected 14 grain yield and/or root morphology QTLs. Tightly linked or pleiotropic QTL underlying the surface area of fine roots (1-2 mm in diameter) and grain yield were detected at positions 1-7 mega base pairs (Mb) and 71 Mb on chromosome 3, respectively, and a root diameter/grain yield QTL was detected at 3 Mb on chromosome 7. All these QTLs were near sorghum homologs of the rice serine/threonine kinase, *OsPSTOL1*. The *SbPSTOL1* genes on chromosome 3, *Sb03g006765* at 7 Mb and *Sb03g031690* at 60 Mb were more highly expressed in SC283, which donated the favorable alleles at all QTLs found nearby *SbPSTOL1* genes. The Al tolerance gene, *SbMATE*, may also influence a grain yield QTL on chromosome 3. Another *PSTOL1*-like gene, *Sb07g02840*, appears to enhance grain yield via small increases in root diameter. Co-localization analyses suggested a role for other genes, such as a sorghum homolog of the *Arabidopsis ubiquitin-conjugating E2 enzyme*, *phosphate 2* (*PHO2*), on grain yield advantage conferred by the elite parent, BR007 allele. Genetic determinants conferring higher root surface area and slight increases in fine root diameter may favor P uptake, thereby enhancing grain yield under low-P availability in the soil. Molecular markers for *SbPSTOL1* genes and for QTL increasing grain yield by non-root morphology-based mechanisms hold promise in breeding strategies aimed at developing sorghum cultivars adapted to low-P soils.

Keywords: Phosphorus deficiency. Phosphorus stress. Acid soils. Root system architecture.

Introduction

Sorghum is a versatile crop that was domesticated in the tropics, in the northeastern quadrant of the African continent, possibly at least 5000 years ago (De Wet and Harlan 1971). Along with pearl millet, sorghum is the main staple food crop of the West African Savannah zones and in that region, guinea sorghums are broadly adapted to different stresses, including those caused by poor soil fertility (Weltzien et al. 2006). In sub-Saharan Africa, two of the most important abiotic stresses that limit sorghum production are Al toxicity and low-P availability in the soil (Doumbia et al. 1993, 1998; Leiser et al. 2014).

Both types of abiotic stresses share a common chemical basis centered on the prevalence of Al and Fe oxides in the clay fraction of highly weathered tropical soils (Shaw 2001). Under low pH, Al is hydrolyzed into the ionic form, Al^{3+}, which damages plant roots, reducing crop yields (Kochian 1995). Low-P availability, in turn, results from P fixation with Al and Fe oxides (Marschner 1995). Plant roots absorb P from the soil solution in the orthophosphate forms, $H_2PO_4^-$ and HPO_4^{2-} (Vance et al. 2003). However, P fixation into soil clays impairs P diffusion from the soil solution towards the root surface, restricting uptake. Approximately half of the world agricultural lands have low-P availability (Lynch 2011). Even in high input production systems, the non-renewable nature of phosphatic rock fertilizer (Hammond et al. 2004) raises questions regarding the sustainability of continuously increasing rates of P fertilizer applications, which are needed to sustain crop yields. Therefore, in view of the prevalence of low-P soils in agricultural frontiers in which food production needs continuous improvement, such as in Africa, the identification of genetic factors that can be used to facilitate breeding for sorghum adaptation to low-P conditions become of utmost importance for global food security.

Aluminum tolerance in sorghum is due to the action of the Al-induced and Al-activated root citrate transporter, *SbMATE*, which underlies the aluminum tolerance locus,

Alt$_{SB}$, at the terminal region of sorghum chromosome 3 (Magalhaes et al. 2007). Recently, the *SbMATE* allele donated by the guinea sorghum, SC283, has been shown to enhance sorghum grain yield by over 1.0 ton ha^{-1} on an acid, Al toxic soil, with no detectable yield penalty in the absence of Al toxicity (Carvalho et al. 2016). Leiser et al. (2014), using *Alt*$_{SB}$-specific markers, also found strong associations of *SbMATE* with grain yield production, particularly in low-P conditions in many environments in West-Africa (Leiser et al. 2014). This suggests that *SbMATE* confers P use efficiency (PUE) in addition to Al tolerance, possibly via a joint effect of citrate mobilizing P that is fixed on the soil clays (Drouillon and Merckx 2003), and by enhancing root development in Al tolerant genotypes, increasing P uptake (Magalhaes et al. 2018).

The ability of a plant to grow and to produce reasonable levels of grain and biomass under low-P availability, which we designate here as P use efficiency (PUE, or simply P efficiency), can be achieved via different mechanisms acting to optimize utilization of internal P or to enhance P acquisition (Parentoni and De Souza Júnior 2008). From the crop physiology standpoint, these mechanisms may result from the modulation of P transporters, organic acid exudation, phosphatase secretion, mycorrhizae associations and alterations in root system architecture in response to low-P conditions, among other mechanisms (reviewed by López-Arredondo et al. 2014). For maize cultivated on a tropical low-P soil, P acquisition has been reported to be more important than P internal utilization to explain differences in P use efficiency (Parentoni and De Souza Júnior 2008), which was also confirmed by QTL mapping results (Mendes et al. 2014). These studies emphasize the importance of changes in root system architecture and morphology as a mechanism favoring P acquisition (reviewed by Magalhaes et al. 2017). These modifications may involve changes in lateral root growth and angle, presence of shallow roots, in addition to enhanced proliferation of root hairs (Ho et al. 2005; Lynch 2011; López-Arredondo et al. 2014).

There is a recent body of evidence suggesting that genes modulating root morphology may result in increased P efficiency. Overexpression of the rice serine/threonine receptor-like kinase, *phosphorus starvation tolerance1* (*OsPSTOL1*, Gamuyao et al. 2012) has been shown to increase grain yield in rice cultivated on a low-P soil via *OsPSTOL1*-elicited enhancement of early root growth, which favors P uptake in the developing rice plant. Subsequently, association mapping established that allelic variation at homologs of *OsPSTOL1* in sorghum, designated as *SbPSTOL1* genes, was associated with enhanced grain yield production on a low-P soil, likely via changes in root morphology, particularly root diameter and root surface area (Hufnagel et al. 2014). In addition, recent studies in Arabidopsis suggested a role in P efficiency for genes involved with Al tolerance, such as the malate transporter, *ALMT1* (Sasaki et al. 2004) and its regulatory factor, the C_2H_2-type zinc finger, sensitive to proton rhizotoxicity 1 (AtSTOP1, Iuchi et al. 2007), in addition to the ABC-like transporter, aluminum sensitive 3 (*ALS3I*, Larsen et al. 1996, 2005). These genes appear to mediate an iron-dependent mechanism leading to enhancement of lateral root growth (Belal et al. 2015; Müller et al. 2015; Balzergue et al. 2017; Dong et al. 2017; Mora-Macías et al. 2017), which can possibly increase P uptake on acidic soils (Magalhaes et al. 2018).

Using a genetic approach based on multi-trait QTL mapping, the present study aimed at unravelling the genetic architecture of P efficiency in a large sorghum recombinant inbred line population, and to establish links between the genetics and physiology of P efficiency, such as associations between root morphology, P content and sorghum grain yield on soils with low-P availability.

Methods

Genetic Material

A population composed of 396 recombinant inbred lines (RILs, $F_{10:11}$), derived from a

cross between the sorghum lines, BR007 and SC283, was developed by single-seed descent (Johnson and Bernard 1962) at Embrapa Maize and Sorghum (Sete Lagoas - MG, Brazil). Both BR007 (Redbine-type) and SC283 (sorghum converted guinea) were introduced into the Embrapa breeding program in 1972 from the Purdue Breeding Program (West Lafayette - IN, US). BR007 is Al sensitive whereas SC283 is highly tolerant to Al toxicity (Magalhaes et al. 2004). Previous studies indicated that, while SC283 has higher grain yield in a soil with low-P availability compared to BR007, the grain yield increase in BR007 in response to adequate P supply is in turn higher than in SC283 (Schaffert et al. 2001).

Phenotyping for low-P in field conditions

Four field experiments were conducted at the experimental station of Embrapa Maize and Sorghum in Sete Lagoas, State of Minas Gerais, Brazil, during the summer season of 2012 - 2013. The experimental site is a clay and highly weathered tropical soil, with low fertility in natural conditions, low pH, Al toxicity and low-P. Soil P (Mehlich 1) varied from 1 to 6 ppm between 0 and 20 cm of soil depth, and from 1 to 4 ppm at the sub-superficial soil layer (20-40 cm). The minimum and maximum content of available P in the soil (Psoil) was 5.88 kg ha^{-1} and 19.79 kg ha^{-1}.

Each experiment was arranged as a 12 x 10 alpha lattice design, with three complete replicates and ten incomplete blocks per replicate. Each block contained 12 plots, within which ten RILs (regular treatments) and the two parents (common checks) were allocated. Each plot consisted of a three-meter row, with 0.45 m between rows and 8 plants m^{-1}. Fertilization consisted of 150 kg ha^{-1} of 20-00-20 (NPK) at sowing and 200 kg ha^{-1} of urea applied 30 days after sowing.

Grain yield (Gy, kg ha^{-1}), flowering time (FT, days), plant height (PH, cm), phosphorus content in the plant (leaves and stems - Pp, kg ha^{-1}) and phosphorus content in the

grain (Pg, kg ha^{-1}) were evaluated. For P measurements, samples of plant tissues and grains were collected in each plot, weighted and then dried at 65ºC to constant weight. Dry plant tissues and grains were then weighted, grounded and homogenized. Twenty gram - subsamples were used to determine P concentration and total P content (Pt), using inductively-coupled argon plasma emission spectrometry.

The phosphorus efficiency indexes were calculated according to the methodology proposed by Moll et al. (1982), where: 1) phosphorus use efficiency (PUE) is equal to the product between phosphorus acquisition efficiency (PAE) and phosphorus internal utilization efficiency (PUTIL); 2) PAE is the total phosphorus content (Pt = Pp, P content in the plant + Pg, P content in the grain) divided by P content available in the soil; 3) PUTIL is Gy divided by Pt (Moll et al. 1982).

Root system phenotyping in low-P conditions

Root morphology traits were assessed in nutrient solutions as described by de Morais De Sousa et al. (2012) and Hufnagel et al. (2014), using a randomized block design with three replicates. Seeds were surface-sterilized using sodium hypochlorite (5%), washed with distilled water and placed in moistened paper rolls. After four days, uniform seedlings were transferred to moistened blotting papers and placed into paper pouches (24 x 33 x 0.02 cm) as described by Hund et al. (2009).

Each experimental unit consisted of one pouch, with three plants per pouch, whose bottom (3 cm) was immersed in containers filled with 5 l of the nutrient solution described in (Magnavaca et al. 1987), with pH 5.65 and a P concentration of 2.5 µM. The containers were kept in a growth chamber with 27°C day and 20°C night temperatures and a 12-hour photoperiod, under continuous aeration for 13 days.

After 13 days, root images were acquired using a digital camera Nikon D300S SLR.

Images were then analyzed using both the RootReader2D (http://www.plantmineralnutrition.net/software/rootreader2d/) and WinRhizo (http://www.regent.qc.ca/) software. The following traits were measured: root length (RL - cm); root diameter (RD - mm); total root surface area (SA - cm²); surface area of very fine roots between 0-1 mm in diameter (SA1 - cm²); surface area of fine roots between 1-2 mm in diameter (SA2 - cm²); surface area of thicker roots between 2- 4.5 mm in diameter (SA3 - cm²); root volume (RV - cm³) and volume of fine roots between 1-2 mm in diameter (V2 - cm³). Shoot dry matter (SDM) and root dry matter (RDM), phosphorus content in the shoot (Ps) and phosphorus content in the root (Pr) (in grams) were also measured.

Phenotypic analyses

Traits assessed in field and hydroponic experiments were analyzed using mixed models. For field experiments, the following model was used:

$$y_{ijkl} = \mu + E_j + R_{k(j)} + B_{l(kj)} + G_i + \varepsilon_{ijkl}$$

y_{ijkl} is the phenotypic value of individual i in the block l of the k^{th} replicate, within the experiment j; μ is the overall mean; and G_i is the genetic effect of individual i, which can be defined as:

$$G_i = \begin{cases} g_i & i = 1, \dots, n_g \\ t_i & i = n_g + 1, \dots, n_g + n_c \end{cases}$$

g_i is the random effect of RIL i, n_g is the total number of RILs; t_i is the fixed effect of check i; and n_c is the total number of checks; E_j is the fixed effect of the j^{th} experiment ($j = 1, \dots, 4$); $R_{k(j)}$ is the fixed effect of replicate k ($k = 1, \dots, 3$) in experiment j; $B_{l(kj)}$ is the random effect of block l ($l = 1, \dots, 10$) in the replicate k, within the experiment j; and $\boldsymbol{\varepsilon} = \left(\varepsilon_{1111}, \varepsilon_{2111}, \dots, \varepsilon_{IJKL}\right)'$ is a $N_{obs} \times 1$ residual random vector assumed to be normally distributed with mean zero and variance σ_ε^2, in which N_{obs} is the total number of

observations.

The model used for analyzing the hydroponic experiments was:

$$y_{ij} = \mu + B_j + g_i + \varepsilon_{ij}$$

where y_{ij} is the phenotypic value of the RIL i $(i = 1, \dots, n_g)$ in the block j; μ is the overall mean; g_i is the random genetic effect of RIL i; B_j is the fixed effect of block j $(j = 1, \dots, 3)$; and $\boldsymbol{\varepsilon} = \left(\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{IJ}\right)'$ is a $N_{obs} \times 1$ residual random vector assumed to be normally distributed with mean zero and variance σ_ε^2. Fixed and random effects were tested using the Wald statistics (Wald 1943) and the likelihood ratio test (LRT, Neyman and Pearson 1928) respectively, considering a 5% significance level (α).

For both statistical models, the genetic effect of RIL was first taken as random for estimating the genetic variance component (σ_g^2) via restricted maximum likelihood (REML), and the heritability coefficient of each trait. The effect of RIL was then considered as fixed for estimating the adjusted means using best linear unbiased estimators (BLUEs). All the mixed models analyses were performed using the GenStat software (v.17.1.0) (VSN International 2014).

Trait heritabilities were estimated as proposed by Cullis et al. (2006), called generalized heritabilities, using:

$$h^2 = 1 - \frac{\bar{v}BLUP}{2\sigma_g^2}$$

where $\bar{v}BLUP$ is the average variance of the difference between two best linear unbiased predictions (BLUPs). Person's correlation coefficients (Pearson 1895) were estimated based on the adjusted means of genotypes for traits assessed in the field and in the hydroponic experiments, using the package *Hmic* (Harrell Jr 2015) in R (R Core Team 2016).

SNP markers

Genomic DNA was isolated from approximately 500 mg of leaf tissue (eight plants per accession, i.e. RILs and their parents) as described by Saghai-Maroof et al. (1984). DNA samples were genotyped by sequencing according to Elshire et al. (2011). Reads were aligned to the version 1.4 of sorghum reference genome using Burrows-Wheeler Aligner program (BWA - Li and Durbin 2009), and the SNP calling was performed using the GBS pipeline (Glaubitz et al. 2014) implemented in the TASSEL software (Bradbury et al. 2007). Missing genotypes were imputed using the NPUTE software (Roberts et al. 2007). Then, SNP data were filtered for 40% of minor allele frequency (MAF).

QTL Mapping

The final set of traits used for multi-trait QTL mapping was comprised of grain yield (Gy), surface area of fine roots in the 1-2 mm diameter class (SA2) and root diameter (RD). Multi-trait QTL mapping analysis was performed according to the procedures described in Silva et al. (2012) implemented in R (Silva et al. 2012). For that, a multi-locus QTL mapping procedure was considered, using the Haley & Knott regression (Haley and Knott 1992) and the following linear model:

$$y_{ti} = \mu_t + \sum_{r=1}^{m} a_{tr} x_{ir} + \varepsilon_{ti}$$

where $\boldsymbol{y}_{ti}$ is the adjusted mean of RIL i $(i = 1, \dots, n_g)$ for trait t $(t = 1, \dots, T)$; μ_t is the intercept for each trait; a_{tr} is the r^{th} QTL main effect on trait t; x_{ir} represents the genotype of RIL i for the SNP marker r $(r = 1, \dots, n_M)$, being n_M the total number of markers; x_{ir} assumed values equal to 0 or 2 for RILs with homozygous genotypes for the allele donated by BR007B or SC283, respectively; and $\boldsymbol{\varepsilon}_i = (\varepsilon_{1i}, \varepsilon_{2i}, \dots, \varepsilon_{Ti})'$ is a $T \times 1$ random vector assumed to be independent and identically distributed according to a multivariate normal

distribution with mean vector zero and positive definite symmetric variance-covariance matrix Σ_ε, i.e. $\boldsymbol{\varepsilon}_i \sim MVN(0,\ \Sigma_\varepsilon)$. Single-trait QTL mapping analyses were performed for each trait, using the above model (t = 1 gives an univariate regression model).

Multiple QTL models were built based on a forward-selection procedure, testing the significance of a putative QTL main effect at each SNP position along the genome. Significance of QTLs main effects were tested using the Score Statistic (Zou et al. 2004), considering a 10% significance level (α). According to simulations performed by Silva et al. (2012), this significance level maximized the QTL detection power and kept the false discovery rate (i.e. the proportion of spurious QTLs) within an acceptable level.

QTL positions were refined after the inclusion of every new QTL in the model, until no more significant QTL main effects were found. Finally, non-significant QTL effects were removed from the model in a backward-elimination procedure, such as proposed in the seemingly unrelated regression coefficients method (Zellner 1962), considering a 1% significance level.

Quantitative analysis of *SbPSTOL1* gene expression

Sorghum seedlings were grown in a modified Magnavaca nutrient solution (Magnavaca et al. 1987) containing a low-P concentration (2.5 μM P), as described in the section *Root system phenotyping in low-P conditions***.** The experiment was set up in randomized block design with three replicates and three seedlings per experimental unit (paper pouch), giving a total of nine biological replicates per genotype. After 13 days in nutrient solution, the expression profiles of the *SbPSTOL1*-like genes (*Sb03g006765*, *Sb03g031690*, and *Sb07g002840*) were assessed in the roots of the RIL parents, BR007 and SC283. Total RNA was isolated from bulked root tissues (nine roots per bulk), using the SV Total RNA Isolation System kit (Promega Corporation, Madison, WI, USA), according to the

manufacturer's instructions. Total RNA (1 μg) was used for cDNA synthesis using the High Capacity cDNA Reverse Transcription kit (Applied Biosystems, Foster City, CA, USA). Transcripts were quantified by quantitative real-time PCR (qPCR-RT), using SYBR Green technology with the ABI Prism 7500 Fast System (Applied Biosystems, Foster City, CA, USA).

Transcript relative quantification was performed with 20 ng cDNA samples and 0.02 ng for the endogenous constitutive control (18s rRNA). Primers were designed for *SbPSTOL1* and *18s rRNA* sorghum genes using the PrimerQuest tool (https://www.idtdna.com/PrimerQuest/) (Additional file 7). Calculation of relative gene expressions were performed using the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak 2008), with three technical replicates.

Results

Phenotypic analyses in the parents and RIL population

The most important trait for P efficiency within a breeding context, grain yield in the field, was assessed under low-P availability in the soil. We also estimated the relative contributions of the efficiency at which a plant acquires P from the soil (P acquisition efficiency, PAE) and also internal utilization efficiency (PUTIL), on overall P use efficiency (PUE or simply P efficiency, that encompasses both PAE and PUTIL) (Moll et al. 1982; Parentoni and De Souza Júnior 2008). Table 1 shows that PAE was the most important component influencing PUE for sorghum cultivated under low-P availability in the soil. Acquisition efficiency accounted for 82% of the variability in PUE, whereas the contribution of the PUTIL component was comparatively much smaller (18%). Therefore, we also assessed root morphology in hydroponics as changes in root morphology including increased root length can lead to enhanced P uptake and grain yield in soils with low-P availability. To gain insights into sorghum performance in hydroponics, we also assessed dry matter

accumulation (DM) and shoot and root P content.

We observed substantial genetic variance for all traits assessed in the present study, with heritability estimates ranging from 0.3 (root diameter - RD) to 0.8 (plant height - PH, Additional file 1). Traits reflecting sorghum performance grown on low-P growth media measured in nutrient solution (DM and P content) and in the field (grain yield - Gy) showed intermediate to high heritability estimates of between 0.4 and ~0.8, indicating reasonable experimental precision to detect regions of the sorghum genome associated with P efficiency. Marked transgressive segregation for grain yield in the recombinant inbreed line (RIL) population, where a maximum of 4.5 ton ha^{-1} exceeded by more than two-fold the grain yield for either parent (Additional file 1), emphasizes the rather complex, polygenic nature of P efficiency measured in sorghum cultivated under low-P availability in the soil.

We measured total root surface area (SA) of the sorghum root system and also the root surface area of roots within the diameter classes of 0-1, 1-2 and 2-4.5 mm, which are designated hereafter as very fine, fine and thicker roots, respectively. BR007 tended to exhibit greater total root surface area and had thinner roots compared to SC283 (Fig. 1a - 1d), which is due to the prevalence in BR007 of roots in the 0-1 mm diameter class (Fig. 1e - labeled SA1). These very fine roots comprise most of the root system in both parents but are more prevalent in BR007 (80%) compared to SC283 (73%) (Fig. 1e). However, when measured in the different root diameter classes, root surface area turned out to be heterogeneous between the parents, with SC283 showing higher surface area of both fine (SA2) and thicker (SA3) roots (Fig. 1f - 1g) compared to BR007. However, fine roots are still far more prevalent (~17%) than thicker roots (~1%) in the SC283 root system. Finally, the most important trait to reflect P efficiency, grain yield under low-P availability in the soil, was approximately 12% higher for the guinea race parent, SC283, compared to BR007 (Additional file 1 and Fig. 1h).

Trait associations

PAE and PUE were both highly correlated with grain yield (r = 0.85 and 0.97, respectively; Additional file 2), which is consistent with the importance of P acquisition on P use efficiency (Table 1). PUTIL, which we found to be a minor component of PUE compared to acquisition efficiency (Table 1), was less correlated with grain yield (r = 0.4).

Next, we studied the association between root morphology traits and grain yield under low-P availability in the soil via a genetic correlation analysis (Fig. 2). Total root surface area was highly correlated with total root length (correlation coefficient, r = 0.98) and surface area of very fine roots (SA1) (r = 0.99). In addition, surface area of fine roots (SA2) was highly correlated with root volume 2 (r = 1.0). Therefore, among those traits, root surface area was used to gain insights into the role of root morphology on grain yield under low-P availability in the soil. A reduction in root diameter was in general associated with increased total root surface area (r = -0.46), which was driven primarily by very fine roots (RD vs. SA1, r = -0.53) and, to a lesser extent, by thicker roots (RD vs. SA3, r = -0.23). This suggests the existence of some genetic determinants that act to increase root surface area via enhanced development of finer roots. However, the magnitude of the correlation coefficients also indicates that root surface area and root diameter are controlled to some extent independently. Surface area of fine roots was positively but weakly correlated with root diameter (RD vs. SA2, r = 0.1), suggesting that slight increases in root diameter between 1 and 2 mm may result in enhanced surface area. Grain yield under low-P availability in the soil was significantly correlated with the different traits reflecting surface area of fine roots ($r \cong 0.1$, p-value < 0.05), although this association tended to dissipate with thicker roots between 2 and 4.5 mm in diameter (SA3, r = 0.08, p-value = 0.10).

QTL mapping for root morphology and performance traits under low-P

We mapped QTLs underlying P efficiency traits and found that the majority of the QTLs, primarily for PAE (9 out of 10) and PUE (9 out of 10), but also for PUTIL (although to a lesser extent) coincided with those detected for grain yield, with exception of the QTL on chromosome 5 for PUTIL (Additional file 3). This is consistent both with the much higher importance of PAE compared to PUTIL on PUE (Table 1) and with the strong association between grain yield and PAE/PUE (Additional file 2).

Although grain yield was the most informative trait for QTL detection, a PUTIL QTL on chromosome 5 may harbor genes underlying changes in P internal utilization and two chromosome 1 QTLs may jointly underlie PAE and PUTIL. As PUTIL was much less important than P acquisition efficiency for PUE, we thus focused primarily on the genetic mechanisms that enhance P acquisition efficiency via changes in root morphology and their role in increased grain yield on low-P soil.

We initially conducted single-trait QTL mapping with many different traits related to root morphology and sorghum performance under low-P conditions (Additional file 4). This analysis detected a total of 101 QTLs, with the favorable allele of 59 QTLs donated by the guinea parent, SC283, whereas the BR007 alleles increased phenotypic expression for 42 QTLs. Based on the correlation analyses between traits and on the single-trait QTL mapping results, we selected a subset of non-redundant and highly informative traits (i.e. traits repeatedly associated with some QTLs) for multi-trait QTL mapping, focusing primarily on the most important P efficiency trait, namely grain yield under low-P availability in the soil (Fig. 3). P content in the grain (Pg), for example, was highly correlated with grain yield ($r = 0.92$) and we thus we only included grain yield and not Pg for multi-trait QTL mapping. The final set of traits used for multi-trait QTL mapping was comprised of grain yield (Gy), surface area of fine roots in the 1-2 mm diameter class (SA2) and root diameter (RD).

For the selected traits, the majority of the QTLs detected by single-trait QTL mapping (Fig. 3a - 3c) were also detected by multi-trait QTL mapping (Fig. 3d). Exceptions are the QTLs for SA2 on chromosomes 5, 7 and 9 and the Gy QTL on chromosome 8 and 10, which were not detected using multi-trait QTL mapping. Multi-trait mapping detected 14 QTLs (see Fig. 3a - 3c for single-trait mapping results) and revealed ten QTLs related to grain yield (Fig. 3d), within which one QTL was tightly linked to a root morphology QTL (*Gy-3…SA2-3*) and two were possibly pleiotropic with root morphology (*Gy/SA2-3* and *Gy/RD-7*). For all of these QTLs, the favorable allele was donated by SC283 (Additional file 5). In contrast, the favorable alleles for five of the eight grain yield-specific QTLs were donated by BR007.

The different grain yield QTLs explained, in general, approximately 1 to 5% of the genetic variance and increased grain yield by ~120 kg ha^{-1} (Additional file 4 and Additional file 5), except for a Gy QTL at the end region of chromosome 9 (*Gy-9*). This QTL was detected for several different traits (Additional file 4), explained the largest proportion of the genetic variance (~26%, Additional file 4), and was associated with the largest increase in grain yield, of ~400 kg ha^{-1}, with the favorable allele donated by BR007.

Based on single-trait QTL analysis, all grain yield QTLs detected by multi-trait QTL mapping were co-located or were found near QTLs underlying P content and/or dry matter accumulation in hydroponics under low-P (Fig. 3, Additional file 4). The *RD/SA2-2* QTL (Additional file 5), which was the only root morphology QTL not associated with grain yield, co-located with QTLs for root dry matter accumulation, and shoot and root P content assessed in hydroponics via single-trait analyses (Additional file 4).

Multi-trait QTL mapping provided insights into possible pleiotropic QTLs underlying changes in root system morphology and grain yield in the context of genes previously shown to be associated with those traits, such as sorghum homologs of the rice serine/threonine kinase, *OsPSTOL1* (Hufnagel et al. 2014). The physical positions of the *SbPSTOL1* genes and

that of *SbMATE*, which confers sorghum Al tolerance (Magalhaes et al. 2004), in the context of the QTL detected by multi-trait QTL mapping, are shown in Fig. 4. The QTLs *Gy-3* and *SA2-3* were in close physical proximity, between 5.38 and 0.46 Mb, respectively, from the *PSTOL1* gene *Sb03g006765* (Fig. 4a). At the end of chromosome 3, a cluster of four *SbPSTOL1* genes were located ~11 Mb from the *Gy/SA2-3* QTL and this QTL was only 80 Kb from *SbMATE* (Fig. 4b). Finally, the *Gy/RD-7* QTL is located only 0.66 Mb from the *SbPSTOL1* gene, *Sb07g002840* (Fig. 4c).

Expression profile of *SbPSTOL1* genes in the parent's root systems under low-P

Multi-trait QTL mapping results indicated that the favorable alleles at QTLs either tightly linked or possibly pleiotropic with grain yield and root morphology on chromosomes 3 and 7, which were located in the vicinity of *SbPSTOL1* genes, were consistently donated by SC283. Next, we assessed the expression profile of these *SbPSTOL1* genes in roots of the RIL parents, BR007 and SC283, subjected to low-P conditions in hydroponics. *Sb03g006765*, located at the beginning of chromosome 3, and *Sb03g031690*, which is part of a *SbPSTOL1* cluster at position ~60 Mb on chromosome 3 (Hufnagel et al. 2014), were both more highly expressed in the roots of the guinea parent, SC283, in the low-P growth media (Fig. 5). In contrast, expression of *Sb07g002840* was higher in BR007 roots, which donates the inferior allele at the *Gy/RD-7* QTL.

Discussion

Low-P availability in the soil is a major factor that compromises food security in many developing countries in West Africa that rely on the sorghum crop for food production (Weltzien et al. 2006). West Africa is the primary domestication center of the guinea race of sorghum (De Wet 1978), which is used therein as a pivotal staple food in areas with low soil

fertility (Leiser et al. 2012). Thus, sorghum adaptation to soils with low-P availability becomes critical for food security (Leiser et al. 2012, 2014).

Our QTL mapping study emphasized the complex nature of traits related to P efficiency in sorghum, with favorable alleles donated by both parents in rather equal proportions. However, the observed slight overrepresentation of superior QTL alleles derived from the guinea race parent, SC283, may not be coincidental, reflecting local adaptation of guinea sorghums to poor soil fertility and acid soils in West Africa (Weltzien et al. 2006).

QTLs for root morphology coincide with grain yield QTL under low-P availability

The root system of monocotyledonous crop plants consists of one or more seminal roots that originate from the seed embryo after germination, and crown roots that emerge later from nodes along the stem (Uga et al. 2018). Increased root surface area, which can be achieved via enhanced lateral root branching, can enhance P uptake and plant growth (Zhu and Lynch 2004). Among the ten grain yield QTL that were detected by multi-trait mapping, three were either tightly linked (one QTL) or possibly pleiotropic (two QTL) with root morphology traits. Those are: 1) the grain yield QTL, *Gy-3* and the QTL for surface area of fine roots, *SA2-3* at the beginning of chromosome 3, which are only ~6 Mb apart; 2) the pleiotropic *Gy/SA2-3* QTL at position ~71 Mb on chromosome 3; and 3) the *Gy/RD-7* QTL at 3.6 Mb on chromosome 7 (Fig. 3d). Thus, our multi-trait QTL mapping approach established an important role for root system morphology as an entry point for molecular breeding strategies targeting enhanced P uptake and grain yield under low-P availability in the soil.

Specific changes in root morphology are likely important for P efficiency

The grain yield QTLs that are possibly determined by changes in root surface area seem to be more specific to roots between 1-2 mm (SA2) in diameter than to the very fine

roots between 0-1 mm (SA1) or thicker roots (2-4.5 mm, SA3). Via single-trait mapping, a QTL for SA2 (and not for other root diameter classes) was found at the beginning of chromosome 3, tightly linked to a Gy QTL, whereas the grain yield/surface area QTL at the end of chromosome 3 was near SA2 and SA3 QTL. A total surface area (SA) QTL was also found at the end of chromosome 3, but this is expected as total surface area, which is highly correlated with SA1, largely represents the sum of surface area of roots in all diameter classes. Importantly, eight QTLs in total were detected for surface area of fine roots via single-trait mapping, whereas only four and two QTLs were detected for total surface area (that is highly correlated with surface area of very fine roots, r = 0.99) and for surface area of thicker roots, respectively. Very thick roots are not expected to play a major role in nutrient uptake, as plant species with a majority of fine roots in their root systems tend to optimize the ratio between root surface area available for uptake and root weight, reflecting a reduced carbon cost for root biomass formation (Wu et al. 2016).

It is generally thought that the finer the roots, the better the root system can mine the soil for diffusion-limited nutrients, like the phosphate anion on tropical soils. However, although fine roots are the key factor for uptake, particularly for nutrients with very low mobility in tropical soils such as P, our QTL data interestingly suggest there is a trade-off between decreased root diameter and enhanced P uptake. This can be expected as decreased root diameter, beyond a given threshold may, for example, limit root penetration through the soil (Wu et al. 2016) and, possibly, lead to less root longevity (Eissenstat 1992).

SbPSTOL1 genes possibly underlie QTLs for root morphology and grain yield

All three root morphology QTL that were found to be either tightly linked or pleiotropic with grain yield are located in the vicinity of sorghum homologs of the rice serine/threonine kinase, *OsPSTOL1*, which was previously found to enhance early root

growth and grain yield in rice under low-P availability (Gamuyao et al. 2012).

Based on multi-trait mapping, two QTLs underlying grain yield and root surface area, *Gy-3* and *SA2-3*, were found on chromosome 3 at positions 1.6 and 7.4 Mb, respectively. Those QTLs are only ~6 Mb apart and are physically very close to the *SbPSTOL1* gene, *Sb03g006765*, at position 7 Mb. A possible pleiotropic QTL, simultaneously underlying SA2 and Gy (*Gy/SA2-3*), was found approximately ~11 Mb from a *SbPSTOL1* cluster at position ~60 Mb on chromosome 3. The favorable alleles both at the SA2 and Gy QTLs near *Sb03g006765* and at the possible *Gy/SA2-3* pleiotropic QTL at position 71 Mb are derived from the guinea parent, SC283. Although BR007 tended to show greater root surface area compared to SC283, this is due to the prevalence in BR007 of very fine roots, between 0 and 1 mm in diameter (Fig. 1). Compared to BR007, SC283, which donates the positive alleles for the Gy and SA2 QTLs in the vicinity of *SbPSTOL1* genes, has about twice the proportion of fine roots between 1 and 2 mm in diameter, whose surface area gives rise to both SA2 QTL on chromosome 3. Finally, both *Sb03g006765* and *Sb03g031690* (that is part of a *SbPSTOL1* cluster at position ~60 Mb), exhibited significantly higher expression in response to low-P growth conditions specifically in SC283 roots, when compared to BR007, which is in agreement with SC283 donating *SbPSTOL1* alleles that enhance both the surface area of fine roots at the respective chromosome 3 QTL.

Previously, single nucleotide polymorphism (SNP) loci within the *SbPSTOL1* gene *Sb03g006765*, and in the *SbPSTOL1* genes present in the gene cluster at position ~60 Mb, were associated both with variation in root surface area and sorghum performance under low-P (Hufnagel et al. 2014). This suggests that the grain yield QTL on chromosome 3 results, at least in part, from enhanced surface area conferred by *SbPSTOL1* genes. However, the major Al tolerance gene, *SbMATE*, is located at position 71 Mb on the same chromosome, and thus is closer to the pleiotropic *Gy/SA2-3* QTL in the region (Fig. 4). *SbMATE* has been shown to

contribute to grain yield under low-P conditions, possibly via citrate-based enhanced mobilization of P that is bound to the soil clays (Leiser et al. 2014), or simply as an indirect effect of enhanced root development under Al toxicity in the subsoil (Magalhaes et al. 2018). Thus, we cannot rule out that *SbMATE* is responsible for some of the yield advantage that gives rise to the grain yield QTL at the end of sorghum chromosome 3.

We previously reported that allelic variation at the *SbPSTOL1* gene, *Sb07g002840*, influences root diameter, biomass accumulation and P uptake, although associations with grain yield were not found using a sorghum association panel (Hufnagel et al. 2014). A pleiotropic QTL underlying both Gy and RD was found only ~0.6 Mb away from *Sb07g002840*, and the favorable allele for this *Gy/RD-7* QTL was donated by SC283. It is interesting to note that SC283 exhibits overall a larger root diameter compared to BR007, which is in agreement with the SC283 origin of the favorable allele at the *Gy/RD-7* QTL. In barley subjected to low-P conditions, based on the effects of co-localized QTLs, larger root diameter was related to higher grain yield (Gong and McDonald 2017), which is consistent with our results with *Sb07g002840*. A positive relationship has been reported between the size of the apical meristem (reflected by the apical diameter) and important characteristics such as the elongation rate, growth duration and gravitropism (Wu et al. 2016), which could lead to enhanced performance under low-P conditions. It is thus possible that the slight increases in root diameter elicited by *Sb07g002840* lead to an increase in the surface area of laterals that are still fine, generating more physically robust roots without substantial carbon cost. Hence, these roots would be more efficient to optimize P mining in the soil, leading to enhanced P uptake and grain yield under soil low-P availability. Unlike the *SbPSTOL1* genes on chromosome 3, root *Sb07g002840* expression is lower in SC283 compared to BR007. This could suggest an allele-specific repressor effect of *Sb07g002840* on root diameter, with lower expression of the SC283 allele of *Sb07g002840*, leading to a slight increase in root diameter.

Synteny analysis in sorghum and maize supports a role for *PSTOL1* genes in root morphology and reveals other genes possibly involved in P efficiency

We compared the positions of the grain yield QTL detected in sorghum by multi-trait QTL mapping with QTLs related to root morphology and P efficiency reported in the closely related species, maize. The summary shown in Additional file 6 was based primarily on a QTL mapping study that included the same traits used in our sorghum RIL population (Azevedo et al. 2015) and in a comprehensive meta-analysis of QTLs related to low-P tolerance in maize, which defined 23 consensus QTL (*cQTL*, Zhang et al. 2014).

The maize root morphology QTLs, *qMulti3.04*, *qRL8.05* and *qRD4.05*, harbor maize homologs of *Sb03g006765*, *Sb03g031690*, and *Sb07g002840*, respectively (the *SbPSTOL1* genes at chromosomes 3 and 7 that are near grain and root morphology QTLs, Fig 4), in regions that are syntenic between maize and sorghum (Additional file 6). Consistent with our findings that *Sb07g002840* influences root diameter, *qRD4.05* is also associated with changes in root diameter in maize. Four homologs of *OsPSTOL1* were found in maize *cQTL3-1* at bin 3.04 (Zhang et al. 2014), which overlaps with *qMulti3.04* (Azevedo et al. 2015) that is associated with multiple root morphology traits. In this region, the maize *PSTOL1*-like gene, *GRMZM2G412760*, is closely related to *Sb03g006765*. In conjunction with the presence of a functional PSTOL1 protein in rice (*OsPSTOL1*, Gamuyao et al. 2012), this suggests that the PSTOL1 function in modulating root morphology is rather ancient, predating the divergence between maize, sorghum and rice.

This analysis also suggests possible functions for grain yield QTL that are apparently unrelated to root morphology in sorghum. Previously, we detected a major QTL for plant height and flowering time at the end region of sorghum chromosome 9 (Sabadin et al. 2012), which coincides with the QTL underlying multiple traits that was found in the present study. The ubiquitin-conjugating enzyme UBC24, encoded by *phosphate2* (*PHO2*), which is a major

player in P homeostasis and plant responses to P deficiency (Liu et al. 2012), has been recently implicated in tolerance to low-P in maize (Du et al. 2018). We found a highly similar *PHO2* homolog at position ~57 Mb on sorghum chromosome 9, which closely overlaps with multiple QTLs related to root morphology and many other traits, including grain yield (Additional file 4 and Fig. 3). *GRMZM2G381709*, a maize homolog of *PHO2*, is also located in a syntenic position of the maize genome, near a cQTL (*cQTL6-2*) for low-P tolerance in maize defined by meta-analysis (Additional file 6).

Sorghum homologs of the auxin transporters, PIN1 and PIN6, are found near the grain yield QTL at position ~58 Mb on sorghum chromosome 4 (Fig. 3). A related PIN protein in maize is found near another QTL for tolerance to low-P conditions in maize, *cQTL5-5* (Additional file 6). *PIN* genes that encode auxin transporters have been implicated in root architecture changes involving lateral roots under low-P in wheat (Talboys et al. 2014) and may play a significant role in sorghum and maize P efficiency. On chromosome 6, a sorghum homolog of the Al tolerance gene, *ALMT1*, which encodes a root malate efflux transporter and has been recently reported to modulate root growth in response to low-P (Mora-Macías et al. 2017), is found at position ~44 Mb, thus near the grain yield QTL at position 42.5 Mb (Fig. 3 and Additional file 4). In maize, the consensus QTL, *cQTL10-1* (Zhang et al. 2014), is also found near a maize homolog of *ALMT1*, suggesting a role for *ALMT1* on P efficiency in both maize and sorghum. As with SbMATE, it is also possible that root malate release via ALMT1 is involved in solubilizing P that is fixed on the surface of Fe and Al oxide minerals in the soil, making them bioavailable for root uptake.

Conclusions

Phosphorus acquisition efficiency was the major component of P efficiency for sorghum cultivated under low-P availability in the soil and grain yield was highly correlated

with both traits. Although our findings emphasize root system morphology as a major target for molecular approaches aimed at developing P efficient sorghum cultivars, other distinct mechanisms may also play a significant role in sorghum performance on low-P soils via enhanced P acquisition. The molecular determinants of such mechanisms, along with *SbPSTOL1* genes, should power novel, gene-based molecular breeding strategies to enhance food security in tropical regions with low-P availability.

List of abbreviations

Al: Aluminum; *ALMT1*: Aluminum-activated malate transporter in wheat; *ALS3*: Aluminum sensitive 3 (ABC-like transporter); $Alt_{SB:}$ Aluminum tolerance locus in *Sorghum bicolor*; *AtSTOP1*: Sensitive to proton rhizotoxicity in Arabidopsis; BLUE: Best linear unbiased estimation; BLUP: Best linear unbiased predictions; BWA: Burrows-wheeler aligner program; DM: Dry matter; E2: Ubiquitin-conjugating enzyme; FT: Flowering time; GBS: Genotyping by sequencing; Gy: Grain yield; LRT: Likelihood rate test; MAF: Minimum allele frequency; Mb: Mega base pairs; *OsPSTOL1*: Phosphorus starvation tolerance 1 in rice; P: Phosphorus; PAE: Phosphorus use efficiency; PH: Plant height; PHO2: Phosphate 2; Pg: Phosphorus content in grain; Pp: Phosphorus content in the plant (leaves and stem); Pr: Phosphorus content in root tissues; Ps: Phosphorus content in the shoot; Psoil: Phosphorus content available in the soil; PUE: Phosphorus use efficiency; Pt: Total phosphorus content; PUTIL: Phosphorus internal utilization efficiency; QTL: Quantitative trait locus; r: Correlation coefficient; RD: Root diameter; RDM: Root dry matter; REML: Restricted maximum likelihood; RIL: Recombinant inbreed line; RL: Root length; RV: Root volume; SA: Total root surface area; SA1: Surface area of very fine roots between 0-1 mm in diameter; SA2: Surface area of fine roots between 1-2 mm in diameter; SA3: Surface area of thicker roots between 2-4.5 mm in diameter; *SbMATE*: Multidrug and toxic compound extrusion in *Sorghum bicolor*; *SbPSTOL1*: Phosphorus starvation tolerance1 in *Sorghum bicolor*; SDM: Shoot dry matter; SNP: Single nucleotide polymorphism; SRL: Specific root length; UBC24: Ubiquitin-conjugating 24 enzyme; V2: Volume of fine roots between 1-2 mm in diameter.

References

Azevedo GC, Cheavegatti-Gianotto A, Negri BF, et al (2015) Multiple interval QTL mapping and searching for PSTOL1 homologs associated with root morphology, biomass accumulation and phosphorus content in maize seedlings under low-P. BMC Plant Biol 15:172 . doi: 10.1186/s12870-015-0561-y

Balzergue C, Dartevelle T, Godon C, et al (2017) Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. Nat Commun 8:1–16 . doi: 10.1038/ncomms15300

Belal R, Tang R, Li Y, et al (2015) An ABC transporter complex encoded by Aluminum Sensitive 3 and NAP3 is required for phosphate deficiency responses in Arabidopsis. Biochem Biophys Res Commun 463:18–23 . doi: 10.1016/j.bbrc.2015.05.009

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635 . doi: 10.1093/bioinformatics/btm308

Carvalho G, Schaffert RE, Malosetti M, et al (2016) Back to acid soil fields: The citrate transporter SbMATE is a major asset for sustainable grain yield for sorghum cultivated on acid soils. G3 Genes, Genomes, Genet 6:475–484 . doi: 10.1534/g3.115.025791

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381

De Wet JMJ (1978) SPECIAL PAPER: SYSTEMATICS AND EVOLUTION OF SORGHUM SECT. SORGHUM (GRAMINEAE). Am J Bot 65:477–484

De Wet JMJ, Harlan JR (1971) The origin and domestication ofSorghum bicolor. Econ Bot 25:128–135 . doi: 10.1007/BF02860074

Dong J, Piñeros MA, Li X, et al (2017) An Arabidopsis ABC Transporter Mediates Phosphate Deficiency-Induced Remodeling of Root Architecture by Modulating Iron Homeostasis in Roots. Mol Plant 10:244–259 . doi: 10.1016/j.molp.2016.11.001

Doumbia MD, Hossner LR, Onken AB (1993) Variable sorghum growth in acid soils of subhumid west africa. Arid Soil Res Rehabil 7:335–346 . doi: 10.1080/15324989309381366

Doumbia MD, Hossner LR, Onken AB (1998) Sorghum growth in acid soils of West Africa: Variations in soil chemical properties. Arid Soil Res Rehabil 12:179–190 . doi: 10.1080/15324989809381507

Drouillon M, Merckx R (2003) The role of citric acid as a phosphorus mobilization mechanism in highly P-fixing soils. Gayana Bot 60:55–62

Du Q, Wang K, Zou C, et al (2018) The PILNCR1-miR399 regulatory module is important for low-phosphate tolerance in maize. Plant Physiol 1–28 . doi: 10.1104/pp.18.00034

Eissenstat DM (1992) Costs and benefits of constructing roots of small diameter. J Plant Nutr 15:763–782 . doi: 10.1080/01904169209364361

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Gamuyao R, Chin JH, Pariasca-Tanaka J, et al (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature 488:535

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS One 9:e90346 . doi: 10.1371/journal.pone.0090346

Gong X, McDonald G (2017) QTL mapping of root traits in phosphorus-deficient soils reveals important genomic regions for improving NDVI and grain yield in barley. Theor Appl Genet 130:1885–1902

Haley CS, Knott SA (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity (Edinb) 69:315–24 . doi: 10.1038/hdy.1992.131

Hammond JP, Broadley MR, White PJ (2004) Genetic responses to phosphorus deficiency. Ann Bot 94:323–332 . doi: 10.1093/aob/mch156

Harrell Jr FE (2015) With contributions from Charles Dupont and many others. Hmisc: Harrell Miscellaneous. R package version 3.17-4

Ho MD, Rosas JC, Brown KM, Lynch JP (2005) Root architectural tradeoffs for water and phosphorus acquisition. Funct Plant Biol 32:737–748

Hufnagel B, de Sousa SM, Assis L, et al (2014) Duplicate and conquer: multiple homologs of PHOSPHORUS-STARVATION TOLERANCE1 enhance phosphorus acquisition and sorghum performance on low-phosphorus soils. Plant Physiol 166:659–677

Hund A, Trachsel S, Stamp P (2009) Growth of axile and lateral roots of maize: I development of a phenotying platform. Plant Soil 325:335–349 . doi: 10.1007/s11104-009-9984-2

Iuchi S, Koyama H, Iuchi A, et al (2007) Zinc finger protein STOP1 is critical for proton tolerance in Arabidopsis and coregulates a key gene in aluminum tolerance. Proc Natl Acad Sci U S A 104:9900–9905

Johnson HW, Bernard RL (1962) Soybean Genetics And Breeding. Adv Agron 14:149–221 . doi: 10.1016/S0065-2113(08)60438-1

Kochian L V (1995) Cellular mechanisms of aluminum toxicity and resistance in plants. Annu Rev Plant Biol 46:237–260 . doi: 10.1146/annurev.pp.46.060195.001321

Larsen PB, Geisler MJB, Jones CA, et al (2005) ALS3 encodes a phloem-localized ABC transporter-like protein that is required for aluminum tolerance in Arabidopsis. Plant J 41:353–363 . doi: 10.1111/j.1365-313X.2004.02306.x

Larsen PB, Tai C, Kochian L V, Howell SH (1996) Arabidopsis Mutants with lncreased Sensitivity to Aluminum ’. Plant Physiol 110:743–751

Leiser WL, Rattunde HFW, Piepho H-P, et al (2012) Selection strategy for sorghum targeting

phosphorus-limited environments in West Africa: analysis of multi-environment experiments. Crop Sci 52:2517–2527

Leiser WL, Rattunde HFW, Weltzien E, et al (2014) Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West African sorghum. BMC Plant Biol 14:206 . doi: 10.1186/s12870-014-0206-6

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows--Wheeler transform. bioinformatics 25:1754–1760

Liu T-Y, Huang T-K, Tseng C-Y, et al (2012) PHO2-Dependent Degradation of PHO1 Modulates Phosphate Homeostasis in Arabidopsis. Plant Cell 24:2168–2183 . doi: 10.1105/tpc.112.096636

López-Arredondo DL, Leyva-González MA, González-Morales SI, et al (2014) Phosphate Nutrition: Improving Low-Phosphate Tolerance in Crops. Annu Rev Plant Biol 65:95–123 . doi: 10.1146/annurev-arplant-050213-035949

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041–1049 . doi: 10.1104/pp.111.175414

Magalhaes J, Piñeros MA, Maciel L, Kochian L (2018) Emerging pleiotropic mechanisms underlying aluminum resistance and phosphorus acquisition on acidic soils. Front Plant Sci 9:1420

Magalhaes J V, de Sousa SM, Guimaraes CT, Kochian L V (2017) Chapter 7 - The role of root morphology and architecture in phosphorus acquisition: physiological, genetic, and molecular basis. In: Hossain MA, Kamiya T, Burritt DJ, et al. (eds) Plant Macronutrient Use Efficiency. Academic Press, pp 123–147

Magalhaes J V, Garvin DF, Wang Y, et al (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics 167:1905–1914

Magalhaes J V, Liu J, Guimarães CT, et al (2007) A gene in the multidrug and toxic

compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nat Genet 39:1156–1161 . doi: 10.1038/ng2074

Magnavaca R, Gardner CO, Clark RB (1987) Inheritance of aluminum tolerance in maize. In: Genetic aspects of plant mineral nutrition. Springer, pp 201–212

Marschner H (1995) Mineral Nutrition of Higher Plants, 2nd Editio

Mendes FF, Guimarães LJM, Souza JC, et al (2014) Genetic architecture of phosphorus use efficiency in tropical maize cultivated in a low-P soil. Crop Sci 54:1530–1538 . doi: 10.2135/cropsci2013.11.0755

Moll RH, Kamprath EJ, Jackson WA (1982) Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. Agron Journaç 74:

Mora-Macías J, Ojeda-Rivera JO, Gutiérrez-Alanís D, et al (2017) Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate. Proc Natl Acad Sci 114:E3563--E3572

Morais De Sousa S, Clark RT, Mendes FF, et al (2012) A role for root morphology and related candidate genes in P acquisition efficiency in maize. Funct Plant Biol 39:925–935 . doi: 10.1071/FP12022

Müller J, Toev T, Heisters M, et al (2015) Iron-Dependent Callose Deposition Adjusts Root Meristem Maintenance to Phosphate Availability. Dev Cell 33:216–230 . doi: 10.1016/j.devcel.2015.02.007

Neyman J, Pearson ES (1928) On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference : Part I. Biometrika 20A:175–240 . doi: 10.2307/2331945

Parentoni SN, De Souza Júnior CL (2008) Phosphorus acquisition and internal utilization efficiency in tropical maize genotypes. Pesqui Agropecu Bras 43:893–901 . doi: 19.1590/S0100-204X2008000700014

Pearson K (1895) Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. Philos Trans R Soc London Ser A, Contain Pap a Math or Phys Character 187:253–318 . doi: 10.1098/rsta.1896.0007

R Core Team (2016) R: A Language and Environment for Statistical Computing

Roberts A, McMillan L, Wang W, et al (2007) Inferring missing genotypes in large SNP panels using fast nearest-neighbor searches over sliding windows. Bioinformatics 23:401–407 . doi: 10.1093/bioinformatics/btm220

Sabadin PK, Malosetti M, Boer MP, et al (2012) Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. Theor Appl Genet 124:1389–1402 . doi: 10.1007/s00122-012-1795-9

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RWL (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81:8014–8018 . doi: 10.1073/pnas.81.24.8014

Sasaki T, Yamamoto Y, Ezaki B, et al (2004) A wheat gene encoding an aluminum-activated malate transporter. Plant J 37:645–653 . doi: 10.1111/j.1365-313X.2003.01991.x

Schaffert RE, Alves VMC, Pitta GVE, et al (2001) Genetic variability in sorghum for P efficiency and responsiveness. In: Horst WJ, Schenk MK, Bürkert A, et al. (eds) Plant Nutrition. Developments in Plant and Soil Sciences. Springer, Dordrecht, pp 72–73

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3:1101–1108 . doi: 10.1038/nprot.2008.73

Shaw JN (2001) Iron and aluminum oxide characterization for highly-weathered Alabama ultisols. Commun Soil Sci Plant Anal 32:49–64 . doi: 10.1081/CSS-100102992

Silva LDCE, Wang S, Zeng Z-B (2012) Multiple trait multiple interval mapping of quantitative trait loci from inbred line crosses. BMC Genet 13:67 . doi: 10.1186/1471-

2156-13-67

Talboys PJ, Healey JR, Withers PJA, Jones DL (2014) Phosphate depletion modulates auxin transport in Triticum aestivum leading to altered root branching. J Exp Bot 65:5023–5032 . doi: https://doi.org/10.1093/jxb/eru284

Uga Y, Assaranurak I, Kitomi Y, et al (2018) Genomic regions responsible for seminal and crown root lengths identified by 2D & 3D root system image analysis. BMC Genomics 19:273

Vance CP, Uhde-Stone C, Allan DL (2003) Phosphorus acquisition and use: Critical adaptations by plants for securing a nonrenewable resource. New Phytol 157:423–447 . doi: 10.1046/j.1469-8137.2003.00695.x

VSN International (2014) GenStat for Windows 17th Edition

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

Weltzien E, Rattunde H, Clerget B, et al (2006) Sorghum diversity and adaptation to drought in West Africa. In: Jarvis D, Mar I, Sears L (eds) Enhancing the use of crop genetic diversity to manage abiotic stress in agricultural production systems. pp 31–38

Wu Q, Pagès L, Wu J (2016) Relationships between root diameter, root length and root branching along lateral roots in adult, field-grown maize. Ann Bot 117:379–390

Zellner A (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. J Am Stat Assoc 57:348–368

Zhang H, Uddin MS, Zou C, et al (2014) Meta-analysis and candidate gene mining of low-phosphorus tolerance in maize. J Integr Plant Biol 56:262–270 . doi: https://doi.org/10.1111/jipb.12168

Zhu J, Lynch JP (2004) The contribution of lateral rooting to phosphorus acquisition efficiency in maize (Zea mays) seedlings. Funct Plant Biol 31:949–958

Zou F, Fine JP, Hu J, Lin DY (2004) An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. Genetics 168:2307–2316 . doi: 10.1534/genetics.104.031427

Figures

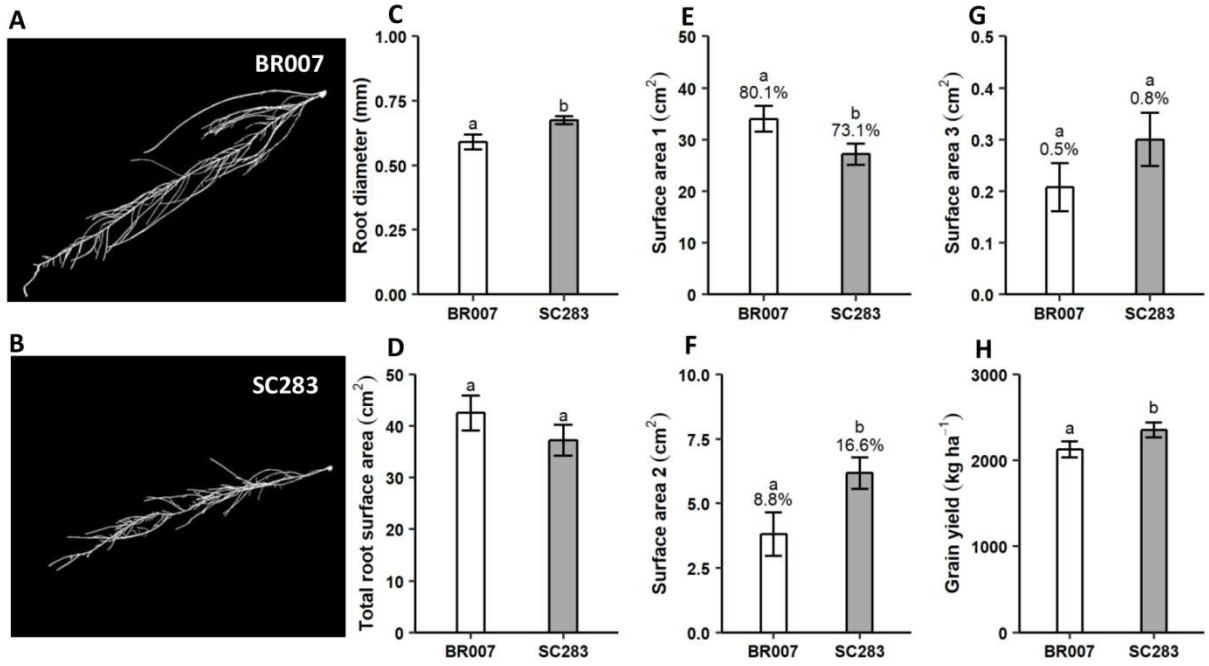


Fig. 1 Phenotypic characterization of the RIL parents. Images of the **(a)** BR007 and **(b)** SC283 root systems. Phenotypic means for **(c)** root diameter (RD), **(d)** total root surface area (SA), surface area of roots in the diameter classes of **(e)** 0-1 mm, (SA1, designated as very fine roots), **(f)** 1-2 mm (SA2, fine roots) and **(g)** 2-4.5 mm (SA3, thicker roots). All surface area measures are in cm^2. Root images and root morphology traits were assessed after 13 days in nutrient solution with low-P. **(h)** Grain yield (Gy) was assessed in a low-P soil with one hundred twenty reps. The proportion of roots within each diameter class relative to total surface area of the root systems are shown as percentages in **e**, **f** and **g.** Error bars are shown. Different letters indicated statistical differences by the t-test (p-values ≤ 0.10).

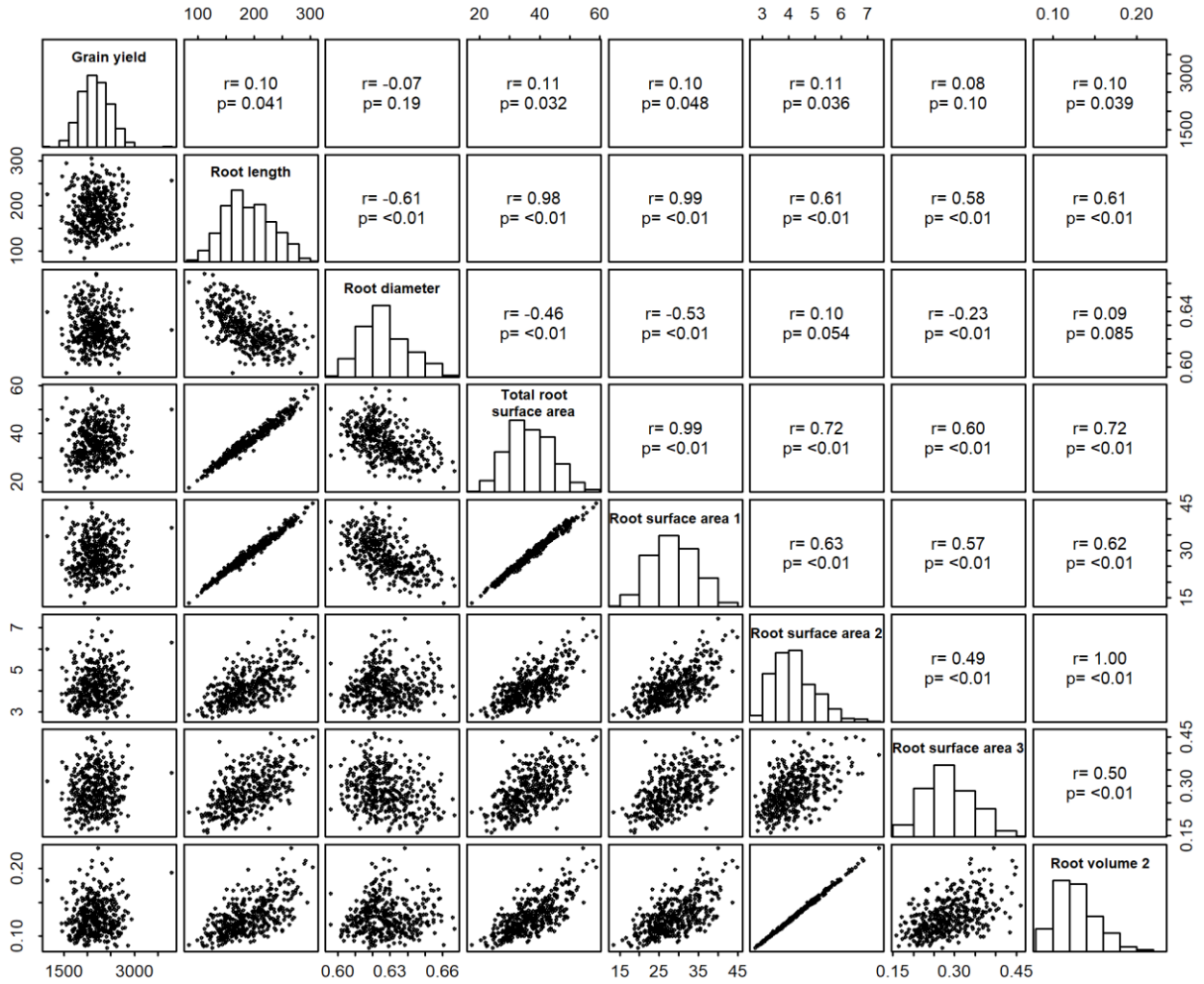


Fig. 2 Genetic correlation analysis of grain yield and root morphology traits. Grain yield data (kg ha^{-1}) was acquired for plants grown on a low-P soil. The root morphology traits assessed after 13 days in nutrient solution with low-P are total root length (cm), root diameter (RD), total root surface area (SA), surface area of roots in the diameter classes of 0-1 mm (SA1, designated as very fine roots), 1-2 mm (SA2, fine roots) and 2-4.5 mm (SA3, thicker roots), and volume of fine roots between 1-2 mm in diameter (root volume 2, in cm^3). Surface area measurements are expressed in cm^2. Pearson correlation coefficients (r) and p-values (p) are shown.

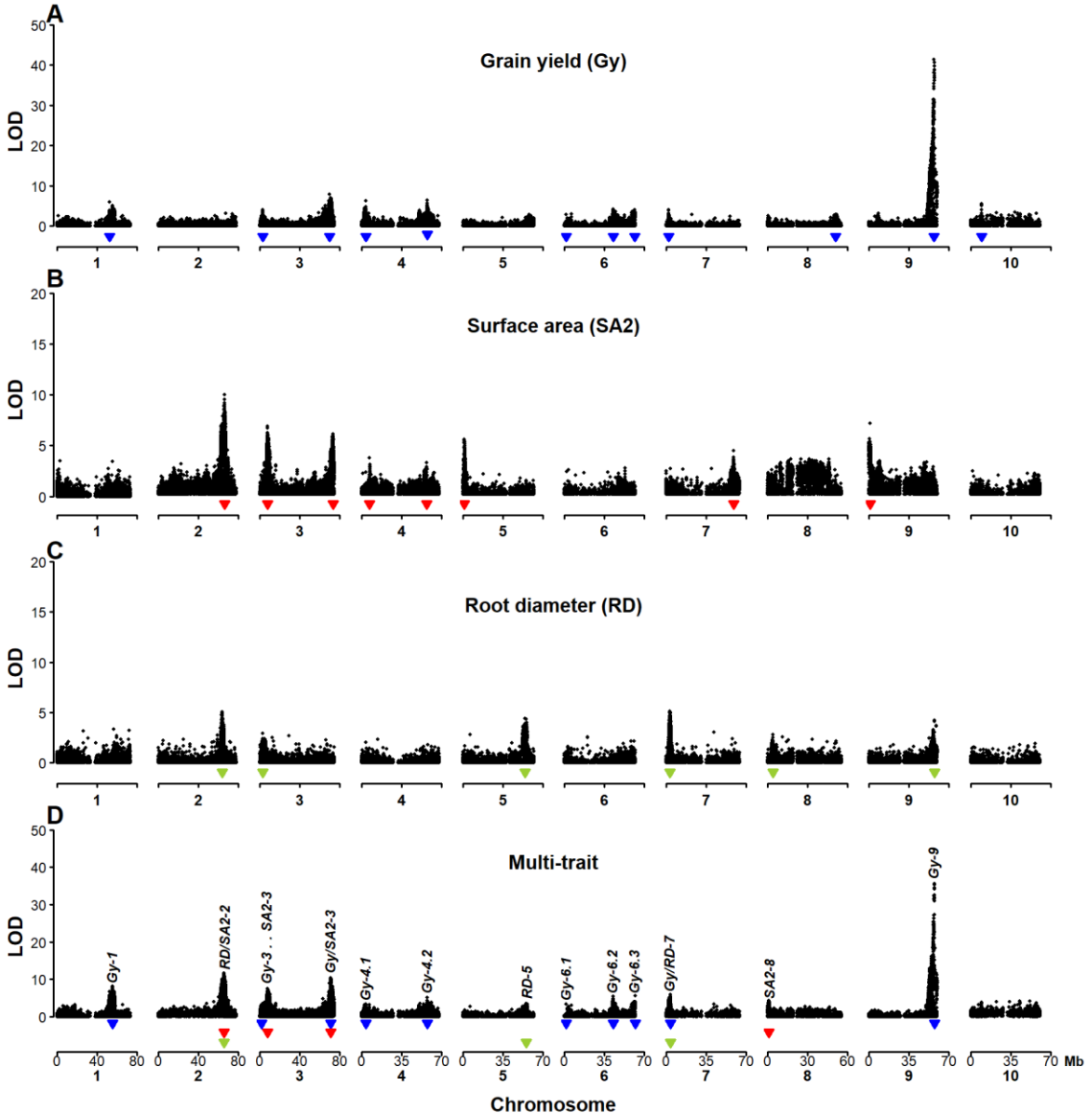


Fig. 3 Single- and multi-trait QTL mapping profiles for grain yield and root morphology traits. The final set of traits used for multi-trait QTL mapping was comprised of grain yield (Gy), surface area of fine roots in the 1-2 mm diameter class (SA2) and root diameter (RD). Grain yield (Gy) data (kg ha^{-1}) was acquired in a low-P soil. The root morphology traits assessed after 13 days in nutrient solution with low-P are root diameter (RD, mm), surface area of fine roots between 1-2 mm in diameter (SA2, in cm^2). QTL profiles obtained with **(a-c)** single- and **(d)** multi-trait QTL mapping are shown. QTLs were designated based on the respective traits followed by the chromosome locations, and are numbered in the case of multiple QTLs within the same chromosome. For example *Gy-6.1* is a grain yield QTL located in the beginning of chromosome 6. Tight linkage between QTL or possible pleiotropy were depicted by double dots and slashes, respectively in the QTL designations. Blue, red and green inverted triangles depict the positions of QTLs for Gy, SA2 and RD, respectively.

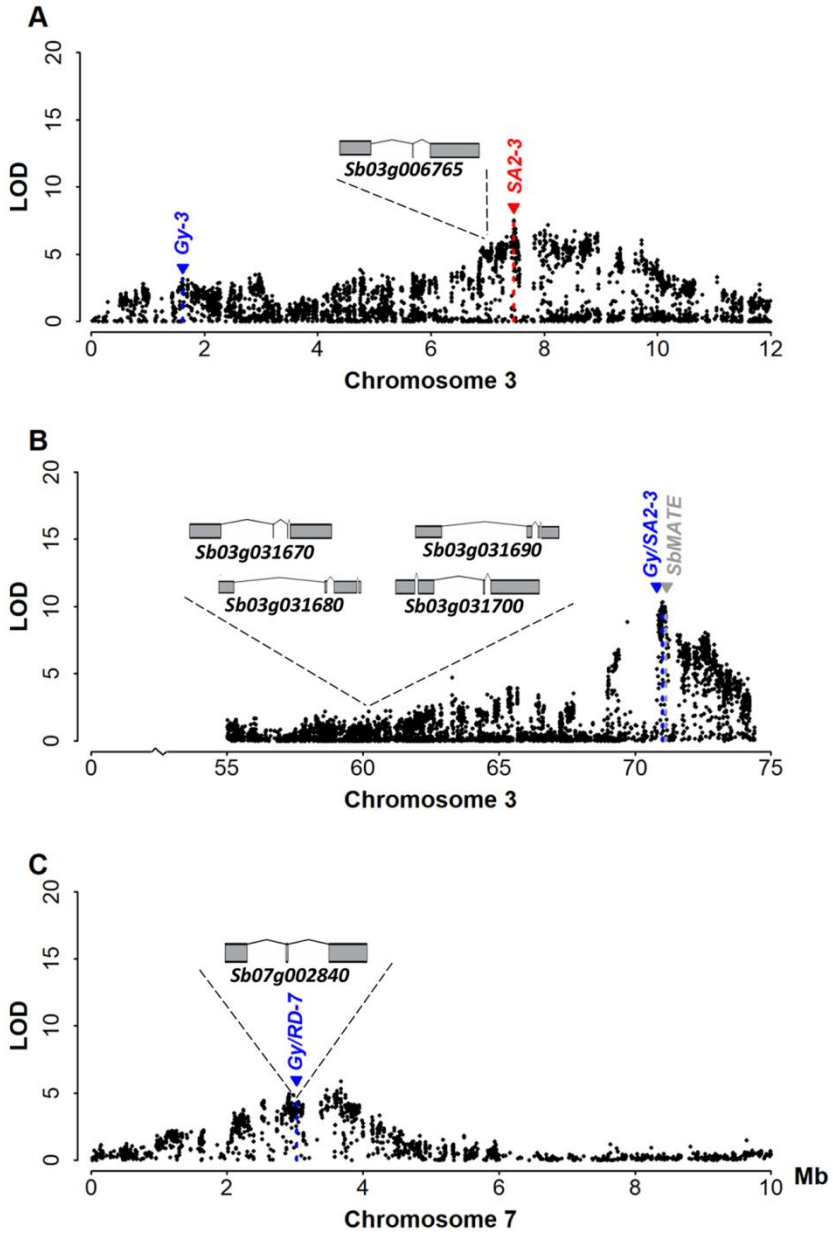


Fig. 4 Physical positions of *SbPSTOL1* genes in the context of the QTL regions detected by multi-trait QTL mapping (**Fig. 3d**). The root morphology traits assessed after 13 days in nutrient solution with low-P are root diameter (RD, mm) and surface area of fine roots between 1 and 2 mm in diameter (SA2, in cm^2). Possible pleiotropy between Gy (grain yield), SA2 and RD QTL, when detected, were depicted by slashes in the QTL designations. Physical positions and gene models for the *SbPSTOL1* genes (https://phytozome.jgi.doe.gov/, v1.4 of the sorghum genome), **(a)** *Sb03g006765*, **(b)** the *SbPSTOL1* cluster including *Sb03g031690* and **(c)** *Sb07g002840* are shown.

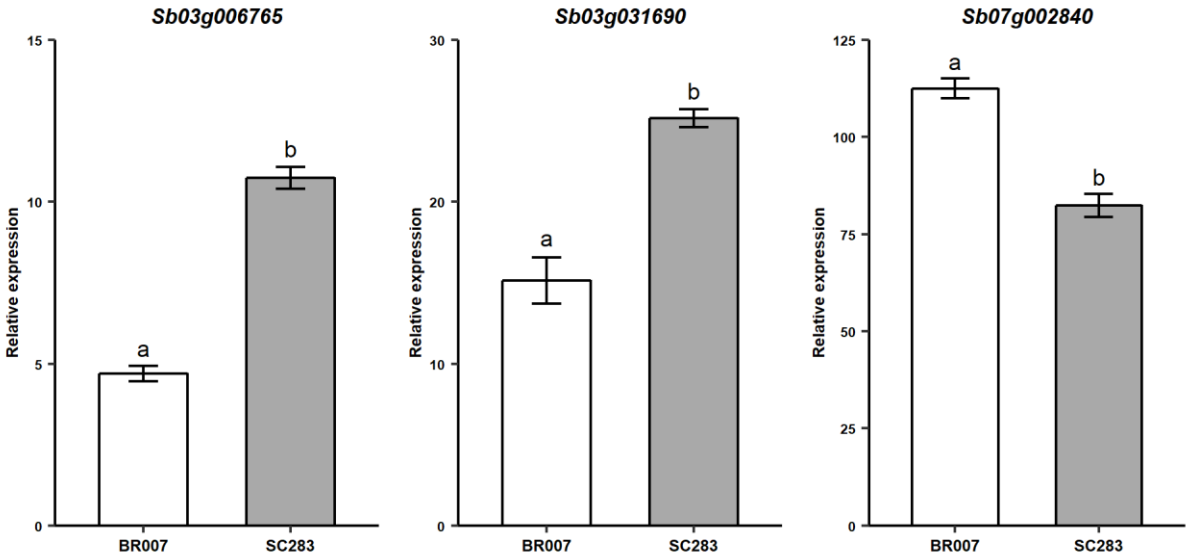


Fig. 5 Expression profile of *SbPSTOL1* genes found near QTLs for grain yield and root morphology (**Fig. 3d**) via quantitative RT-PCR. Whole root systems of the parents cultivated in nutrient solution with low-P for 13 days were collected, frozen, and used for quantitative RT-PCR. Expression was assessed using the $2^{-\Delta\Delta CT}$ method (Schmittgen and Livak 2008). Different letters indicated statistical differences by the t-test (p-values ≤ 0.05). Error bars are shown.

Tables

Table 1 Relative importance of PAE and PUTIL over PUE assessed in low-P conditions field

Trait	Correlation $(r_{xiy})^a$	Standard Deviation (S)	Sx_i/S_y	Relative importance
PAE (x1)	0.9216	0.2285	0.8868	0.82
PUTIL (x2)	0.4763	0.0999	0.3878	0.18
PUE (y)		0.255		

[a]*rxiy*: Phenotypic correlation among Phosphorus acquisition efficiency (PAE) and Phosphorus internal utilization efficiency (PUTIL) and Phosphorus use efficiency (PUE).

Additional files

Additional file 1 Descriptive statistics and variance components for traits assessed in low P conditions (field and hydroponics)

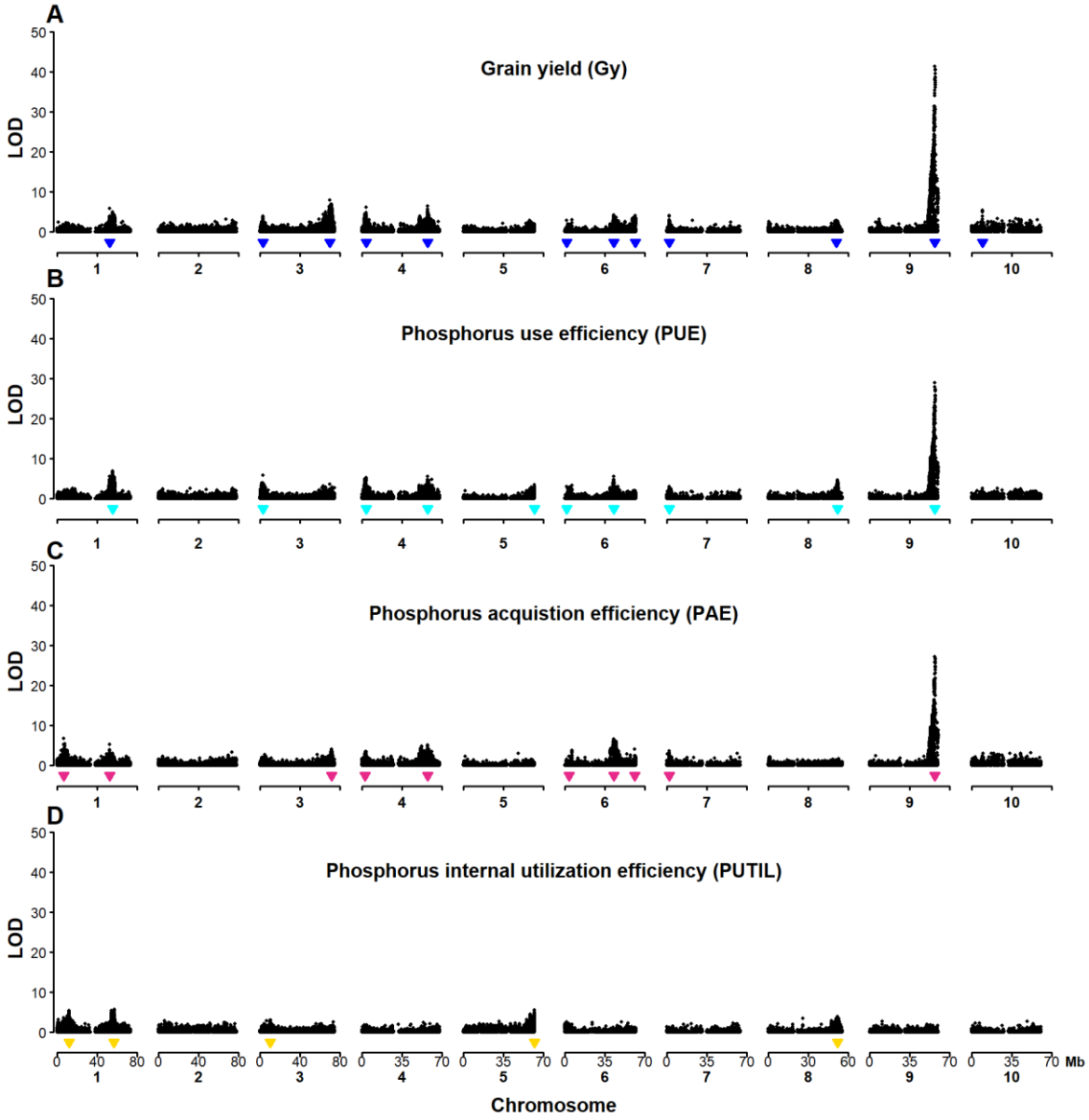
Variance components	Field traits					Traits assessed in hydroponics											
						Root morphology								Dry matter / P content			
	Gy (kg ha^{-1})	FT (days)	PH (cm)	Pp (kg ha^{-1})	Pg (kg ha^{-1})	RL (cm)	RD (mm)	SA (cm²)	SA1 (cm²)	SA2 (cm²)	SA3 (cm²)	RV (cm³)	V2 (cm3)	SDM (g)	RDM (g)	Ps (g)	Pr (g)
Genetic variance	187809	3.62	222.20	0.09	0.59	2350.00	5.36×10^{-4}	71.47	43.87	1.16	0.01	0.01	1.15×10^{-3}	1.33×10^{-5}	7.73×10^{-6}	1.42×10^{-4}	1.52×10^{-4}
Residual variance	467203	3.98	149.40	0.44	1.90	1963.00	3.08×10^{-3}	67.33	42.41	2.60	0.02	0.02	2.47×10^{-3}	1.11×10^{-5}	2.64×10^{-5}	3.40×10^{-4}	1.60×10^{-4}
RIL population																	
Mean	2081	70.41	141.10	1.26	4.24	190.00	0.63	36.85	28.71	4.23	0.29	0.58	0.13	0.02	0.02	0.06	0.05
Minimum	132	66.63	75.20	0.24	0.53	54.50	0.46	11.47	8.46	1.61	0.02	0.19	0.05	0.01	0.01	0.02	0.02
Maximum	4562	77.64	222.00	3.37	7.96	337.30	0.74	65.51	50.16	9.84	0.63	1.03	0.31	0.03	0.03	0.12	0.10
h²	0.54	0.74	0.83	0.38	0.48	0.77	0.34	0.75	0.75	0.56	0.53	0.65	0.57	0.78	0.46	0.55	0.73
Parents																	
BR007	1965	71.35	147.00	1.14	3.78	233.50	0.59	42.52	33.99	3.81	0.21	0.63	0.11	0.02	0.02	0.07	0.04
SC283	2196	69.47	135.20	1.38	4.70	176.60	0.68	37.18	27.15	6.17	0.30	0.63	0.19	0.02	0.02	0.07	0.05

Gy: grain yield; FT: flowering time; PH: plant height; Pp: phosphorus content in the plant (leaves and stem); Pg: phosphorus content in the grain; RL: root length; RD: root diameter; SA: total root surface area; SA1: surface area of very fine roots between 0-1 mm in diameter, SA2: surface area of fine roots between 1-2 mm in diameter; SA3: surface area of thicker roots between 2-4.5 mm in diameter; RV: root volume; V2 volume of fine roots between 1-2 mm in diameter; SDM: shoot dry matter; RDM: root dry matter; Ps: phosphorus content in the shoot; Pr: phosphorus content in the root. All genetic variance components were significant at p-values ≤ 0.05 by the Likelihood Ratio Test. h²: heritability. Best linear unbiased predictors (BLUEs) are shown for the parents.

Additional file 2 Correlations and p-values among all traits assessed in low P conditions (field and hydroponics)

p-values			Field traits								Traits assessed in hydroponics												
											Root morphology									Dry matter / P content			
Correlations			Gy	HI	FT	PH	Pp	Pg	PAE	PUTIL	PUE	RL	SA	RD	RV	V2	SA1	SA2	SA3	SDM	RDM	RSR	Ps	Pr
Field traits		Gy		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.03	0.19	0.03	0.04	0.05	0.04	0.10	0.02	0.01	0.98	0.01	0.03
		HI	0.42		0.69	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	0.07	0.55	0.07	0.04
		FT	-0.14	-0.02		0.00	0.06	0.00	0.00	0.52	0.01	0.04	0.07	0.02	0.17	0.13	0.02	0.18	0.13	0.42	0.10	0.18	0.04	0.03
		PH	0.43	0.25	-0.28		0.24	0.00	0.00	0.00	0.00	0.15	0.31	0.08	0.63	0.28	0.19	0.35	0.51	0.01	0.00	0.99	0.04	0.00
		Pp	0.23	-0.52	-0.10	0.06		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.23	0.01	0.55
		Pg	0.92	0.38	-0.14	0.40	0.21		0.00	0.00	0.00	0.07	0.06	0.28	0.06	0.06	0.08	0.05	0.07	0.00	0.00	0.48	0.00	0.00
		PAE	0.85	0.11	-0.15	0.34	0.56	0.90		0.20	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.00	0.80	0.00	0.01
		PUTIL	0.40	0.63	-0.03	0.22	-0.52	0.17	-0.06		0.00	0.02	0.02	0.04	0.05	0.14	0.02	0.15	0.01	0.13	0.07	0.94	0.14	0.15
		PUE	0.97	0.40	-0.14	0.37	0.25	0.89	0.87	0.38		0.06	0.05	0.14	0.05	0.07	0.07	0.06	0.16	0.05	0.03	0.92	0.01	0.27
Traits assessed in hydroponics	Root morphology	RL	0.10	-0.22	-0.10	0.07	0.23	0.09	0.15	-0.12	0.10		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34	0.00	0.01
		SA	0.11	-0.21	-0.09	0.05	0.22	0.10	0.15	-0.11	0.10	0.98		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00
		RD	-0.07	0.17	0.12	-0.09	-0.20	-0.05	-0.13	0.10	-0.07	-0.61	-0.46		0.00	0.09	0.00	0.05	0.00	0.20	0.00	0.02	0.00	0.00
		RV	0.11	-0.18	-0.07	0.02	0.20	0.10	0.14	-0.10	0.10	0.92	0.98	-0.28		0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00
		V2	0.10	-0.15	0.08	-0.05	0.17	0.10	0.13	-0.07	0.09	0.61	0.72	0.09	0.82		0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00
		SA1	0.10	-0.21	-0.11	0.07	0.21	0.09	0.14	-0.11	0.09	0.99	0.99	-0.53	0.95	0.62		0.00	0.00	0.00	0.00	0.28	0.00	0.00
		SA2	0.11	-0.15	0.07	-0.05	0.16	0.10	0.13	-0.07	0.10	0.61	0.72	0.10	0.82	1.00	0.63		0.00	0.00	0.00	0.01	0.00	0.00
		SA3	0.08	-0.18	-0.08	0.03	0.22	0.09	0.14	-0.13	0.07	0.58	0.60	-0.23	0.60	0.50	0.57	0.49		0.00	0.00	0.30	0.00	0.00
	Dry matter / P content	SDM	0.12	-0.04	-0.04	0.13	0.11	0.15	0.14	-0.08	0.10	0.62	0.68	-0.06	0.72	0.64	0.64	0.65	0.55		0.00	0.00	0.00	0.00
		RDM	0.14	-0.09	-0.08	0.15	0.17	0.15	0.17	-0.09	0.11	0.72	0.77	-0.20	0.78	0.66	0.74	0.67	0.62	0.78		0.01	0.00	0.00
		RSR	0.00	-0.03	-0.07	0.00	0.06	-0.04	0.01	0.00	0.00	-0.05	-0.08	-0.12	-0.10	-0.14	-0.05	-0.14	-0.05	-0.49	0.125		0.00	0.17
		Ps	0.14	-0.09	-0.10	0.10	0.14	0.15	0.16	-0.07	0.12	0.61	0.63	-0.18	0.64	0.56	0.60	0.56	0.47	0.74	0.64	-0.29		0.00
		Pr	0.11	0.10	-0.11	0.17	0.03	0.17	0.13	-0.07	0.06	0.12	0.18	0.16	0.24	0.29	0.15	0.30	0.21	0.43	0.52	0.07	0.40	

Gy: grain yield; HI: harvest index; FT: flowering time; PH: plant height; Pp: phosphorus content in the plant (leaves and stem); Pg: phosphorus content in the grain; PAE: phosphorus acquisition efficiency; PUTIL: phosphorus internal utilization efficiency; PUE: phosphorus use efficiency; RL: root length; SA: total root surface area; RD: root diameter; RV: root volume; V2 volume of fine roots between 1-2 mm in diameter; SA1: surface area of very fine roots between 0-1 mm in diameter. SA2: surface area of fine roots between 1-2 mm in diameter; SA3: surface area of thicker roots between 2-4.5 mm in diameter; SDM: shoot dry matter; RDM: root dry matter; RSR: root shoot ratio; Ps: phosphorus content in the shoot; Pr: phosphorus content in the root.



Additional file 3 Single-trait QTL mapping profiles for grain yield (Gy), phosphorus use efficiency (PUE), phosphorus acquisition efficiency (PAE) and phosphorus internal utilization efficiency (PUTIL). Blue, light blue, pink and yellow inverted triangles depict the positions of QTLs for Gy, PUE, PAE and PUTIL respectively.

Additional file 4 Detailed information and confidence interval of the QTLs detected by single-trait QTL mapping

Trait	Chr	Position (bp)	Inferior limit (SNP - bp)[a]	Upper limit (SNP - bp)[b]	Criterion[c]	Effect[d]	h^2 QTL (%)[e]	LOD[f]
Gy	1	52867138	51470058	53030036	LOD-5	137.30	3.09	5.93
	3	2633838	2416138	3632469	LOD-3	108.68	2.00	4.00
	3	69721261	65415968	72349800	LOD-5	157.29	4.17	7.96
	4	4048460	2664697	4528042	LOD-5	-137.30	3.20	6.26
	4	57829755	57268360	61161923	LOD-5	144.53	3.45	6.45
	6	1515389	1511565	1515571	Adjacent SNP	-97.59	1.63	2.83
	6	42531422	42299754	44786145	LOD-1.5	-120.80	2.50	4.25
	6	61158715	59524891	61968905	LOD-1.5	-111.34	2.11	4.11
	7	2546541	1651645	3469312	LOD-3	110.28	2.05	4.04
	8	51004690	49657000	52239880	LOD-1.5	-93.27	1.48	2.93
	9	56918847	56803551	56918855	Adjacent SNP	-401.08	26.40	41.37
	10	9665482	7519015	19977971	LOD-5	127.62	2.78	5.46
FT	1	57228393	55854323	55854323	LOD-5	-0.43	3.81	5.33
	3	59659567	58614157	63559502	LOD-5	0.64	8.50	11.29
	4	64650431	63762999	65706745	LOD-5	0.65	8.43	11.34
	6	45286099	44786824	45909796	LOD-5	-0.67	9.16	11.98
	8	48894693	47593920	53098121	LOD-5	-0.51	5.32	7.34
	9	57240359	56214458	59010231	LOD-1,5	0.67	8.71	11.72
PH	1	55802290	55802244	55802294	Adjacent SNP	7.02	4.81	15.24
	3	70956129	70946538	70971448	Adjacent SNP	7.20	5.12	16.62
	4	2659321	2508776	4106307	LOD-3	-4.16	1.70	5.82
	6	1515389	1511565	1515571	Adjacent SNP	-2.87	0.83	2.52
	6	42659282	42618767	43223581	LOD-5	-7.04	4.95	14.07
	8	54133247	52495190	55155213	LOD-5	3.94	1.55	5.34
	9	2163799	1079606	3391743	LOD-5	4.07	1.64	5.66
	9	57236934	57236931	57239857	Adjacent SNP	-22.09	47.34	92.69
	10	51375639	51240178	51375654	Adjacent SNP	4.83	2.33	7.85
Pp	3	10562562	9336091	13099549	LOD-5	0.15	8.89	8.60
	8	1253340	563258	2452080	LOD-5	0.09	3.28	3.37
	9	56373414	53687659	59576115	LOD-3	-0.11	4.22	4.19

Trait	Chr	Position (bp)	Inferior limit (SNP - bp)[a]	Upper limit (SNP - bp)[b]	Criterion[c]	Effect[d]	h^2 QTL (%)[e]	LOD[f]
Pp	1	10213690	6639971	11126848	LOD-5	0.33	4.87	9.02
	1	52867138	50014371	56899882	LOD-5	0.26	2.84	5.44
	3	941295	505834	2793197	LOD-3	0.21	1.87	3.56
	3	70940992	65424826	72581671	LOD-5	0.27	3.26	6.16
	4	2775590	2365712	6426489	LOD-5	-0.25	2.62	5.10
	4	57829755	56105112	59077791	LOD-5	0.28	3.35	6.12
	6	1515389	1391651	2762632	LOD-1.5	-0.17	1.21	2.04
	6	42514160	39909230	42750973	LOD-3	-0.29	3.82	6.26
	6	61156724	58784270	61582802	LOD-5	-0.26	2.98	5.63
	7	2111675	2009233	3047986	LOD-3	0.23	2.29	4.44
	9	57236931	56918855	57236934	Adjacent SNP	-0.87	32.27	47.73
RL	3	72723595	72589876	73753094	LOD-3	11.11	4.04	4.11
	5	55148812	48628059	57871373	LOD-5	-12.81	5.45	5.49
	9	56918847	56803551	56918855	Adjacent SNP	-19.14	11.64	11.41
RD	2	63987239	62344000	67454922	LOD-5	9.04×10^{-03}	4.99	5.07
	3	2860525	2852321	2908010	LOD-1.5	6.75×10^{-03}	2.81	2.96
	5	54117286	53419779	56726018	LOD-3	8.45×10^{-03}	4.31	4.43
	7	3610920	2132016	5137309	LOD-5	9.07×10^{-03}	5.05	5.13
	8	4004272	3548990	6744458	LOD-1	-6.59×10^{-03}	2.68	2.79
	9	57158460	55184769	59419466	LOD-3	8.24×10^{-03}	4.08	4.24
SA	2	66100293	65500047	67462841	LOD-3	1.99	4.19	4.51
	3	72723595	72581704	73409388	LOD-3	2.10	4.60	4.90
	5	55168534	48689027	56072846	LOD-5	-2.12	4.76	5.06
	9	56918847	56803551	56918855	Adjacent SNP	-3.35	11.40	11.74
SA2	2	66111836	63085373	67357397	LOD-5	0.41	7.96	9.91
	3	7458033	5818022	10371376	LOD-5	0.33	5.40	6.78
	3	73358414	70806957	74172322	LOD-5	0.31	4.69	6.02
	4	7258258	2545983	8039551	LOD-3	0.24	2.81	3.66
	4	57245536	54450257	58810251	LOD-3	0.22	2.38	3.15
	5	1122951	463766	2005717	LOD-3	-0.30	4.28	5.49
	7	59183269	56895432	61024033	LOD-3	0.26	3.34	4.35
	9	1100753	266961	1655752	LOD-5	-0.34	5.61	7.07

Trait	Chr	Position (bp)	Inferior limit (SNP - bp)[a]	Upper limit (SNP - bp)[b]	Criterion[c]	Effect[d]	h^2 QTL (%)[e]	LOD[f]
SA3	2	65377085	62297769	67860200	LOD-5	0.03	5.94	5.33
	3	71000920	69297704	74017132	LOD-3	0.03	4.51	4.07
RV	2	66100293	64485159	67673105	LOD-5	0.04	6.07	6.47
	3	72723595	68939930	74204072	LOD-5	0.03	4.87	5.19
	5	55168534	46393055	56332735	LOD-3	-0.03	3.79	4.07
	9	56918847	56874994	56918855	Adjacent SNP	-0.05	10.37	10.77
V2	2	66111836	64442866	67224055	LOD-5	1.28×10^{-02}	8.16	10.17
	3	7458033	6829237	10165361	LOD-5	1.02×10^{-02}	5.20	6.56
	3	73358414	70806957	74172322	LOD-5	9.71×10^{-03}	4.67	6.02
	4	7258258	4515244	10231254	LOD-3	7.27×10^{-03}	2.60	3.41
	4	57245536	53534478	58992822	LOD-3	7.22×10^{-03}	2.59	3.42
	5	1122991	1065045	1449415	LOD-1.5	-9.46×10^{-03}	4.34	5.57
	7	59183269	56896540	61324427	LOD-3	8.25×10^{-03}	3.37	4.40
	9	1100753	266961	1960233	LOD-5	-1.08×10^{-02}	5.64	7.12
SDM	1	15827061	7446178	18272551	LOD-5	9.22×10^{-04}	4.97	5.79
	3	8064255	6539868	11803909	LOD-5	9.85×10^{-04}	5.75	6.33
	3	72967699	70340800	74088666	LOD-3	7.07×10^{-04}	2.93	3.35
	4	62047113	59822466	62352489	LOD-3	-7.70×10^{-04}	3.50	3.98
	7	57911569	57449895	58997016	LOD-3	8.25×10^{-04}	4.00	4.63
	9	9104920	6811794	12054464	LOD-3	-7.73×10^{-04}	3.54	4.12
	9	56918847	56449050	59619557	LOD-5	-1.03×10^{-03}	6.01	6.75
RDM	1	10768850	9381773	13906880	LOD-3	7.44×10^{-04}	3.33	4.38
	1	61202644	60641269	61573808	LOD-3	-7.63×10^{-04}	3.42	4.46
	2	58349902	57662551	61662915	LOD-3	-7.42×10^{-04}	3.32	4.21
	2	66100293	64442874	67258251	LOD-5	1.10×10^{-03}	7.21	9.12
	7	7032670	5624118	9033989	LOD-3	6.90×10^{-04}	2.80	3.66
	9	9844358	8553113	10312896	LOD-3	-7.40×10^{-04}	3.31	4.25
	9	56918847	56874994	56918855	Adjacent SNP	-1.49×10^{-03}	12.92	15.64
Ps	1	13492478	13289672	14688167	LOD-3	3.97×10^{-03}	6.13	7.12
	1	61629352	60640129	63750829	LOD-3	-3.17×10^{-03}	3.77	4.49
	2	64352754	62944225	67359000	LOD-3	2.99×10^{-03}	3.47	4.11
	3	8064255	7974609	8757559	LOD-3	3.46×10^{-03}	4.67	5.54

Trait	Chr	Position (bp)	Inferior limit (SNP - bp)[a]	Upper limit (SNP - bp)[b]	Criterion[c]	Effect[d]	h² QTL (%)[e]	LOD[f]
Ps	7	55196250	54739623	56920600	LOD-1,5	2.36×10^{-03}	2.16	2.59
	9	8613429	6837044	10417582	LOD-3	-2.83×10^{-03}	3.12	3.70
	9	56652039	56651990	56652040	Adjacent SNP	-4.50×10^{-03}	7.68	8.80
Pr	1	9381773	7130200	12766880	LOD-5	3.34×10^{-03}	5.96	6.69
	2	66356099	65316522	67309204	LOD-5	3.97×10^{-03}	7.88	8.72
	6	47732349	47412388	49954158	LOD-5	-3.27×10^{-03}	5.14	5.81
	9	56874994	56874838	57236931	Adjacent SNP	-4.69×10^{-03}	10.57	11.55

[a]Inferior limit for confidence interval. [b]Upper limit for confidence interval. [c]Criterion adopted to determine of confidence interval. [d]Positive effect indicates that favorable allele was donated by SC283, and negative effect indicates favorable allele donated by BR007. [e]The proportion of the genetic variance explained by each QTL (h² QTL, %). [f]Log-of-the odds. Chr: chromosome; Gy: grain yield (kg ha-1); FT: flowering time (days); PH: plant height (cm); Pp: plant phosphorus content (kg ha-1); Pg: grain phosphorus content (kg ha-1); RL: root length (cm); RD: root diameter (mm); SA: total root surface area (cm²); SA2: surface area of fine roots between 1-2 mm in diameter (cm²); SA3: surface area of thicker roots between 2-4.5 mm in diameter (cm²); RV: root volume (cm³); V2: volume of fine roots between 1-2 mm in diameter (cm³); SDM: shoot dry matter (g); RDM: root dry matter (g); Ps: shoot phosphorus content (g); Pr: root phosphorus content (g).

Additional file 5 Detailed information of the QTLs detected by multi-trait QTL mapping.

QTL	Position (bp)	Inferior limit (SNP - bp)[a]	Upper limit (SNP - bp)[b]	Criterion[c]	Effect[d]			h^2_{QTL} (%)[e]			LOD[f]
					Gy	RD	SA2	Gy	RD	SA2	
Gy-1	55802244	55133968	56732121	LOD-3	0.22	-	-	4.76	-	-	8.18
RD/SA2-2	65506207	63980832	67203635	LOD-5	-	0.20	0.29	-	3.96	8.25	11.56
Gy-3	1617914	1240416	2028420	LOD-1.5	0.13	-	-	1.75	-	-	3.19
SA2-3	7458660	6872695	7960595	LOD-5	-	-	0.27	-	-	7.19	7.52
Gy/SA2-3	71015638	70804761	71852763	LOD-5	0.21	-	0.23	4.26	-	5.30	10.34
Gy-4.1	4059198	4059197	4060107	Adjacent SNP	-0.13	-	-	1.83	-	-	3.29
Gy-4.2	57829755	56152948	60859112	LOD-5	0.18	-	-	3.03	-	-	5.13
RD-5	5541663	55259386	55655665	LOD-1.5	-	0.20	-	-	3.80	-	3.44
Gy-6.1	1515389	1471969	1574949	LOD-3	-0.15	-	-	2.17	-	-	3.38
Gy-6.2	42531422	42316732	44569495	LOD-3	-0.19	-	-	3.58	-	-	5.46
Gy-6.3	61839737	61183372	61968898	LOD-3	-0.18	-	-	3.24	-	-	5.59
Gy/RD-7	3672580	2132016	4813345	LOD-5	0.14	0.15	-	2.01	2.36	-	5.90
SA2-8	1093993	891531	1206164	LOD-1.5	-	-	-0.20	-	-	3.97	4.31
Gy-9	57236934	56918847	57337475	LOD-3	-0.50	-	-	24.73	-	-	35.66

[a]Inferior limit for confidence interval. [b]Upper limit for confidence interval. [c]Criterion adopted to determine of confidence interval. [d]Positive effect indicates that favorable allele was donated by SC283, and negative effect indicates favorable allele donated by BR007. Trait values were standardized by subtracting from each value the trait mean and dividing by the trait standard deviation. [e]The proportion of the genetic variance explained by each QTL (h^2 QTL, %). [f]Log-of-the odds .Chr: chromosome; Gy: grain yield (kg ha^{-1}); RD: root diameter (mm); SA2: surface area of fine roots between 1-2 mm in diameter (cm^2). The first number after of the trace (-) in the codification of the QTLs indicates the chromosome where they were mapped and the second number distinguish QTLs mapped in the same chromosome.

Additional file 6 Synteny between sorghum, Arabidopsis and maize for the main QTLs detected in this study

SORGHUM							ARABDOPSIS			MAIZE			Maize QTL mapping					
Chr	Trait	Pos QTLs single-trait (Mb)	Pos QTLs multi-trait (Mb)	Gene ID[a]	Pos (Mb)[a]	Candidate gene	Gene ID	Synteny sorghum × Arabdopsis[b] Score	Synteny sorghum × Arabdopsis[b] Similarity (%)	Gene ID	Synteny sorghum × maize[b] Score	Synteny sorghum × maize[b] Similarity (%)	Chr	QTLs	*Bin*	Pos (Mb)	Markers / Pos (Mb)		References
3	Gy	2.63	1.62	*Sb03g006765*	7.00	*SbPSTOL*	-	-	-	*GRMZM2G412760 (ZmPSTOL3.04)*	622	87.8	3	*qMulti3.04*	3.04	20.17	ZmPSTOL3.04 20.20	PHM5502_31 67.20	Azevedo et al. (2015)
	SA2	7.46	7.46																
3	Gy	69.72	71.02	*Sb03g031690*	60.12	*SbPSTOL*	-	-	-	*GRMZM2G451147 (ZmPSTOL8.05_1)*	590	48.4	8	*qRL8.05*	8.05	152.04	PHM934_19 116.80	ZmPSTOL8.05_1 152.00	Azevedo et al. (2015)
	SA2	73.36																	
4	SA2	57.25	57.83	*Sb04g028170*	58.26	*PIN1/PIN6*	*AT1G73590*	826	80.8	*GRMZM2G074267*	1114	96.4	5	*cQTL5-5*	5.07	206.78	- 207.16	- 208.07	Zhang et al. (2014)
				Sb04g028550	58.62	*CRE1*	*AT2G01830*	1088	73.9	*GRMZM2G151223*	1872	95.3	5	*cQTL5-5*	5.07	205.60	- 207.16	- 208.07	Zhang et al. (2014)
	Gy	57.83																	
				Sb04g026610	56.5	*ARF10*	*AT2G28350*	647	62.9	*Zm00008a022376*	1196	94.6	5	*cQTL5-3*	5.05	183.94	- 175.33	- 180.15	Zhang et al. (2014)
6	Gy	42.53	42.53	*Sb06g016260*	44.73	*ALMT1*	*AT1G08430*	338	59.4	*GRMZM5G858653*	765	93.0	10	*cQTL10-1*	10.04	114.98	- 113.90	- 115.20	Zhang et al. (2014)
7	Gy	2.55	3.67	*Sb07g002840*	3.00	*SbPSTOL*	-	-	-	*AC193632.2 (ZmPSTOL4.05)*	502	46.4	4	*qRD4.05*	4.05	39.79	PHM15427_11 33.90	PHM3587-6 59.40	Azevedo et al. (2015)
	RD	3.61																	
9	Gy	56.92	57.24	*Sb09g028110*	57.08	*PHO2*	*AT2G33770*	723	65.4	*GRMZM2G381709*	1641	96.7	6	*cQTL6-2*	6.07	164.04	- 162.23	- 163.82	Zhang et al. (2014)

[a] Gene ID and physical positions were based on version 1.4 of the sorghum genome.[b] Synteny analysis in Arabidopsis and Maize was undertaken in phytozome with v3.1 of the sorghum genome using a key word search (based on v1.4 sorghum gene IDs) or the amino acid sequence for Sb03g006765. Chr: chromosome; Pos: Physical position; Gy: grain yield (kg ha-1); SA2: surface area of fine roots between 1-2 mm in diameter (cm²); RD: root diameter (mm).

Additional file 7 Primers used for qPCR-RT assays

Gene ID	Primer ID	Sequence 5'- 3'
Sb03g006765	RTPupSbH602	CGCCGACGATGAACATCTC
	RTPupSbH701	TTGGCTCTGCTGAAGACGAA
Sb03g031690	RTPupSb504	CGCTCCTCCTTGCTGTCTTG
	RTPupSb601	TGTAATCGTCGTCGGAAGGAT
Sb07g002840	RTPupSbH102	CACCAGCCTCGATTTCATACAA
	RTPupSbH103	AGCCGCACCGGAAGTAGAC
18s rRNA	Sb18s_F	AATCCCTTAACGAGGATCCATTG
	Sb18s_R	CGCTATTGGAGCTGGAATTACC

CAPÍTULO 2

ASSOCIATION MAPPING AND GENOMIC SELECTION FOR SORGHUM ADAPTATION TO TROPICAL SOILS IN A SORGHUM MULTIPARENTAL RANDOM MATING POPULATION

RESUMO

Bernardino, Karine da Costa, D.Sc., Universidade Federal de Viçosa, agosto de 2019. **Mapeamento associativo e seleção genômica para adaptação a solos tropicais em uma população multiparental de acasalamento ao acaso de sorgo.** Orientador: Pedro Crescêncio Souza Carneiro. Coorientador: Jurandir Vieira de Magalhães.

Solos tropicais ácidos com baixa disponibilidade de fósforo (P) e toxidez por alumínio (Al) prejudicam a produção de sorgo [*Sorghum bicolor* (L.) Moench] nos países em desenvolvimento. Para lidar com isso, nós relatamos as propriedades genéticas da BRP13R, uma população multiparental de cruzamentos aleatórios, parcialmente endogâmica (RMP), que é comumente usada na seleção recorrente de sorgo. A recombinação intensiva dissipou grande parte da subestrutura da população original da BRP13R. O desequilíbrio de ligação (LD) decaiu atingindo níveis basais notavelmente homogêneos a aproximadamente 2,5 Mb, estabelecendo a BRP13R como um meio termo entre populações biparentais e painéis de associação de sorgo. A partir da análise de mapeamento associativo (GWAS) identificou-se QTLs conservados em outros estudos, notadamente de efeito aditivo para morfologia radicular e de efeito dominante para rendimento de grãos em condições de baixo P. Ao sobrepor regiões consenso de QTL, mapeamos dois genes de eficiência de P, candidatos nas regiões de ~ 5 Mb nos cromossomos 6 (*ALMT*) e 9 (*PHO2*). Concluímos que apenas 200 progênies genotipadas com cerca de 45.000 marcadores na BRP13R podem levar à clonagem posicional, baseada nos resultados do GWAS, de alelos raros, como os observados para tolerância a Al, condicionada pelo gene *SbMATE*. Verificou-se que a seleção genômica (GS) é útil em tais RMPs, particularmente se os marcadores em LD com genes principais são ajustados como efeitos fixos em modelos GBLUP que acomodam dominância. Mudanças nas frequências alélicas em progênies contrastantes para o rendimento de grãos indicaram que os genes de menor efeito na eficiência de P, como os genes *SbPSTOL1*, podem entrar no pré-melhoramento baseado em GS via mineração de alelos. Portanto, as RMPs, como a BRP13R, emergem como recursos polivalentes, permitindo a descoberta e a implantação eficiente de genes, beneficiando a segurança alimentar global através de cultivares de sorgo com ampla adaptação a solos tropicais.

Palavras-chave: Solos ácidos. Tolerância ao alumínio. Eficiência de fósforo. Estresse abiótico. GWAS.

ABSTRACT

Bernardino, Karine da Costa, D.Sc., Universidade Federal de Viçosa, August, 2019. **Association mapping and genomic selection for sorghum adaptation to tropical soils in a sorghum multiparental random mating population**. Adviser: Pedro Crescêncio Souza Carneiro. Co-advisers: Jurandir Vieira de Magalhães.

Tropical soils where low phosphorus (P) and aluminum (Al) toxicity limit sorghum [Sorghum bicolor (L.) Moench] production are widespread in the developing world. We report here on BRP13R, a multiparental random mating population (MP-RMP), which is commonly used in sorghum recurrent selection targeting tropical soil adaptation. Intensive recombination dissipated much of BRP13R´s original population structure and average linkage disequilibrium (LD) persisted up to 2.5 Mb, establishing BRP13R as a middle ground between bi-parental populations and sorghum association panels. Genome-wide association mapping (GWAS) identified conserved QTL from previous studies such as for root morphology and grain yield under low-P. By overlapping consensus QTL regions, we mapped two candidate P efficiency genes to a ~5 Mb region on chromosomes 6 (*ALMT*) and 9 (*PHO2*). Remarkably, we find that only 200 progeny genotyped with ~45,000 markers in BRP13R can lead to GWAS-based positional cloning of naturally rare, subpopulation-specific alleles, such as for *SbMATE*-conditioned Al tolerance. Genomic selection (GS) was found to be useful in such MP-RMP, particularly if markers in LD with major genes are identified by GWAS and fitted as fixed effects into GBLUP models accommodating dominance. Shifts in allele frequencies in progeny contrasting for grain yield indicated that intermediate to minor-effect genes on P efficiency, such as *SbPSTOL1* genes, can be employed in pre-breeding via allele mining in the base population. Therefore, MP-RMPs such BRP13R emerge as multipurpose resources for efficient gene discovery and deployment, benefiting global food security via sorghum cultivars with broad adaptation to tropical soils.

Keywords: Acidic soils. Aluminum tolerance. Phosphorus efficiency. Abiotic stress. GWAS.

Introduction

Crop adaptation to tropical soils relies on tolerance to multiple abiotic stresses rather than to a single stress condition. Hence, populations amenable for the simultaneous detection of favorable alleles at multiple tolerance loci and for selecting transgressive progeny are needed. Here, we explore the potential of using BRP13R, a sorghum multiparental random mating population (MP-RMP) constructed based on the nuclear male sterility gene, *ms*$_3$ (Webster 1965), for such an endeavor. Populations such as BRP13R are commonly used in recurrent selection schemes in crop pre-breeding, which may potentially narrow the gap between gene discovery and applications in cultivar development.

Acidic soils (pH $\leq$ 5) are prevalent in the tropics and sub-tropics, occupying more than half of the world arable lands (Von Uexküll and Mutert 1995). In sub-Saharan Africa, where sorghum is a staple food, 25% of the soils are acidic (FAO 2015; Tully et al. 2015). The highly weathered nature of acidic soils results in enrichment of aluminum (Al) and iron (Fe) oxides in the soil clay fraction (Shaw 2001), which is a central aspect leading to multiple abiotic stresses that significantly reduce crop yields, and hence food security worldwide (reviewed by Magalhaes et al. (2018)).

Under low pH, Al solubilizes into its ionic form, Al^{3+}, which damages the root system and impairs root growth (Kochian 1995) into deeper soils layers. Therefore, Al toxicity reduces grain yield due to restricted uptake of mineral nutrients and water (Foy et al. 1993). Due to Al toxicity in an acidic soil, we showed that sorghum grain yield was reduced by about 23% or one ton ha^{-1}, compared to an adjacent non-Al toxic field site (Carvalho et al. 2016). Phosphorus (P) diffusion on tropical soils is strongly constrained due to the formation of stable complexes between P and soil Al and Fe oxides (Marschner 1995; Lynch 2011), leading to very low P availability for crop uptake. Furthermore, P diffusion is severely limited by reductions in soil water content on tropical soils, even when those that are still not nearly

enough to cause drought stress (Ruiz et al. 1988). Hence, for non-irrigated crops cultivated on acidic soils, P stress is also a common limiting factor during the crop growth cycle. Therefore, acidic soil impact on crop yields results from a rather complex interplay of different abiotic stresses, which is further worsened by the often ubiquitous occurrence of drought stress.

Some of the molecular determinants and related physiological mechanisms that contribute to sorghum adaptation to acidic tropical soils have been revealed. The Al-activated citrate transporter, *SbMATE*, which mediates sorghum Al tolerance by promoting Al detoxification via citrate release into the rhizosphere (Magalhaes et al. 2007), has been shown to increase grain yield by over one ton ha^{-1} for both sorghum lines and hybrids harboring superior *SbMATE* alleles, when grown in an Al-toxic soil (Carvalho et al. 2016). In addition, *SbMATE*-specific single nucleotide polymorphism (SNP) markers have been associated with grain yield under low-P availability in West Africa, suggesting a pleiotropic effect of *SbMATE* also enhancing P acquisition (Leiser et al. 2014). We also found that sorghum homologs of the rice (*Oryza sativa*) serine/threonine receptor kinase, *phosphorus starvation tolerance1* (*OsPSTOL1*) (Gamuyao et al. 2012), were associated with root morphology changes, such as increased root surface area, leading to grain yield increases under low-P availability in the soil (Hufnagel et al. 2014; Bernardino et al. 2019). In addition, either quantitative trait loci (QTLs) or anonymous SNP loci associated with abiotic stress tolerance in sorghum (Mace and Jordan 2011; Leiser et al. 2014; Parra-Londono et al. 2018; Mace et al. 2019), including stay green QTLs that enhance grain yield under drought stress (Harris et al. 2006; Sabadin et al. 2012), are expected to lead to the isolation of novel abiotic stress tolerance genes in sorghum.

In order to efficiently integrate multiple abiotic stress tolerance loci into sorghum breeding, detection strategies and appropriate target populations should be carefully designed. Provided that proper attention is directed to the occurrence of false positives, the population

flexibility provided by association mapping approaches (Yu and Buckler 2006) can facilitate tolerance loci detection directly on the breeder's germplasm, within a multi-allelic context, which can facilitate progeny selection. In the case where inferences are made directly in the target population, more readily available applications for crop improvement can be expected (Breseghello and Sorrells 2006). Although they explore a narrower allelic range with in general less resolution, genetic mapping using bi-parental crosses, such as with recombinant inbred lines (RILs), is an important complementary approach to association mapping, particularly by providing higher detection power for quantitative trait locus (QTL) (Breseghello and Sorrells 2006). A middle ground between bi-parental crosses and association panels in terms of population structure, genetic diversity, the number of traits that can be investigated, resolution and power are provided by multiparental populations, such as Multiparent Advanced Generation Intercross populations (MAGIC) (Mackay and Powell 2007; Stadlmeier et al. 2018). Eight-parent MAGIC populations have been shown to capture a high proportion of the allelic diversity available in the German wheat breeding gene pool (Stadlmeier et al. 2018) and have been deemed adequate for high-resolution mapping of quantitative trait loci (Mackay et al. 2014). Nested association mapping (NAM) approaches, where diverse founders are crossed to a common parent to produce sets of mapping populations, minimize genetic background effects on QTL detection and increase detection power (Yu et al. 2008). Such approaches have been shown to lead to more consistent detection of phenology QTL compared to association mapping in sorghum (Bouchet et al. 2017), and to enhance detection of putative multiple small effect alleles influencing flowering time (Mace et al. 2013).

We focus here on exploring the consequences of enhanced recombination via randomly mating multiple parents repeatedly throughout the genesis of BRP13R, focusing on simultaneously detecting loci related to abiotic stress tolerance by GWAS and deploying

previously identified tolerance loci into a pre-breeding pipeline. In the context of BRP13R, we also investigate the potential of genomic prediction as a tool to assist sorghum breeding efforts with the final goal of selecting progeny with broad adaptation to tropical soils with low P availability and Al toxicity.

Materials and Methods

Genetic material

Male sterile plants (*ms3ms3*) from the Nebraska Random Mating Population 3 (NRP3R) were crossed with 100 sorghum fertility restorer (R) lines from the world collection selected for grain protein content, giving rise to the Purdue Population 3R (PP3R, Robert Schaffert, personal communication). PPR3 was then subjected to 6 cycles of recombination to generate the Brazilian Random Mating Population 3 (BRP3R); subsequently, male-sterile (*ms3ms3*) plants from BRP3R were crossed with 24 R-lines selected for Al tolerance, P efficiency, and other desirable traits (Table S1). This random mating population was designated BRP13R. Fertile F1 plants were self-pollinated to produce F2 seeds, which were recombined (first recombination cycle). Seeds of sterile plants were harvested in bulk for the second recombination cycle. After the third recombination cycle, approximately 600 seedlings were phenotyped for Al tolerance in nutrient solution.

Al tolerant plants were selected and transplanted to pots in the greenhouse. Fertile and sterile plants were then self-pollinated or crossed with a composite pollen sample of the population, respectively. Seeds were harvested in bulk and planted in an isolated field for recombination purposes. Seeds derived from the sterile plants produced BRP13R S_0 progeny. Two hundred and ten fertile S_0 plants, with plant height between 100 and 150 cm (*Ms3ms3*) were self-pollinated producing $S_{0:1}$ progeny. One fertile plant of each $S_{0:1}$ progeny was self-pollinated, originating $S_{0:2}$ progeny.

Phenotyping

Phenotyping in a low-P soil

Two field trials were conducted at the experimental station of Embrapa Maize and Sorghum, in Sete Lagoas, Minas Gerais, Brazil, during the summer season of 2014. The experimental area is a weathered tropical soil with low-P availability, containing 2.57 ppm P (± 0.57 standard deviation, s.d.) (Mehlich 1) in the top soil (0 -20 cm) and 1.25 ppm P (± 0.30 s.d.) in the subsoil (20 - 40 cm). Two hundred $S_{0:2}$ progeny were arranged in two 12 (progeny) x 10 (incomplete block) alpha lattice designs with three replicates. Each block contained 10 progeny and BR007 and SC283, which were used as checks. Each plot consisted of two 3-m rows, with 0.45 m between rows and 8 plants m^{-1}. Fertilization was applied as 150 kg ha^{-1} of 20-00-20 (NPK) at sowing and 200 kg ha^{-1} of urea 30 days after.

The traits measured were: grain yield (Gy, ton ha^{-1}), flowering time (FT, days), plant height (PH, cm), plant phosphorus content (Pp, ton ha^{-1}), grain phosphorus content (Pg, ton ha^{-1}), total phosphorus content (Pt, ton ha^{-1}), plant dry matter (PDM, ton ha^{-1}) and grain dry matter (GDM, ton ha^{-1}). Plant and grain tissues, collected by plot, were dried at 65^{o} C until constant weight, ground and homogenized, and P content was assessed in 20 g subsamples using inductively-coupled argon plasma emission spectrometry.

Phenotyping of root system morphology in nutrient solution with low P availability

Assessment of root system morphology under low-P was undertaken in nutrient solution as described by Sousa et al. (2012) and Hufnagel et al. (2014) in a randomized block design with three replicates. Seeds were sterilized with sodium hypochlorite (5%), washed with distilled water and germinated in paper rolls. After four days, uniform seedlings of each progeny were transferred to moistened germination papers placed in paper pouches (24 x 33 x 0.02 cm) (Hund et al. 2009).

Each experimental unit consisted of one pouch with three seedlings per pouch, whose bottom (3 cm) was immersed in containers with 5 L of nutrient solution as described by Magnavaca et al. (1987) at pH 5.65 and 2.5 µM P. The containers were kept in a growth chamber for 13 days with 12 hours of photoperiod, 27 °C day and 20 °C night, and continuous aeration.

After 13 days, the root system was photographed with a digital camera Nikon D300S SLR, and the obtained images were analyzed with the RootReader2D (http://www.plantmineralnutrition.net/software/rootreader2d/) and WinRhizo (http://www.regent.qc.ca/) softwares. The traits measured were: root length (RL, cm); root diameter (RD, mm); total root surface area (SA, cm^2); surface area of super fine roots (SA1, cm^2 - 0 mm < RD ≤ 1 mm); surface area of fine roots (SA2, cm^2 - 1 mm < RD ≤ 2 mm); surface area of thicker roots (SA3, cm^2 - 2 mm < RD ≤ 4.5 mm); root volume (RV, cm^3); volume of fine roots (V2, cm^3 - 1 mm < RD ≤ 2 mm); shoot dry matter (SDM, g); root dry matter (RDM, g); shoot phosphorus content (Ps, g); and root phosphorus content (Pr, g).

Al tolerance in nutrient solution

Al tolerance was assessed in nutrient solution by measuring Al-inhibition of root growth as described in Caniato et al. (2007). Seed sterilization and germination was as described above but with a 3-day germination period. After germination, uniform seedlings were transferred to containers (49 seedlings per container) in a growth chamber with a photoperiod of 12 hours, 27 °C day and 20 °C night temperatures under continuous aeration without stress.

After 24 hours, the nutrient solution of half of the trays was replaced by an identical solution without Al (control containers) whereas the remaining trays received nutrient solution with {27} µM Al^{3+} (braces indicate Al^{3+} activity). Aluminum was supplied as

$AlK(SO_4)_2$ and the solution pH was adjusted for 4.0 with HCl. The experimental design was an augmented block, in which seven seedlings constituted one experimental plot. Each tray represented one block with seven plots, containing also four seedlings Al-sensitive (ATF13) and three Al-tolerant (ATF14) as controls.

The initial root length (IRL), the final root length after five days (FRL_{5d}), and net root growth (NRG = FRL_{5d} - IRL) were recorded and relative net root growth (RNRG) was calculated by dividing the NRG Al treatment by the NRG without Al.

Statistical analysis

The model adopted for traits assessed in a low-P soil was:

$$y_{ijkl} = \mu + E_j + R_{k(j)} + B_{l(kj)} + G_i + \varepsilon_{ijkl}$$

y_{ijkl} is the phenotypic value of progeny i in the block l of the k^{th} replicate, within the experiment j; μ is the overall mean; E_j is the fixed effect of the j^{th} experiment $(j = 1, 2)$; $R_{k(j)}$ is the fixed effect of replicate k $(k = 1, \dots, 3)$ in experiment j; $B_{l(kj)}$ is the random effect of block l $\left(l = 1, \dots, 10,\ b_l \sim N(0, \sigma_b^2)\right)$ in the replicate k, within the experiment j; G_i is the genetic effect of progeny i, which can be defined as:

$$G_i = \begin{cases} g_i & i = 1, \dots, n_g \\ t_i & i = n_g + 1, \dots, n_g + n_c \end{cases}$$

g_i is the random effect of progeny i with n_g being the total number of progeny $(g_i \sim N(0, \sigma_g^2)$; t_i is the fixed effect of check i with n_c being the total number of checks. ε_{ijkl} is the experimental error for progeny i in the block l of the k^{th} replicate within the experiment j, assuming $\varepsilon_{ijkl} \sim N(0, \sigma_e^2)$.

The model used for analyzing the hydroponic experiments with low-P conditions was:

$$y_{ij} = \mu + B_j + g_i + \varepsilon_{ij}$$

where y_{ij} is the phenotypic value of the progeny i $(i = 1, \dots, n_g)$ in the block j; μ is the

overall mean; B_j is the fixed effect of block j ($j = 1, \ldots, 3$); g_i is the random genetic effect of progeny i ($g_i \sim N(0, \sigma_g^2)$; and ε_{ij} is the experimental error for progeny i in the block j ($\varepsilon_{ij} \sim N(0, \sigma_e^2)$).

The model used for analyzing the hydroponic experiments with aluminum stress was:

$$y_{ij} = \mu + B_j + G_i + \varepsilon_{ij}$$

where y_{ij} is the phenotypic value of the progeny i ($i = 1, \ldots, n_g$) in incomplete block j; μ is the overall mean; B_j is the fixed effect of incomplete block j ($j = 1, \ldots, 35$), G_i is the genetic effect of progeny i, which can be defined as:

$$G_i = \begin{cases} g_i & i = 1, \ldots, n_g \\ t_i & i = n_g + 1, \ldots, n_g + n_c \end{cases}$$

g_i is the random effect of progeny i with n_g being total number of progeny ($g_i \sim N(0, \sigma_g^2)$; t_i is the fixed effect of check i with n_c being the total number of checks; and ε_{ij} is the experimental error for progeny i in the block j, assuming $\varepsilon_{ij} \sim N(0, \sigma_e^2)$.

Fixed and random effects were tested using the Wald statistics (Wald 1943) and the likelihood ratio test (LRT) (Neyman and Pearson 1928) respectively, considering a 5% significance level (α). For all statistical models, the genetic effect of progeny was first taken as random for estimating the genetic variance component (σ_g^2) via restricted maximum likelihood (REML), and the heritability coefficient of each trait. The effect of progeny was then considered as fixed for estimating the adjusted means using best linear unbiased estimators (BLUEs) using the ASReml-R package (Butler et al. 2009). Generalized heritabilities (h^2) were estimated as proposed by Cullis et al. (Cullis et al. 2006):

$$h^2 = 1 - \frac{\bar{v}BLUP}{2\sigma_g^2}$$

where $\bar{v}BLUP$ is the average variance of the difference between two best linear unbiased predictions (BLUPs). Person's correlation coefficients (Pearson 1895) were estimated based

on adjusted means using the package *Hmisc* (Harrell Jr 2015) in R software (R Core Team 2016).

Genotyping

Genomic DNA was isolated from 500 mg of vegetal tissue (eight plants per progeny), as described by Saghai_Maroof et al. (Saghai-Maroof et al. 1984). DNA samples were genotyped by sequencing, according to Elshire et al. methodology (Elshire et al. 2011). DNA fragments ("reads") obtained during genotyping were aligned against the sorghum reference genome (version 2.1), using the Burrows Wheller Aligner (BWA) (Li and Durbin 2009) software and SNP calling was performed with the GBS pipeline (Glaubitz et al. 2014) in TASSEL V (Bradbury et al. 2007). $S_{0:2}$ progeny were genotyped with *SbPSTOL1*- and *SbMATE-specific* markers (Caniato et al. 2014; Hufnagel et al. 2014) using the Allele Specific PCR genotyping system (KASP, LGC genomics) (Robinson 2006).

Marker imputation

Missing data were imputed with Beagle (Browning and Browning 2007), which has been reported to show higher imputation accuracy for heterozygous populations and reduced computation time compared to other procedures (Nothnagel et al. 2009; Swarts et al. 2014).

At least two reads from different sister chromatids are needed for correctly calling a heterozygous genotype (Swarts et al. 2014). Thus, the probability of miscalling heterozygous genotypes is related to the read-depth and can be estimated as $P(\mathrm{AA}|\mathrm{Aa}) + P(\mathrm{aa}|\mathrm{Aa}) = 0.5^n + 0.5^n$, where AA and aa represents genotypes homozygous for the most and least frequent alleles and n is the sequencing depth (Swarts et al. 2014). Based on that, for heterozygous genotypes with read-depth 5, 6 and 7, miscalling percentages are 6.25%, 3.125%, and 1.5625, respectively. As the median read-depth prior to imputation in BRP13R was 5, while selecting

genotypes with read-depth > 5 leads to enhanced imputation accuracy for heterozygotes, it also decreases the total number of markers left for GWAS. We thus set out to identify the imputation conditions that would balance the trade-off between imputation accuracy and the total number of markers left. Imputation accuracy was first calculated selecting loci with read-depths ≥ 5, 6 and 7, window sizes (i.e. physical distance used for haplotype inference) between 10 Kb to 10 Mb and with no filtering for missing data or selecting loci with at most 25%, 50%, 75% missing data.

Accuracy tests were performed with a total of 146,306 biallelic and polymorphic GBS SNPs. Masking was undertaken by randomly replacing twenty percent of data that had genotypic information (homozygous and heterozygous classes) for missing data. Upon imputation, accuracy was calculated by comparing imputed genotypes with the "real", observed genotypes. Finally, loci with MAF < 0.01 were eliminated.

Population structure and relatedness

The genetic relationship or kinship matrix (K) was obtained by the Identity-by-state approximation, proposed by Endelman and Jannink (2012), with TASSEL V (Bradbury et al. 2007). Genetic divergence between progeny was calculated in R based on the Euclidean distance and clustering was undertaken with the unweighted pair group method with arithmetic mean (UPGMA) method (Sokal and Sneath 1963). We also undertook a principal component analysis based on 43,825 SNP markers to investigate the degree of population structure remaining in BRP13R after recombination using the *pcaMethods* package (Stacklies and Redestig 2016) in R (R Core Team 2016).

Linkage disequilibrium

Linkage disequilibrium (LD) was calculated for each sorghum chromosome using squared genotypic correlations between pairs of loci (r^2) (Weir 2008) with the *Hmisc* package in R. To assess the extent of LD per chromosome, we first selected SNPs under significant LD based on a t-test ($\alpha = 0.05$) corrected for multiple tests based on the Bonferroni correction for the total number of pairs of SNP loci (0.05 / [total number of SNPs × (total number of SNPs-1)/2]). We then plotted the r^2 values of SNPs under significant LD as a function of the physical distance between pairs of SNPs. The extent of LD was determined as the physical region beyond which average r^2 reached constant, basal levels.

Association mapping

Adjusted means (BLUEs) for the different traits were used for GWAS. The association mapping analyses were performed only with markers whose genotypic classes showed frequencies above 0.05, totaling 24,485 markers. We fitted models with no correction for population structure nor relatedness (naïve model), including the kinship matrix (K), population structure (Q, with PC1 scores), or jointly incorporating population structure and relatedness (Q + K). The best model was chosen based on the Akaike information criterion (AIC) (Akaike 1973), the Bayesian information criterion (BIC) (Schwarz and others 1978) and on Type-I error simulation via inspection of the quantile-quantile (q-q) plots of the *p*-values from association analysis plotted against cumulative *p*-values. The significance threshold for GWAS was determined with a Bonferroni correction (Bland and Altman 1995), calculated by dividing an alpha level of 0.05 by the number of independent genome blocks based on the estimated LD extent per chromosome.

Genomic selection (GS)

Genomic selection was undertaken for grain yield (Gy, ha^{-1}), plant height (PH, cm), plant dry matter (PDM, ha^{-1}), and aluminum tolerance (RNRG). The genomic best linear unbiased prediction (GBLUP) models examined were:

i) Model 1 – GBLUP with the Additive genomic relationship matrix (GBLUP-A) estimated by the method proposed by Van Raden (VanRaden 2008);

ii) Model 2 – Model 1 incorporating the Dominance genomic relationship matrix (GBLUP-AD) estimated by the method proposed by Vitezica et al. (Vitezica et al. 2013);

iii) Model 3 – GBLUP-A incorporating Gene-specific SNP markers, for *SbPSTOL1* and *SbMATE* (Caniato et al. 2014; Hufnagel et al. 2014), as Fixed cofactors with additive genetic effects (GF-GBLUP-A);

iv) Model 4 – Model 3 incorporating the dominance genomic relationship matrix, and gene-specific SNP markers, for *SbPSTOL1* and *SbMATE*, as fixed cofactors with both additive and dominance genetic effects (GF-GBLUP-AD);

v) Model 5 – GBLUP-A with SNP markers associated to different traits by GWAS as fixed cofactors with additive genetic effects (GWAS-GBLUP-A);

vi) Model 6 – Model 5 incorporating the dominance genomic relationship matrix and SNP markers associated to different traits by GWAS as fixed cofactors with both additive and dominance genetic effects (GWAS-GBLUP-AD).

Additive and dominance genomic relationship matrices were calculated using the R package *AGHmatrix* (Amadeu et al. 2016). For models incorporating GWAS or gene-specific SNPs as fixed cofactors (models 3 to 6), SNPs within the same LD block were removed from the estimation process of the genomic additive and dominance relationship matrices. LD

blocks were defined based on the estimated LD extent per chromosome. The general model fitted was:

$$y = \mu 1 + X_1 a_f + X_2\, d_f + Z_1 a_r + Z_2 d_r + e$$

where $y(p \times 1)$ is a vector of adjusted means via BLUE, obtained by correcting the phenotypic progeny means for nuisance variables from the experimental design, for p progeny; μ is the overall mean; a_f and d_f are the vectors of additive and dominance fixed effects, respectively, for g *SbPSTOL1* and *SbMATE* genes, $a_f\ (g \times 1)$, or s GWAS SNPs, $a_f\ (s \times 1)$; $a_r(p \times 1)$ is the vector of random additive genetic effects of p progeny, with $a_r \sim N(0, A\sigma_a^2)$; $d_r(p\ \times 1)$ is the vector of random dominance genetic effects of p progeny, with $d_r \sim N(0, D\sigma_d^2)$; and $e(p\ \times 1)$ is the vector of residuals, with $e \sim N(0, I\sigma_e^2)$. $X_1(p \times g\ or\ p \times s)$, $X_2(p \times g\ or\ p \times s)$, $Z_1(p \times p)$ and $Z_2(p \times p)$ are incidence matrices for their respective effects, 1 is a vector of ones $(p \times 1)$, A and D are $p \times p$ additive and dominance genomic relationship matrixes, respectively, and I is a $p \times p$ identity matrix.

To avoid bias that could artificially inflate accuracy estimates, the SNP markers included as fixed effect cofactors in the genomic selection models v-vi were selected based on one hundred rounds of GWAS, using different populations constructed from randomly sampled 160 BRP13R progeny in each round. SNP loci with association signals exceeding the Bonferroni threshold in more than half of the GWAS rounds were selected as fixed cofactors. For genomic selection, BRP13R was randomly split into training/validation sets, and four distinct population sizes were compared: 100/100, 120/80, 140/60 and 160/40. The different GS models were fitted considering 100 replicates for each size of the training/validation sets. Then, the optimum size of training/validation sets was selected based on the maximization of the predictive accuracy, which was calculated as the correlation between the adjusted means obtained by BLUE and the predicted means via GBLUP models. For that, the adjusted means via BLUE of 40, 60, 80 or 100 progeny (dependent of the validation population size) taken at

random were masked and compared to the predicted means using the GBLUP models. All analyses were performed with ASReml-R (Butler et al. 2009). Only markers whose genotypic classes showed frequencies above 5% were used.

Results

Phenotypic analysis of BRP13R

We focused on the following performance traits on a low-P tropical soil field site: grain yield, plant dry matter and plant height assessed under low-P availability in the soil. In addition, we assessed root morphology traits related to P acquisition and Al tolerance in nutrient solution. Grain yield (Gy) under low-P was found to largely reflect P acquisition efficiency, which is the most important P efficiency component in sorghum (Bernardino et al. 2019). Among root traits, we previously established the importance of total root surface area (SA) and root diameter (RD) on grain yield under low P availability in the soil (Hufnagel et al. 2014; Bernardino et al. 2019). Finally, Al tolerance was assessed based on relative net root growth (RNRG) in hydroponics. Heritability estimates for root diameter and total root surface area were 0.37 and 0.51, respectively, 0.61 for grain yield and 0.57 for plant dry matter (PDM) (Table S2). Grain yield was highly correlated both with grain P content (Pg, r = 0.76) and total P (Pt, r = 0.63), as well as with plant dry matter (PDM, r = 0.56) (Table S3). Additionally, Al tolerance assessed in controlled conditions was highly heritable (0.73).

Genotyping-by-sequencing (GBS) and marker imputation

Single nucleotide polymorphism (SNP) markers distributed genome-wide were genotyped via GBS (Elshire et al. 2011). Before imputation, the median number of reads per genotype (read-depth) was 5. About half of the reads had a depth between 1 and 5, and 43% of the reads had depths exceeding 6 (Fig. S1). There was a tendency of higher read-depth

towards the end of the sorghum chromosomes compared to the centromeres, and average minor allele frequency (MAF) was in general low.

For genotypes covered by 5 reads, the probability of GBS to mistakenly call as homozygous a heterozygous genotype is 0.0625 (Swarts et al. 2014). Due to the partially heterozygous nature of half-sib progeny in BRP13R, we only kept homozygous genotypes with read-depth ≥ 6 to minimize miscalling of heterozygotes (expected miscalling frequency = 0.03), while keeping adequate marker coverage in the genome. We then replaced 20% of known genotypes by missing data (i.e. masking) and imputed missing data with Beagle (Browning and Browning 2007). Accuracy for all genotypic classes (global accuracy) was very high, about 97%, irrespective of the imputation window size (Table S4), and was the highest for genotypes homozygous for major alleles at SNP loci. For imputation, we selected a window size of 500 Kb, which maximized imputation accuracy for heterozygous genotypes. Inspection of our masking procedure indicated that imputation errors for heterozygous genotypes occasionally caused them to be imputed as genotypes homozygous for the major allele. Therefore, the main effect of the incorrect imputation of heterozygotes, which was likely due to their lower frequency and sparse distribution in the genome, was to reduce the frequency of this class. After imputation, BRP13R was found to consist of 20% heterozygous genotypes with and average MAF of 0.14. Imputation more than doubled the number of markers, totaling 43,825 markers, and improved genome coverage, especially in centromeric regions (Fig. 1).

Population structure and relatedness

Population structure may absorb phenotypic variance and reduce the detection power in association mapping (Kang et al. 2008). We used an Identical by State (IBS)-based method (Endelman and Jannink 2012) implemented in TASSEL V (Bradbury et al. 2007) to assess

genetic relatedness between half-sib progeny in BRP13R (Fig. 2a). The progeny kinship coefficients were tightly clustered around a mean value of 0.5 (0.5 ± 0.02). Consequently, the kinship heatmap was rather homogeneous, with absence of strongly differentiated groups in the population. Based on the kinship heatmap and on UPGMA clustering, five clusters were detected, but average kinship for these groups was in general only slightly above the population mean (between 0.53 and 0.58). Group 1 showed an average kinship (0.75) that was higher than the population average, but this group had only three progeny. Next, we conducted a principal component analysis with 43,825 SNP markers and plotted progeny scores for the first two principal components (Fig. 2b). The groups detected by UPGMA were in general separated by the two principal components, with groups 3 and 4 tending to overlap. Group 2 was the most well defined group whereas progeny within the other groups were rather disperse. The 24 restorer lines used in the formation of BRP13R comprised different morphological races and geographical origins, which largely govern population structure in sorghum (Caniato et al. 2011; Bouchet et al. 2012), as well as breeding materials (Table S1). In conjunction, the population structure and relatedness results indicate that the recombination cycles to which BRP13R was subjected resulted in little structure left in the population.

Linkage disequilibrium

Linkage disequilibrium (LD) was measured using squared genotypic correlations between pairs of loci (r^2) (Weir 2008). The number of SNPs under significant LD was plotted as a function of physical distance between pairs of loci, and LD extent per chromosome was determined as the physical distance after which average r^2 values reached constant, basal levels (Fig. 3). Based on this method, LD was found to decay in a remarkable homogeneous way across chromosomes, with LD extending to 2.5 Mb on average (± 0.5 Mb). LD persisted the longest on chromosome 6 (3.5 Mb), probably due to selection during breeding (Bouchet et

al. 2017) acting on the linked plant height and maturity loci, *Dw2* and *Ma1* (Sabadin et al. 2012), respectively. The shortest LD extent of 2 Mb was found for chromosomes 1, 2 and 3.

Association mapping

Model selection

We initially conducted a series of model selection steps to define the most adequate model for GWAS. Inspection of the Bayesian Information Criterion (BIC) (Schwarz and others 1978) determined that the naïve model (without correction for population structure nor relatedness) performed poorly in terms of goodness-of-fit to grain yield on low-P soil data (BIC = 10229, Fig. S2). The model including the kinship matrix (K) produced a slight decrease in model performance compared to the naïve model, which is likely due to the highly homogeneous relatedness among BRP13R progeny (Fig. 2a). The best performing model included the first PC of our principal component analysis (PC1, Fig. 2b). The PC + K model showed reduced performance compared to the PC model, indicating that including progeny scores for PC1 alone efficiently captured the remaining population structure in BRP13R. Next, for each tested model, the probability distribution under the null hypothesis was inspected based on the quantile-quantile (q-q) plots of the *p*-values from association analysis plotted against cumulative *p*-values (Fig. S2). Consistent with the model fitness results, we observed substantial inflation of type-I error in the naïve model compared to the PC, K and PC + K model. Both the K and PC + K models showed below-diagonal *p*-values, indicating reduction in detection power caused by the K matrix. Therefore, we selected the PC model for GWAS.

Genome wide association mapping

The significance threshold for GWAS was based on the Bonferroni correction for multiple tests (Bland and Altman 1995). The number of independent tests was defined based on the extent of linkage disequilibrium estimated for each sorghum chromosome (Fig. 3) and the resulting –log(*p*) threshold was 3.74 (*alpha* = 0.05). As an additional false positive control, GWAS was performed only with markers whose genotypic classes showed frequencies above 5%. The GWAS profiles for the selected traits are shown in Fig. 4 and the additive and dominance effects for the respective associated SNPs, in addition to effects for SNPs associated with different auxiliary traits, are shown in Table S5. We found in total 78 significant SNP loci (Fig. 4), within which 18 SNPs associated with grain yield on low-P soil were distributed across all sorghum chromosomes, except for chromosome 7. The grain yield effect for the associated SNPs varied from 110 to 430 kg ha^{-1} and explained 6.99 to 12.41 % of the phenotypic variance. About 65% of the SNPs whose effect could be partitioned between dominance and additivity (i.e. all three genotypic classes were present) had predominantly dominant effects (Table S5). This strongly contrasts with Al tolerance, which was controlled largely by SNPs acting additively. The SNP with the highest association signal and effect on Al tolerance is located at position 71.1 Mb on chromosome 3 and co-localizes with the major Al tolerance gene, *SbMATE* (Magalhaes et al. 2007). *SbMATE* has been previously shown to control Al tolerance in hydroponics and in the field in a semi-dominant and additive way, respectively (Magalhaes et al. 2004; Carvalho et al. 2016). The effects for SNPs associated with total root surface area were evenly partitioned into additive and dominance, whereas those associated with root diameter (RD) acted mostly in an additive manner.

Association mapping with gene-specific markers

We genotyped BRP13R with a set of *SbMATE*-specific markers and SNPs within the Alt_{SB} locus where *SbMATE* is located (Magalhaes et al. 2004, 2007), which were strongly associated with Al tolerance (Caniato et al. 2014). Genotyping was also performed with markers tagging *SbPSTOL1* genes, which are sorghum homologs of rice *phosphorus-starvation tolerance1* that have been previously associated with root morphology and/or grain yield under low P (Hufnagel et al. 2014). *SbMATE* SNPs (in red in Fig. 4), in addition to other GBS SNPs near *SbMATE* (*Sb03g043890*) at position 71.1 Mb, were highly associated with Al tolerance (Fig. 4, Table S5). In BRP13R, SNPs within *SbPSTOL1* genes were below our significance threshold. However, inspection of adjusted phenotypic means indicated that the *SbPSTOL1* alleles increasing grain yield reported in Hufnagel et al. (2014) were consistently associated with grain yield advantage in BRP13R. For example, within the *SbPSTOL1* gene *Sb03g006765*, the SNP loci (favorable allele in parenthesis), 1912 (A), 1998 (C), 2042 (G), 2067 (G), 2073 (C) and 2141 (T) were in complete LD in Hufnagel et al. (2014). In BRP13R, grain yield means for all the favorable alleles was consistently higher than that of the respective alternate alleles (Fig. S3). In addition, the A allele at the 1.541 SNP within *Sb03g031680* increased grain yield in Hufnagel et al. (2014) and the grain yield mean for this allele was again higher than the alternative allele in BRP13R (Fig. S3). This strongly suggested that these *SbPSTOL1* genes contribute to grain yield in BRP13R, although the narrower allelic diversity in the random mating population compared to the association panel (Hufnagel et al. 2014) may have reduced their association signals.

Next, we compared the positions of the QTLs detected in BRP13R to those previously detected in a large RIL population (Bernardino et al. 2019) and observed many instances of likely QTL conservation in the two populations (Table S6). Co-localized QTLs for grain yield within a 15 Mb window were found on chromosomes 4 (55 – 70 Mb), 6 (35 – 45 Mb), 8 (55 –

60 Mb), 9 (45 – 60 Mb) and 10 (5 – 20 Mb). We also found co-localized QTL for grain yield and P content under low P availability, such as a QTL for grain P content (Pg) on chromosome 1 (5 – 25 Mb), and grain yield/P content QTL on chromosomes 1 (55 – 65 Mb), 3 (0 – 5 Mb), 4 (0 – 10 Mb), 6 (0 – 5 Mb), 7 (0 – 10 Mb), 8 (55 – 65 Mb), 9 (55 – 60 Mb) and 10 (5 – 20 Mb). Over half of the grain yield QTL detected in BRP13R co-localize with root morphology QTL, mainly with QTL for root surface area and occasionally with root diameter QTL, which supports the importance of root morphology in P acquisition on low P soils (Bernardino et al. 2019). We detected more QTLs for root surface area in BRP13R in comparison to the RIL analyses, but clear instances of conserved QTL were also observed, for example, at 60 – 70 Mb on chromosome 2 and at 5 – 15 Mb and 65 – 75 Mb on chromosome 3, among other cases.

Genetic makeup of selected progeny

Next, we explored the genetic constitution of BRP13R progeny selected for grain yield using a 10% selection pressure for high (designated henceforth as top 10% for simplicity) and low (bottom 10%) grain yield (Fig. 5). This analysis was undertaken with markers associated with grain yield via GWAS (Fig. 5a) as well as with our gene-specific markers for *SbSPTOL1* genes and *SbMATE* (Fig. 5b). The frequency of heterozygotes for SNP loci associated with grain yield was much higher in the top 10% of the BRP13R progeny compared to low yielding progeny, whereas fewer loci homozygous for the favorable alleles were present in top 10% progeny (Fig. 5a). In addition, the top 10% group had much fewer progeny that were homozygous for the inferior allele. In conjunction, these results are consistent with the predominance of dominance effects for SNPs associated with grain yield and suggest possible relevance of overdominance on grain yield (Additional file 7). In contrast, for gene-specific markers, both loci that are homozygous for the favorable allele as well as heterozygotes were

more frequent in top 10% progeny compared to low yielding progeny (Fig. 5b). The frequency of loci in homozygosity for the inferior allele was, in turn, higher in the low-yielding progeny. We also looked at allele frequencies per gene and calculated frequency shifts between high- and low-yielding BRP13R progeny as Δ_f (Fig. S4). Based on this analysis, all *SbPSTOL1* genes showed increased frequency of the favorable allele in high yielding progeny. This frequency divergence was the highest in *Sb03g006765* and *Sb03g31680* and neglectable for the Al tolerance gene, *SbMATE*. An analysis of molecular variance confirmed that the high and low-yielding groups differed for allele frequencies ($p < 0.10$) for both GWAS- and gene-specific loci within the *SbPSTOL1* genes, *Sb03g006765*, *Sb03g031680*, which are associated with the highest Δ_f for grain yield (Fig. S4).

Genomic selection

Due to the intrinsically quantitative nature of sorghum adaptive traits to abiotic stresses on tropical soils, we explored adequacy of BRP13R for genomic selection (GS), targeting grain yield under low-P availability in the soil. For that, we used models that accommodate dominance effects to take advantage of the residual heterozygosity in the multiparental population. We also studied whether the inclusion of loci associated with grain yield by GWAS as fixed effects (GWAS-SNPs) could increase prediction accuracies via GBLUP. To avoid artificially inflating accuracy estimates, the population used for GWAS to identify SNP cofactors was different than that used for genomic selection. First, SNP loci most frequently associated with grain yield after multiple rounds of GWAS, conducted in different BRP13R subsets of randomly sampled 160 progeny, were selected as cofactors. Then, for genomic selection, BRP13R was randomly split into training and prediction sets, which consisted of 100/120 and 100/80 progeny, respectively. Genomic selection was also undertaken in multiple rounds, varying the constitution of the training and the prediction sets

across rounds. Accuracy was calculated as the correlation between the adjusted and the predicted means via BLUE and GBLUP, respectively, for grain yield.

In the absence of dominance effects and fixed cofactors (GBLUP-A), prediction accuracy varied from 0.22 for grain yield to 0.35 for Al tolerance (RNRG) and the traits with the highest heritability (plant height and Al tolerance, h^2 = ~0.75) also showed the highest accuracies (Fig. 6 and Table S7). Inclusion of dominance effects (GBLUP-AD) increased prediction accuracies for grain yield and plant dry matter, but only slightly. There was no advantage in using gene-specific markers for *SbPSTOL1* and *SbMATE* as cofactors (GF-GBLUP-A and -D) except for Al tolerance, where accuracy was increased in GF-GBLUP-A. In general, when used as fixed cofactors, SNPs associated with the different traits via GWAS (GWAS-GBLUP) increased prediction accuracies, except for plant dry matter. The highest prediction accuracies of 0.28, 0.53 and 0.45 for grain yield, plant height and Al tolerance, respectively, resulted from the GBLUP model which included dominance effects and GWAS-SNPs as cofactors (GWAS-GBLUP-AD). The strongest impact of including GWAS-SNPs as cofactors was observed for plant height and appears to be closely related to the presence of underlying loci with dominance effects. Although there was no advantage in including dominance effects in the absence of GWAS-SNPs (GBLUP-A vs GBLUP-AD), dominance increased prediction accuracies from 0.40 to 0.53 in the presence of GWAS-derived cofactors (GWAS-GBLUP-A vs. AD). Strikingly, accuracies increased by 90% after inclusion of GWAS-SNP cofactors in the presence of dominance effects (GBLUP-AD vs. GWAS-GBLUP-AD).

Discussion

Different from populations such as some recombinant inbred lines and diverse association panels, the multiparental, partially-selfed random mating population (MP-RMP), BRP13R, is intrinsically a breeding resource. BRP13R has been designed to dynamically incorporate into a pre-breeding pipeline new sources of alleles for desirable agronomic traits and to allow for the identification of transgressive progeny accumulating favorable alleles at multiple loci, particularly those related to sorghum adaptation to tropical soils, where abiotic stresses are common.

One significant advantage of BRP13R emerges from its power to positionally clone abiotic stress tolerance genes whose favorable alleles are present in rather low frequencies and are specific to certain subgroups, which is an enormous challenge for GWAS approaches (Brachi et al. 2011). The Al tolerance gene, *SbMATE*, has been shown to increase grain yield by over one ton ha^{-1} when grown on an Al-toxic acid soil (Carvalho et al. 2016) and is the major determinant of Al tolerance in sorghum. Favorable alleles of *SbMATE* are rather rare and mostly specific mostly to guinea sorghums from their primary and secondary domestication centers, in West and South East Africa, respectively (Caniato et al. 2011). A GWAS approach targeting Al tolerance was performed in a diverse and highly structured association panel (Melo et al. 2019), with a model jointly including population structure and relatedness (Yu et al. 2006). Accordingly, many SNP loci distributed within most of the sorghum chromosomes showed association signals either similar to or even higher than the GBS-SNP that showed the highest association signal in the *SbMATE* region (Melo et al. 2019). Therefore, without previous knowledge, GWAS in the association panel used by Melo et al. (2019) would have been rather inefficient to directly positionally clone *SbMATE*, as many other candidate regions would have to be considered for gene discovery and further validation.

Roughly 265,000 SNPs have been considered adequate for GWAS approaches in sorghum, even in a highly diverse association panel, which was likely based on an average LD extent estimated to be under 10 Kb (Morris et al. 2013). Nevertheless, this assertion should be viewed with extreme caution, as LD coefficients typically show extremely high variance (Hedrick 1987) and LD fluctuations in the genome are also common due to heterogeneous recombination (Flint-Garcia et al. 2003). For example, in sorghum, de Alencar Figueiredo et al. (2008) reported on remarkably variable within-gene LD, encompassing whole genes (> 4 Kb) for *Opaque2*, intensive intragenic recombination happening within only 244 bp in *Waxy,* and very weak LD along *Brittle2*. Lack of salient associations of GBS-SNPs near *SbMATE* with Al tolerance is influenced by the need for population structure cofactors in the association model, the low frequency of Al tolerance and low LD in the *SbMATE* region, which was found to persist to up to around 500 bp, resulting in intragenic recombination and great haplotype diversity for the Al tolerance gene (Caniato et al. 2014; Hufnagel et al. 2018). Because rare alleles are not efficiently sampled by the skim sequencing of GBS using moderate population sizes, *SbMATE*-specific markers identified via a targeted associating mapping approach (Caniato et al. 2014), rather than any GBS SNP marker, showed by far the highest association signals for Al tolerance in the diverse association panel used by Melo et al. (2019). In contrast, using BRP13R, an extremely strong and prominent probability peak for association between GBS markers and Al tolerance was observed kilobases away *SbMATE* and, in fact, two GBS-SNPs overlapping with *SbMATE* SNPs (in red in Fig. 4) showed even stronger association signals compared to the *SbMATE*-specific markers.

Because r^2 reflects statistical power to detect LD (Balding 2006) and is inversely proportional to the sample size required for detection (Zondervan and Cardon 2004; Wang et al. 2005), we infer that a much larger population size would have been needed to positionally clone *SbMATE* only with GBS markers via GWAS in the diverse association panel used by

Melo et al. (2019). Alternatively, a GWAS approach with only 200 individuals, genotyped with ~44,000 SNPs in BRP13R, would have been enough to directly positionally clone *SbMATE* as previously discussed, which contrasts with the ~235,000 markers and 254 accessions used for GWAS by Melo et al. (2019). Hence, BRPR13R appears to offset some of the hurdles of other related multiparental designs such as MAGIC and NAM populations, some of which require long development time and large population sizes (Mackay et al. 2014; Bouchet et al. 2017) that can significantly constrain phenotyping, particularly in field trials. The reduced population size in BRP13R leading to these results is remarkable, as in the original positional cloning of *SbMATE*, a 354-member RIL population and over 2,085 F_2 individuals had to be screened (Magalhaes et al. 2007). This extraordinary advantage of BRP13R over other populations is likely associated with its multi-allelic nature, intermediate levels of LD compared to highly diverse association panels and biparental populations, highly reduced population structure (that may also reduce detection power (Kang et al. 2008)) via random mating and selection for the target trait, which were all but natural consequences of the breeders´ effort to identify transgressive segregants for hybrid development.

Our predictive ability for grain yield in the absence of dominance effects or fixed cofactors, in the range of 0.22, was slightly lower than what was reported in previous publications on this in sorghum (Velazco et al. 2019). However, the inclusion of dominance effects and GWAS-derived fixed cofactors raised grain yield accuracies to ~0.3. This indicates that the mutiparental, random mating nature and the residual heterozygosity in BRP13R do not preclude its effective use for genomic selection approaches, particularly if markers in LD with major genes with dominant effects are identified by GWAS and included in the GS model as fixed effects. The most dramatic increase in accuracy was achieved with a GBLUP model including dominance effects and GWAS cofactors for plant height, yielding maximum accuracy of 0.53, which is value similar to that in previous reports (Velazco et al.

2019). In general, modelling SNPs in LD with major genes as having fixed instead of random effects has been shown to improve accuracies for traits such as plant height and flowering time in rice (Spindel et al. 2016), rust resistance in wheat (Rutkoski et al. 2014) and carotenoid levels (Owens et al. 2014).

The fraction of the genetic variance jointly explained by the GWAS-derived cofactors in the GBLUP models applied to BRP13R varied from 0.13 for grain yield to 0.32 for plant height, which showed the highest heritability among all traits ($h^2 = 0.75$). Thus, our results in sorghum, particularly based on the substantial improvement in plant height prediction accuracy with GWAS-derived cofactors, agree with simulations by Bernardo et al. (Bernardo 2014) in maize. Accordingly, these authors concluded that adding fixed cofactors to the GS models is helpful, particularly for oligogenic traits and when each major gene explains more than 10% of the genetic variance.

Many important traits in plant breeding are quantitative in nature and are controlled by several genes, each with modest effects on the phenotype. Due to the large number of hypothesis to be tested and the consequent need to correct for multiple tests, detecting minor-effect loci is a substantial limitation of GWAS approaches, and population sizes in the range of thousands may be needed, even for alleles with a frequency of 0.15 (Hirschhorn and Daly 2005). From this perspective, we set out to explore if and how SNPs discovered by targeted approaches and explaining a much smaller portion of the genetic variance could be integrated into a pre-breeding pipeline including genomic selection under random mating.

Sorghum homologs of rice *Phosphorus-starvation tolerance1* (*OsPSTOL1)* (Gamuyao et al. 2012) were found to be associated with root morphology traits that have been shown to enhance root P acquisition and grain yield under low-P availability in the soil (Hufnagel et al. 2014), and co-localized with the respective QTLs in a sorghum RIL population (Bernardino et al. 2019). Six SNPs in total LD within the *SbPSTOL1* gene, *Sb03g006765*, in addition to one

SNP within *Sb03g031680*, which individually explained ~3 – 4 % of the genetic variance, were associated with increases in grain yield of about 154 – 200 kg ha^{-1} in a diverse sorghum association panel cultivated in a low-P soil, which likely results from increases in root surface area leading to enhanced P uptake (Hufnagel et al. 2014). Although the association probabilities for these SNPs were not nearly as close as our GWAS threshold in BRP13R, the same six alleles within *Sb03g006765* and the single allele in *Sb03g031680* that increased grain yield in the Hufnagel et al. (2014) study also resulted in higher grain yield average compared to the alternative alleles (Fig. S3), suggesting that the *SbPSTOL1* effect on grain yield in BRP13R is sound, albeit below significance. In support, we found that the allele frequency for all *SbPSTOL1* SNPs shifted upwards, favoring a higher frequency of favorable alleles (positive Δ_f) (Fig. S4) in the top 10% lines selected for grain yield compared to low-yielding BRP13R progeny. Furthermore, this shift was most dramatic for *Sb03g006765* (0.13) and *Sb03g031680* (0.10), which are exactly the two *SbPSTOL1* genes that had been previously associated with grain yield under low P availability in the soil via targeted association mapping (Hufnagel et al. 2014). In addition, on average, we observed enrichment of genotypes homozygous for favorable alleles over progeny homozygous for the unfavorable alleles in high-yielding progeny. This pattern, which was primarily driven by *Sb03g006765*, was reversed in low-yielding progeny, where progeny homozygous for the unfavorable alleles predominate. Collectively, our findings confirm that genes with more subtle effect on the target traits, consequently explaining smaller fractions of the phenotypic variance, such as *SbPSTOL1* genes, are in principle not useful for boosting accuracies in genomic selection approaches as previously predicted (Bernardo 2014). Notwithstanding, our genotype and allele frequency analyses indicate that such genes identified via targeted association mapping and other approaches, should enter into GS pipelines via allele mining and characterization of the founder germplasm and, if needed, marker-assisted introgression in the base population.

The allele shift for the Al tolerance gene, *SbMATE*, was negligible between high- and low-yielding progeny. This can be explained by the fact that Al saturation in the soil surface was below toxicity levels for sorghum, as we wanted our low-P site to isolate the effect of low-P availability from Al toxicity on BRP13R performance. A previous study with *SbMATE*-specific markers suggested that citrate release mediated by the root plasma membrane SbMATE protein may also benefit P uptake and grain yield under low-P availability in West Africa (Leiser et al. 2014). However, P stress on tropical soils occurs via P fixation on the surfaces of Al and iron oxides in the soil clays, impairing the diffusive flux of P from the soil towards the root surface (reviewed by Magalhaes et al. (2018)). Since BRP13R was assessed for grain yield in a clay low-P soil under irrigation in Brazil, it is possible that a stronger P stress occurred in the West Africa trials, which might have potentiated the effect of citrate release mediated by *SbMATE* on P uptake. In addition, the *SbMATE* effect on grain yield in the Leiser et al. (2014) study may result from a combined effect of *SbMATE* enhancing both P acquisition and Al tolerance, due to the possibly higher Al^{3+} activity in the West African sandy soils where grain yield was assessed (Leiser et al. 2014; Magalhaes et al. 2018). Nevertheless, in by far the most widespread situation where Al toxicity and low P availability co-exist on acidic soils, we anticipate that progeny combining favorable alleles of both *SbMATE* and *SbSPTOL1* genes will be more adapted and hence show enhanced yield stability on tropical soils.

We have found reassuring evidence for QTL conservation between BRP13R and a large RIL population (Bernardino et al. 2019), which were both phenotyped for grain yield under low-P availability and root morphology traits (Table S6). For example, we have previously reported on the presence of a sorghum homolog of the wheat Al tolerance gene, *ALMT*, within a grain yield QTL on chromosome 6, and of a *PHOSPHATE2* (*PHO2*) homolog co-localized with a grain yield QTL on chromosome 9 (Bernardino et al. 2019). The

aluminum-activated malate transporter, ALMT1, has been recently shown to influence root growth in low-P conditions in Arabidopsis (Mora-Macías et al. 2017), whilst *PHO2* has been implicated in maize P efficiency (Du et al. 2018). In BRP13R, the QTL with the strongest association signal for grain yield was located on chromosome 6 at 37 – 45 Mb. This region overlaps at 40 – 45 Mb with the grain yield QTL detected both by single- and multi-trait mapping in the RIL population, and the overlapping region includes the sorghum homolog of *ALMT* at position ~44 Mb. At position ~57 Mb, *PHO2* is near a grain yield QTL in the end region of chromosome 9 and is within overlapping QTL for P content, grain, root and plant dry matter, in addition to plant height in BRP13R.

The cloning of genes important for crop breeding, particularly the challenging ones with rather minor effects, will benefit from integrative approaches that explore complementarities between different types of populations, such as recombinant inbred lines, diverse association panels and multiparental populations. We can anticipate that such integrative resources may balance advantages and drawbacks of each type of population taken alone, arising from historical aspects influencing genetic structure, demography and diversity, which ultimately translate into variable levels of linkage disequilibrium (Nordborg and Tavaré 2002). Our study with BRP13R indicates that this type of random mating population - where many founders and derived progeny were intensively recombined - emerges as a multipurpose resource useful both for genomics and breeding applications. Such highly recombined multiparental populations increase the chances of cloning important genes by GWAS, serving as a vehicle for bridging gene discovery and cultivar development via deployment of gene-specific markers into pre-breeding efforts. Finally, boosted by genomic selection, this approach benefits cultivar development via selection of progeny transgressively accumulating favorable alleles at many loci that are important for a broader adaptation to acidic soils, such as those conferring Al tolerance and P efficiency.

Conclusions

MP-RMPs such BRP13R emerge as multipurpose resources for efficient gene discovery and deployment, benefiting global food security via sorghum cultivars with broad adaptation to tropical soils.

List of abbreviations

Δ_f: Difference of favorable allele frequencies between the top and bottom BRP13R progeny; AIC: Akaike information criterion; Al: Aluminum; *ALMT*: Aluminum tolerance locus in *Sorghum bicolor*; $Alt_{SB:}$ Aluminum tolerance locus in *Sorghum bicolor*; ATF13: Aluminum tolerant female 13; ATF14: Aluminum tolerant female 14; BIC: Bayesian Information Criterion; BLUEs: Best linear unbiased estimators; BRP3R: Brazilian random mating population 3 restore; BWA: Burrows-wheeler aligner program; Fe: Iron; FRL_{5d}: Final root length after five days; FT: Flowering time; GBLUP: Genomic best linear unbiased prediction; GBLUP-A: Genomic best linear unbiased prediction with random additive matrix; GBLUP-AD: Genomic best linear unbiased prediction with random additive matrix and random dominance matrix; GBS: Genotyping-by-sequencing; GDM: Grain dry matter; GF-GBLUP-A: Genomic best linear unbiased prediction with random additive matrix and *SbPSTOL1* and *SbMATE* markers as fixed-effect covariates; GF-GBLUP-AD: Genomic best linear unbiased prediction with random additive matrix and random dominance matrix, and *SbPSTOL1* and *SbMATE* genes as fixed-effect covariates; GS: Genomic selection; GWAS: Genome-wide association mapping; GWAS-GBLUP-A: Genomic best linear unbiased prediction with random additive matrix and markers selected via GWAS as fixed-effect covariates; GWAS-GBLUP-AD: Genomic best linear unbiased prediction with random additive matrix and random dominance matrix, and including GWAS markers as fixed-effect covariates; Gy: Grain yield; IRL: Initial root length; K: Kinship matrix; KASP: Allele Specific PCR genotyping system; LD: Linkage disequilibrium; LRT: Likelihood ratio test; MAF: Minor allele frequency; MAGIC: Multiparent Advanced Generation Intercross; ms_3: Male sterile plants; NAM: Nested association mapping; NRG: Net root growth; NRP3R: Nebraska Random Mating Population 3; *OsPSTOL1:phosphorus starvation tolerance1 in rice*; P: Phosphorus; PC1:

Principal component 1; PDM: Plant dry matter; Pg: Grain phosphorus content; PH: Plant height; *pho2: Phosphate 2*; Pp: Plant phosphorus content; PP3R: Purdue Population 3R; Pr: Root phosphorus content; Ps: Shoot phosphorus content; Pt: Total phosphorus content; Q: Population structure; q-q: Quantile-quantile; QTL: Quantitative trait locus; r^2: Correlation square; RD: Root diameter; RD: Root diameter; RDM: Root dry matter; REML: Restricted maximum likelihood; RILs: Recombinant inbred lines; RL: Root length; RMP: Multiparental random matting population; RNRG: Relative net root growth; RV: Root volume; SA: Total root surface area; SA1: Surface area of super fine roots between 0 - 1 mm in diameter; SA2: Surface area of fine roots between 1 - 2 mm in diameter; SA3: Surface area of thicker roots between 2 - 4.5 mm in diameter; SbMATE: Multidrug and toxic compound extrusion in *Sorghum bicolor*; SbPSTOL1: *Phosphorus starvation tolerance1* in *Sorghum bicolor;* SDM: Shoot dry matter; SNP: Single nucleotide polymorphism.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. 267–281

Amadeu RR, Cellon C, Olmstead JW, et al (2016) AGHmatrix: R package to construct relationship matrices for autotetraploid and diploid species: a blueberry example. Plant Genome 9:

Balding DJ (2006) A tutorial on statistical methods for population association studies. Nat Rev Genet 7:781

Bernardino KC, Pastina MM, Menezes CB, et al (2019) The genetic architecture of phosphorus efficiency in sorghum involves pleiotropic QTL for root morphology and grain yield under low phosphorus availability in the soil. BMC Plant Biol 19:87

Bernardo R (2014) Genomewide selection when major genes are known. Crop Sci 54:68–75

Bland JM, Altman DG (1995) Multiple significance tests: the Bonferroni method. Bmj 310:170

Bouchet S, Olatoye MO, Marla SR, et al (2017) Increased power to dissect adaptive traits in global sorghum diversity using a nested association mapping population. Genetics 206:573–585

Bouchet S, Pot D, Deu M, et al (2012) Genetic structure, linkage disequilibrium and signature of selection in sorghum: lessons from physically anchored DArT markers. PLoS One 7:e33470

Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants the missing heritability is in the field. Genome Biol 12:232

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635 . doi: 10.1093/bioinformatics/btm308

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165–1177

Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81:1084–1097

Butler DG, Cullis BR, Gilmour AR, Gogel BJ (2009) ASReml-R reference manual. State Queensland, Dep Prim Ind Fish Brisbane

Caniato FF, Guimarães CT, Hamblin M, et al (2011) The relationship between population structure and aluminum tolerance in cultivated sorghum. PLoS One 6:e20830

Caniato FF, Guimaraes CT, Schaffert RE, et al (2007) Genetic diversity for aluminum tolerance in sorghum. Theor Appl Genet 114:863–876

Caniato FF, Hamblin MT, Guimaraes CT, et al (2014) Association mapping provides insights into the origin and the fine structure of the sorghum aluminum tolerance locus, AltSB. PLoS One 9:e87438

Carvalho G, Schaffert RE, Malosetti M, et al (2016) Back to acid soil fields: The citrate transporter SbMATE is a major asset for sustainable grain yield for sorghum cultivated on acid soils. G3 Genes, Genomes, Genet 6:475–484 . doi: 10.1534/g3.115.025791

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. J Agric Biol Environ Stat 11:381

De Alencar Figueiredo LF, Calatayud C, Dupuits C, et al (2008) Phylogeographic evidence of crop neodiversity in sorghum. Genetics 179:997–1008

de Sousa SM, Clark RT, Mendes FF, et al (2012) A role for root morphology and related candidate genes in P acquisition efficiency in maize. Funct Plant Biol 39:925–935

Du Q, Wang K, Zou C, et al (2018) The PILNCR1-miR399 Regulatory Module Is Important for Low Phosphate Tolerance in Maize. Plant Physiol 177:1743–1753

Elshire RJ, Glaubitz JC, Sun Q, et al (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379

Endelman JB, Jannink J-L (2012) Shrinkage estimation of the realized relationship matrix. G3 Genes, Genomes, Genet 2:1405–1413

FAO I (2015) Status of the world's soil resources (SWSR)--main report. Food Agric Organ United Nations Intergov Tech panel soils, Rome, Italy 650:

Flint-Garcia SA, Thornsberry JM, Buckler IV ES (2003) Structure of linkage disequilibrium in plants. Annu Rev Plant Biol 54:357–374

Foy CD, Duncan RR, Waskom RM, Miller DR (1993) Tolerance of sorghum genotypes to an acid, aluminum toxic Tatum subsoil. J Plant Nutr 16:97–127 . doi: 10.1080/01904169309364517

Gamuyao R, Chin JH, Pariasca-Tanaka J, et al (2012) The protein kinase Pstol1 from traditional rice confers tolerance of phosphorus deficiency. Nature 488:535

Glaubitz JC, Casstevens TM, Lu F, et al (2014) TASSEL-GBS: A high capacity genotyping

by sequencing analysis pipeline. PLoS One 9:e90346 . doi: 10.1371/journal.pone.0090346

Harrell Jr FE (2015) With contributions from Charles Dupont and many others. Hmisc: Harrell Miscellaneous. R package version 3.17-4

Harris K, Subudhi PK, Borrell A, et al (2006) Sorghum stay-green QTL individually reduce post-flowering drought-induced leaf senescence. J Exp Bot 58:327–338

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95

Hufnagel B, de Sousa SM, Assis L, et al (2014) Duplicate and conquer: multiple homologs of PHOSPHORUS-STARVATION TOLERANCE1 enhance phosphorus acquisition and sorghum performance on low-phosphorus soils. Plant Physiol 166:659–677

Hufnagel B, Guimaraes CT, Craft EJ, et al (2018) Exploiting sorghum genetic diversity for enhanced aluminum tolerance: Allele mining based on the Alt SB locus. Sci Rep 8:10094

Hund A, Trachsel S, Stamp P (2009) Growth of axile and lateral roots of maize: I development of a phenotying platform. Plant Soil 325:335–349 . doi: 10.1007/s11104-009-9984-2

Kang HM, Zaitlen NA, Wade CM, et al (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kochian L V (1995) Cellular mechanisms of aluminum toxicity and resistance in plants. Annu Rev Plant Biol 46:237–260 . doi: 10.1146/annurev.pp.46.060195.001321

Leiser WL, Rattunde HFW, Weltzien E, et al (2014) Two in one sweep: aluminum tolerance and grain yield in P-limited soils are associated to the same genomic region in West

African sorghum. BMC Plant Biol 14:206 . doi: 10.1186/s12870-014-0206-6

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows--Wheeler transform. bioinformatics 25:1754–1760

Lynch JP (2011) Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. Plant Physiol 156:1041–1049 . doi: 10.1104/pp.111.175414

Mace E, Innes D, Hunt C, et al (2019) The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. Theor Appl Genet 132:751–766

Mace ES, Hunt CH, Jordan DR (2013) Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theor Appl Genet 126:1377–1395

Mace ES, Jordan DR (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. Theor Appl Genet 123:169

Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends Plant Sci 12:57–63

Mackay IJ, Bansept-Basler P, Barber T, et al (2014) An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: creation, properties, and validation. G3 Genes, Genomes, Genet 4:1603–1610

Magalhaes J, Piñeros MA, Maciel L, Kochian L (2018) Emerging pleiotropic mechanisms underlying aluminum resistance and phosphorus acquisition on acidic soils. Front Plant Sci 9:1420

Magalhaes J V, Garvin DF, Wang Y, et al (2004) Comparative mapping of a major aluminum tolerance gene in sorghum and other species in the Poaceae. Genetics 167:1905–1914

Magalhaes J V, Liu J, Guimarães CT, et al (2007) A gene in the multidrug and toxic

compound extrusion (MATE) family confers aluminum tolerance in sorghum. Nat Genet 39:1156–1161 . doi: 10.1038/ng2074

Magnavaca R, Gardner CO, Clark RB (1987) Inheritance of aluminum tolerance in maize. In: Genetic aspects of plant mineral nutrition. Springer, pp 201–212

Marschner H (1995) Adaptation of plants to adverse chemical soil conditions. Miner Nutr High plants

Melo JO, Martins LGC, Barros BA, et al (2019) Repeat variants for the SbMATE transporter protect sorghum roots from aluminum toxicity by transcriptional interplay in cis and trans. Proc Natl Acad Sci 116:313–318 . doi: 10.1073/PNAS.1808400115

Mora-Macías J, Ojeda-Rivera JO, Gutiérrez-Alanís D, et al (2017) Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate. Proc Natl Acad Sci 114:E3563--E3572

Morris GP, Ramu P, Deshpande SP, et al (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proc Natl Acad Sci 110:453–458

Neyman J, Pearson ES (1928) On the use and interpretation of certain test criteria for purposes of statistical inference: Part II. Biometrika 20A:263–294 . doi: 10.2307/2331945

Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. TRENDS Genet 18:83–90

Nothnagel M, Ellinghaus D, Schreiber S, et al (2009) A comprehensive evaluation of SNP genotype imputation. Hum Genet 125:163–171

Owens BF, Lipka AE, Magallanes-Lundback M, et al (2014) A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. Genetics 198:1699–1716

Parra-Londono S, Kavka M, Samans B, et al (2018) Sorghum root-system classification in

contrasting P environments reveals three main rooting types and root-architecture-related marker--trait associations. Ann Bot 121:267–280

Pearson K (1895) Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. Philos Trans R Soc London Ser A, Contain Pap a Math or Phys Character 187:253–318 . doi: 10.1098/rsta.1896.0007

R Core Team (2016) R: A Language and Environment for Statistical Computing

Robinson PS (2006) EP1726664B1 - Detection system for PCR assay

Ruiz HA, Fernandes B, Novais RF, Alvaes V (1988) VH Efeito da umidade do solo sobre o volume e o conteúdo de fósforo no exsudato xilemático de soja. R. Bras Ci Solo 12:39–42

Rutkoski JE, Poland JA, Singh RP, et al (2014) Genomic selection for quantitative adult plant stem rust resistance in wheat. Plant Genome 7:

Sabadin PK, Malosetti M, Boer MP, et al (2012) Studying the genetic basis of drought tolerance in sorghum by managed stress trials and adjustments for phenological and plant height differences. Theor Appl Genet 124:1389–1402

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RWL (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81:8014–8018 . doi: 10.1073/pnas.81.24.8014

Schwarz G, others (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Shaw JN (2001) Iron and aluminum oxide characterization for highly-weathered Alabama ultisols. Commun Soil Sci Plant Anal 32:49–64 . doi: 10.1081/CSS-100102992

Sokal RR, Sneath PHA (1963) Principles of Numerical Taxonomy. W.H. Freeman, San Francisco

Spindel JE, Begum H, Akdemir D, et al (2016) Genome-wide prediction models that

incorporate de novo GWAS are a powerful new tool for tropical rice improvement. Heredity (Edinb) 116:395

Stacklies W, Redestig H (2016) The pcaMethods Package. Diambil dari< https//www. bioconductor. org/packages/3.3/bioc/manuals/pcaMethods/man/pcaMethods. pdf

Stadlmeier M, Hartl L, Mohler V (2018) Usefulness of a multiparent advanced generation intercross population with a greatly reduced mating design for genetic studies in winter wheat. Front Plant Sci 9:1825

Swarts K, Li H, Romero Navarro JA, et al (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Genome 7.

Tully K, Sullivan C, Weil R, Sanchez P (2015) The state of soil degradation in Sub-Saharan Africa: Baselines, trajectories, and solutions. Sustainability 7:6523–6552

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Velazco JG, Malosetti M, Hunt CH, et al (2019) Combining pedigree and genomic information to improve prediction quality: an example in sorghum. Theor Appl Genet 1–13

Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195:1223–1230

Von Uexküll HR, Mutert E (1995) Global extent, development and economic impact of acid soils. Plant Soil 171:1–15

Wald A (1943) Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans Am Math Soc 54:426–482

Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109

Webster OJ (1965) Genetic Studies in Sorghum vulgare (Pers.) 1. Crop Sci 5:207–210

Weir BS (2008) Linkage disequilibrium and association mapping. Annu Rev Genomics Hum Genet 9:129–142

Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. Genetics 178:539–551

Yu J, Pressoir G, Briggs WH, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203

Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. Nat Rev Genet 5:89

Figures

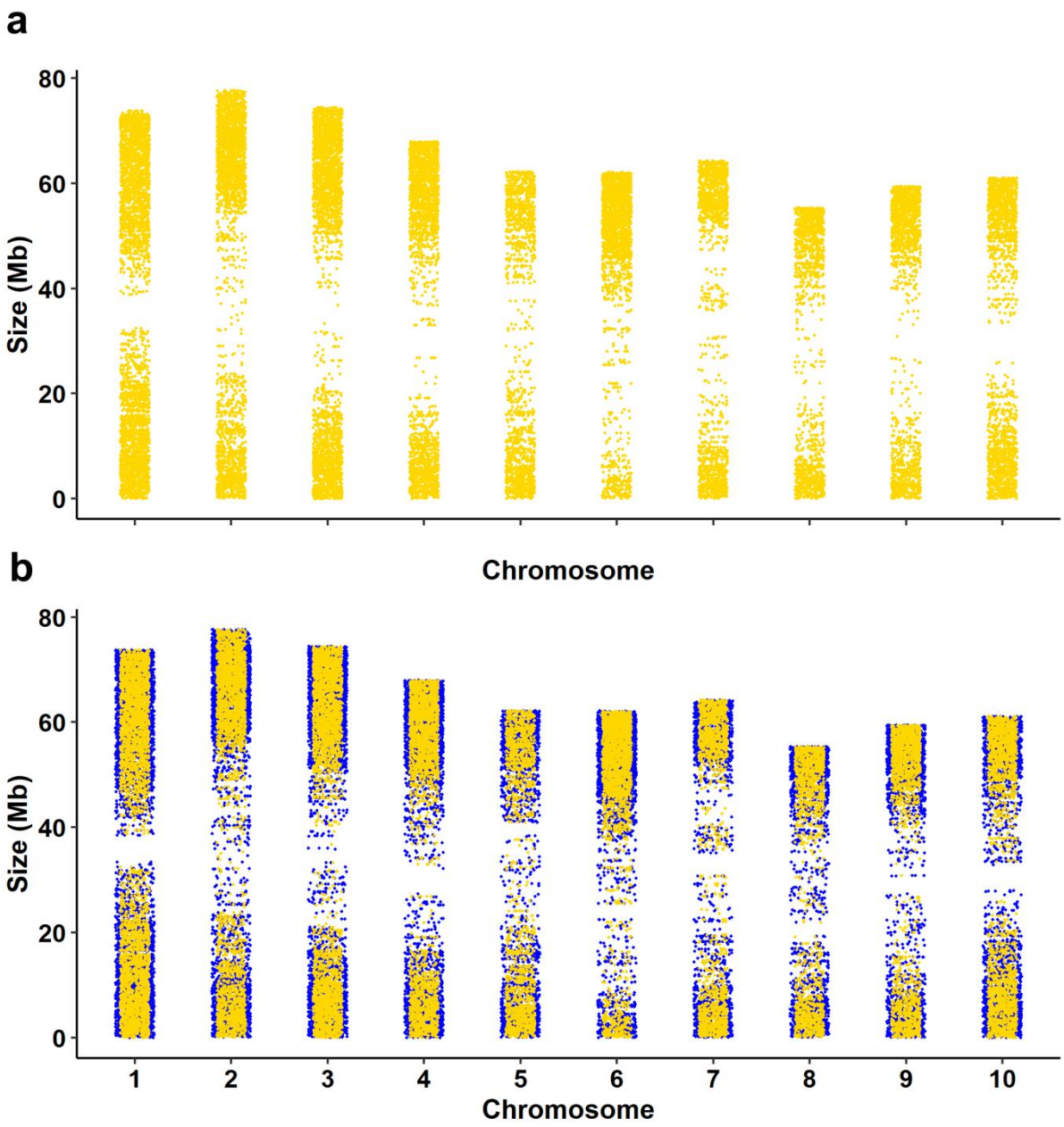


Fig. 1 Chromosome distribution of SNP loci before and after imputation. **(a)** Unimputed data. A maximum of 20% missing data per site and read depth ≥ 6 were allowed and the dataset contained 20,506 SNPs. **(b)** Imputed dataset with Beagle (Browning and Browning 2007). A maximum of 50% missing data per site was allowed and the window size was 500kb. The imputed data set contained 43,827 SNPs. Blue points in panel **(b)** depict imputed markers. Both panels show the distribution of biallelic, polymorphic loci, without insertions and deletions, with a read depth ≥ 6 and MAF ≥ 0.01. Mb: megabase pairs

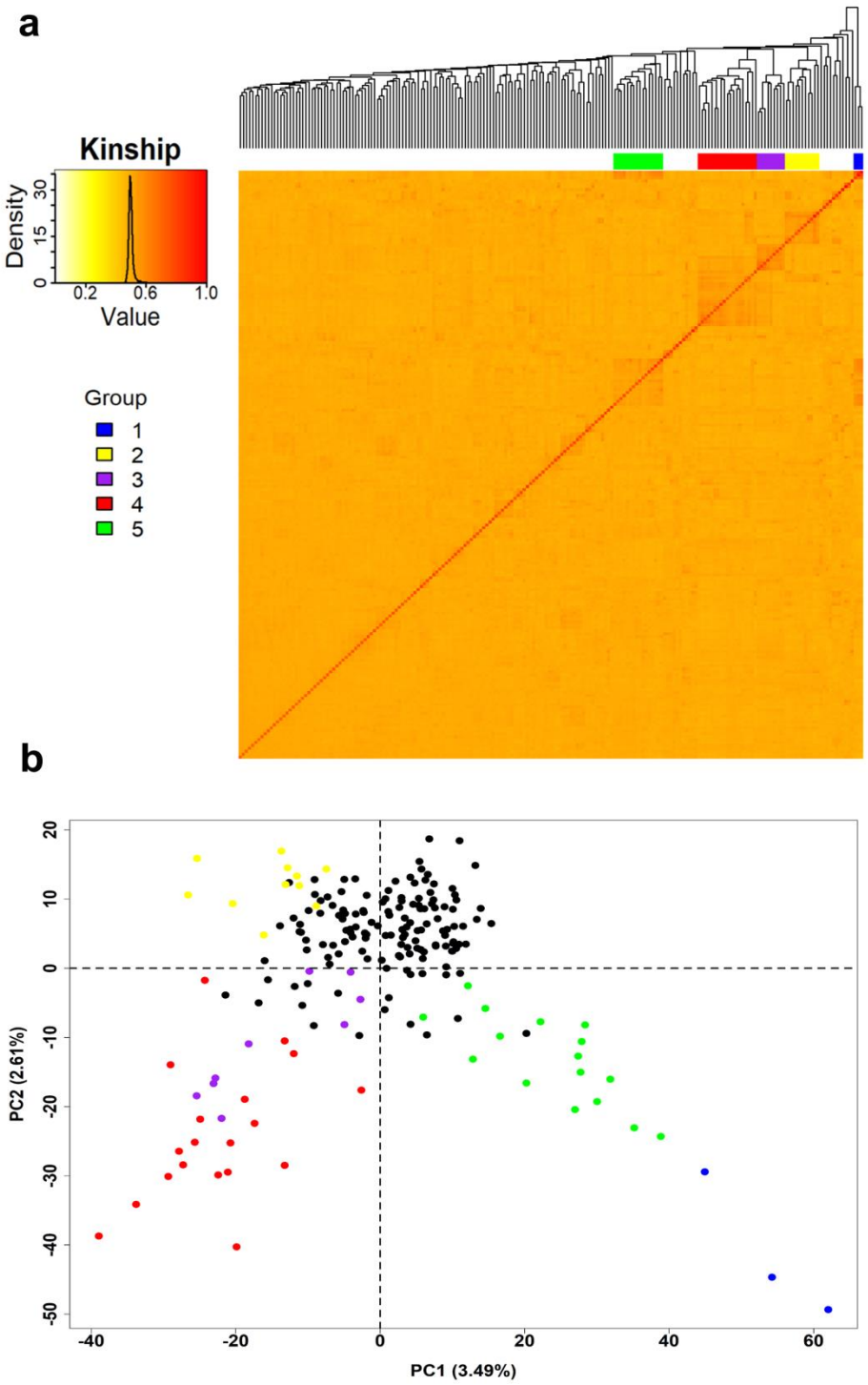


Fig. 2 Genetic relationship and population structure in 200 BRP13R progeny estimated with 43,827 SNP markers. **(a)** The kinship matrix was calculated with TASSEL (Bradbury et al. 2007) using an identity by state (IBS, Endelman and Jannink 2012) method and displayed as a heatmap and the frequency distribution of genetic relationship values are depicted (left). The unweighted pair group method with arithmetic mean (UPGMA) clustering of BRP13R progeny based on Euclidian distances is shown above the kinship heatmap. A colored scale was used to depict five differentiated groups. **(b)** Graphical display of progeny scores obtained by principal component analyses (PCA). Progeny belonging to the five groups identified in **(a)** were depicted by the same colors. The percentages of variance explained by the two PCs are shown in the axis titles.

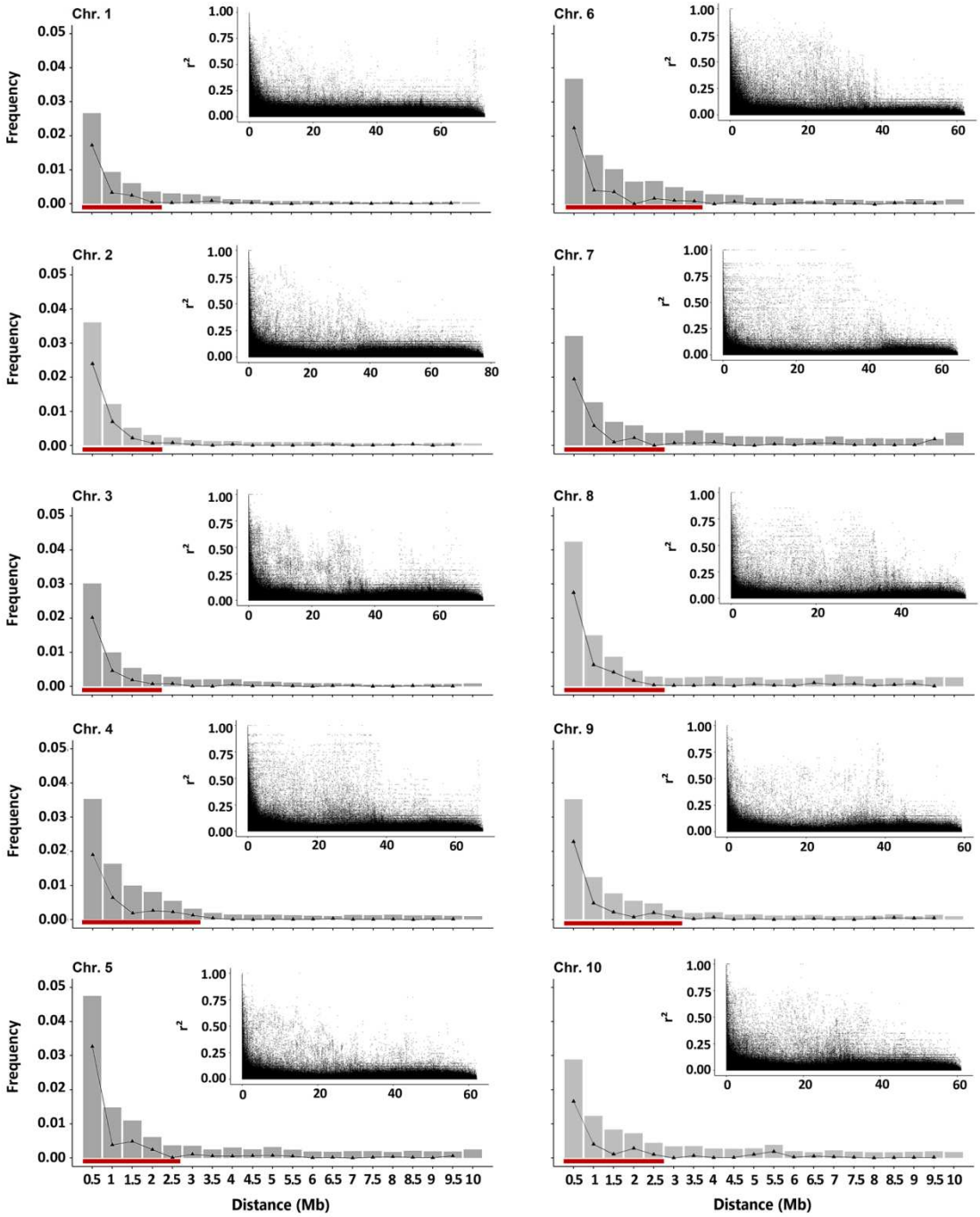


Fig. 3 Linkage disequilibrium decay per sorghum chromosome. The larger graphs show the frequency of loci under significant LD, which was measured as squared genotypic correlations (r^2) between pairs of SNP loci (Weir 2008) . A thin line was used to connect the difference between the proportion of significant LD measures of the current and previous chromosome physical interval (in Mb). The Fisher exact test was used to assess significance followed by a Bonferroni ($\alpha = 0.05$) multiple test correction per-chromosome. The thick red line above the x-axis indicates the LD extent, defined as the physical distance where the average r^2 values reached constant, basal levels. The inset graphs show the r^2 values between SNPs per chromosome (43,827 markers total). Chr: chromosome; Mb: megabase pairs.

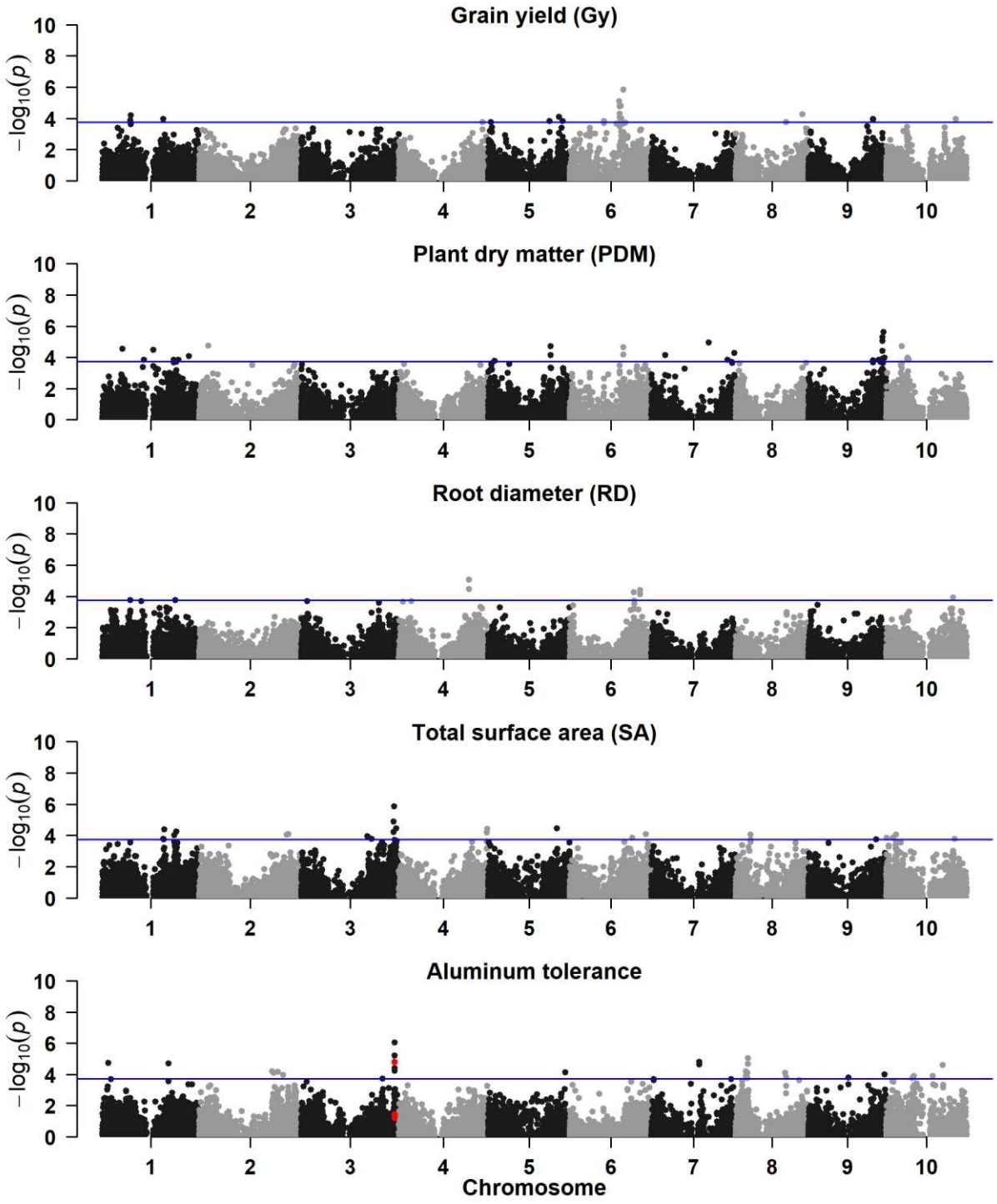


Fig. 4 GWAS profiles for grain yield (Gy, ton ha^{-1}), plant dry matter (PDM, ton ha^{-1}), root morphology traits and Al tolerance. The root morphology traits, root diameter (RD, in mm) and total surface area (SA, in cm²), were assessed after 13 days in nutrient solution with low-P. Al tolerance was measured by relative net root growth after five days of +/-Al exposure in nutrient solution with an Al^{3+} activity of {27} µM at pH 4.0. Colored in red are SNPs within the Alt_{SB} locus where *SbMATE* is located and within *SbMATE* itself (Caniato et al. 2014). The negative log of *p*-values ($-\log_{10}(p)$) were obtained with a GWAS model including principal component 1 (PC1, **Fig. 2b**). The horizontal line in blue depicts the significance threshold based on the Bonferroni correction for multiple, independent tests (*alpha* = 0.05), which were defined based on the extent of LD for each sorghum chromosome.

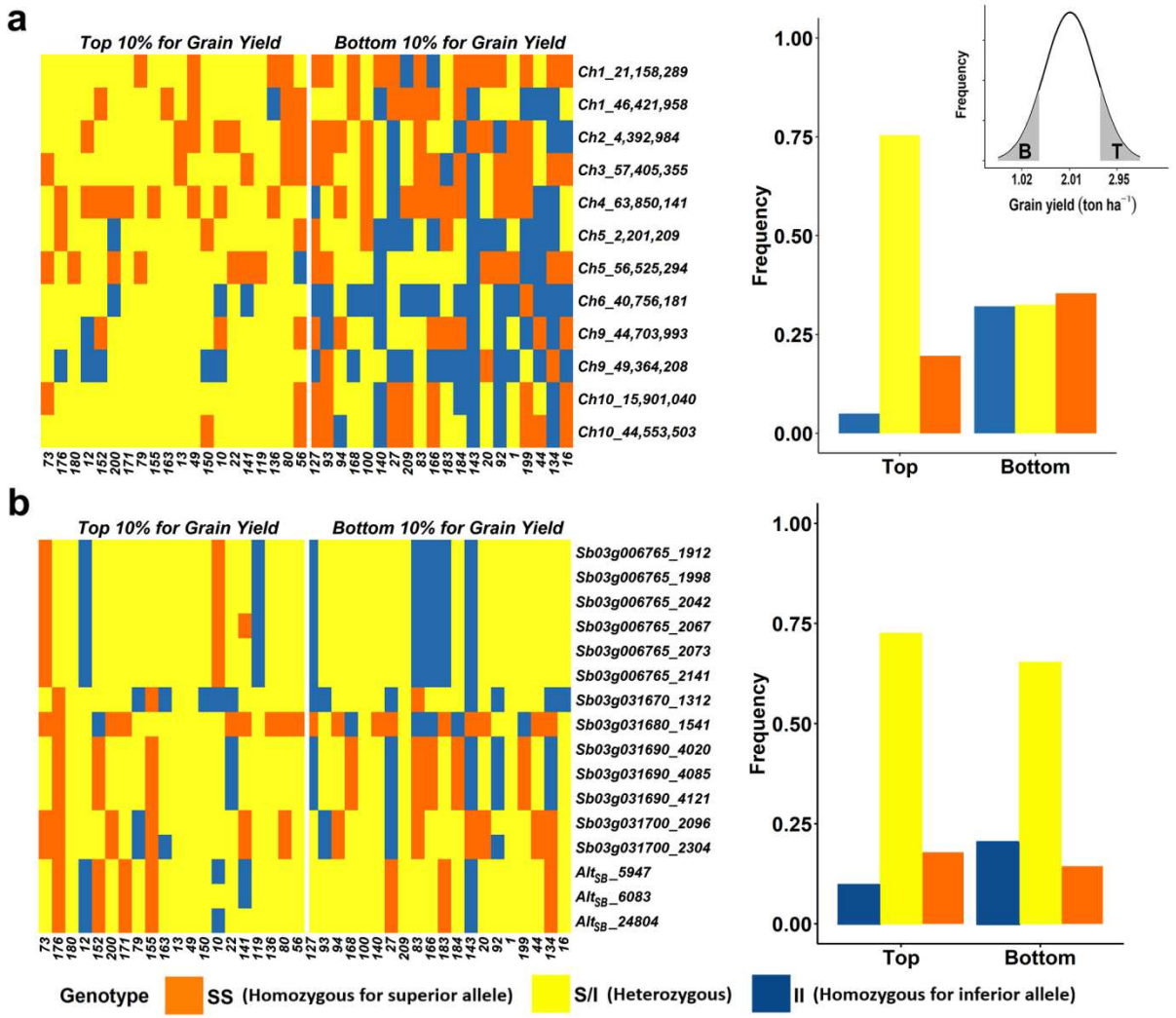


Fig. 5 Genotypic makeup of BRP13R progeny selected based on superior and inferior grain yield assessed on a low-P soil. A 10% selection threshold was imposed to select 20 progeny each with the highest (top 10%) and lowest (bottom 10%) grain yields on a low-P soil. The adjusted grain yield means for BRP13R and for the progeny in the top (T) and bottom (B) groups are shown in the histogram cartoon (top right). For each BRP13R progeny, homozygous genotypes for superior (SS) and inferior (II) alleles and heterozygotes (S/I) are depicted for: **(a)** SNP loci significantly associated with grain yield (**Fig. 4**) and **(b)** SNPs previously associated with Al tolerance (Caniato et al. 2014) and grain-yield under low-P (Hufnagel et al. 2014) within the *Alt$_{SB}$* locus or *SbPSTOL1* genes (*Sb03g006765*, *Sb03g031670*, *Sb03g031680* and *Sb03g031690*), respectively. Genotypic frequencies for the "top" and "bottom" progeny are shown in the respective bar charts.

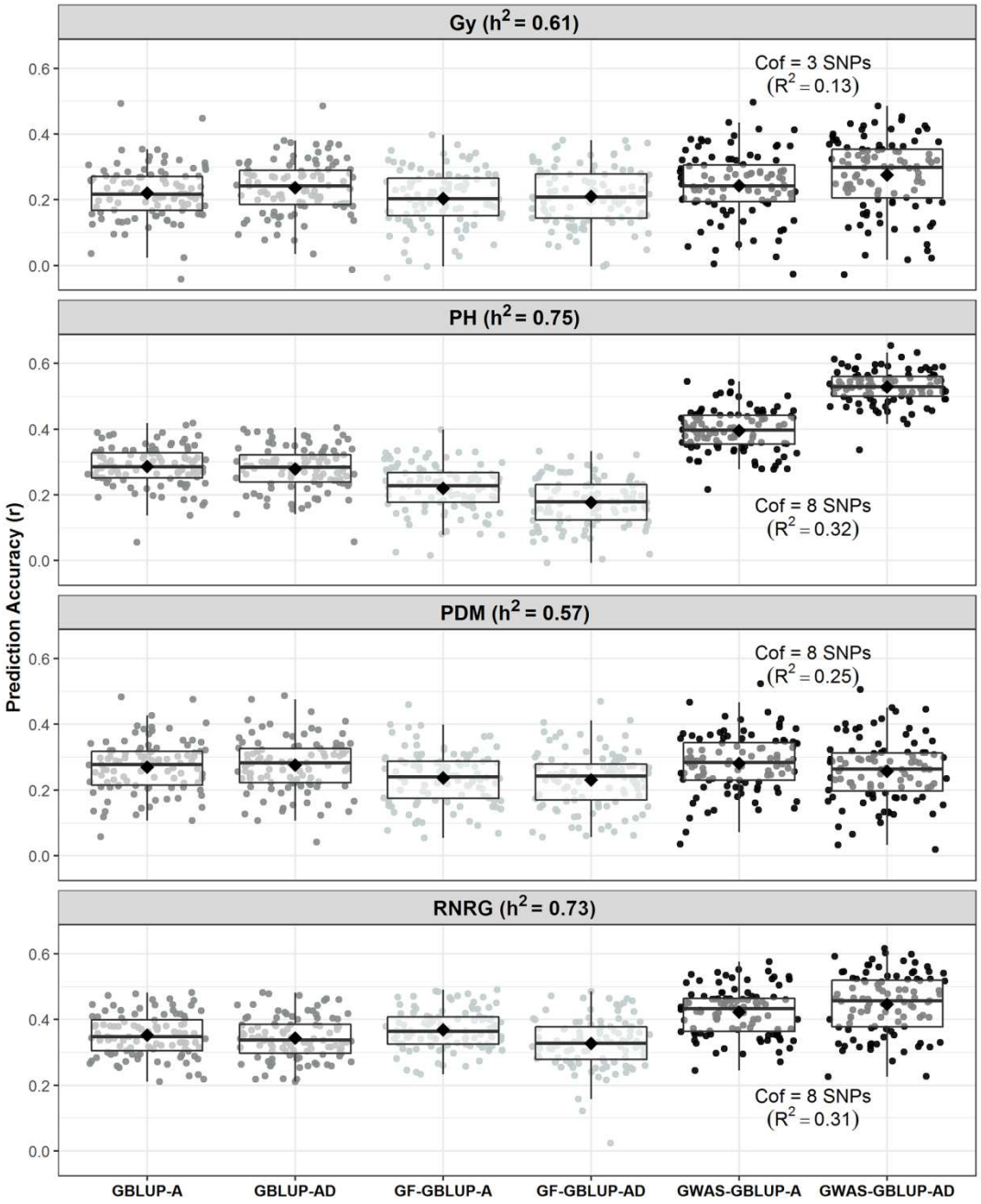


Fig. 6 Prediction accuracy (r) for genomic selection for grain yield (Gy), plant height (PH), plant dry matter (PDM) and Al tolerance (RNRG). Heritability (h^2) coefficients are shown for each trait. GBLUP-A is a GBLUP model with an additive genomic relationship matrix. The inclusion of a dominance genomic relationship matrix to GBLUP-A gives rise to GBLUP-AD. Gene-specific markers for *SbPSTOL1* and *SbMATE* (GF) and SNPs associated with the different traits by GWAS (GWAS-SNPs) were included as fixed cofactors both in the presence of an additive genomic relationship matrix (GF-GBLUP-A and GWAS-GBLUP-A) or including a dominance genomic relationship matrix (GF-GBLUP-AD and GWAS-GBLUP-AD) associated to the progeny random effect. Prediction accuracies were calculated as the correlation between the grain yield adjusted means via BLUE and the predicted means using GBLUP models. Cof corresponds to the number of cofactors, that is, the number of GWAS-SNP markers fitted as fixed effects whereas R^2 is the coefficient of determination of the full GWAS model including all the selected fixed effect cofactors for each traits.

Supplemental Figures

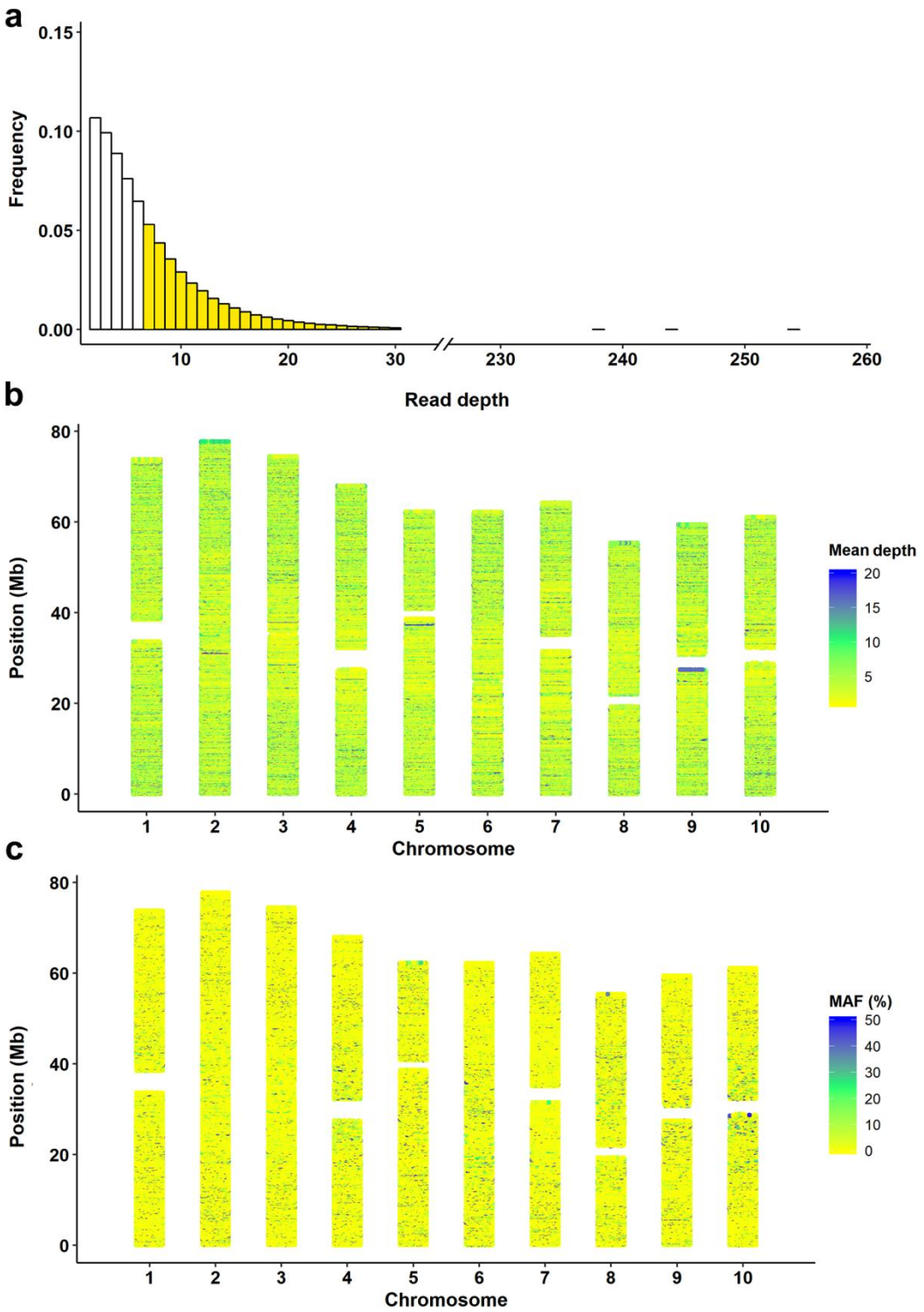


Fig. S1 Genotyping-by-sequencing (GBS) for the multiparental population BRP13R. **(a)** Frequency distribution of number of reads per genotype (read-depth). Sites with a number of reads ≥ 6 (yellow) were retained. **(b)** Chromosome distribution of SNP loci with average sequencing-depth between one and twenty, and **(c)** chromosome distribution of minor allele frequencies (MAF) for those sites. Mb: megabase pairs.

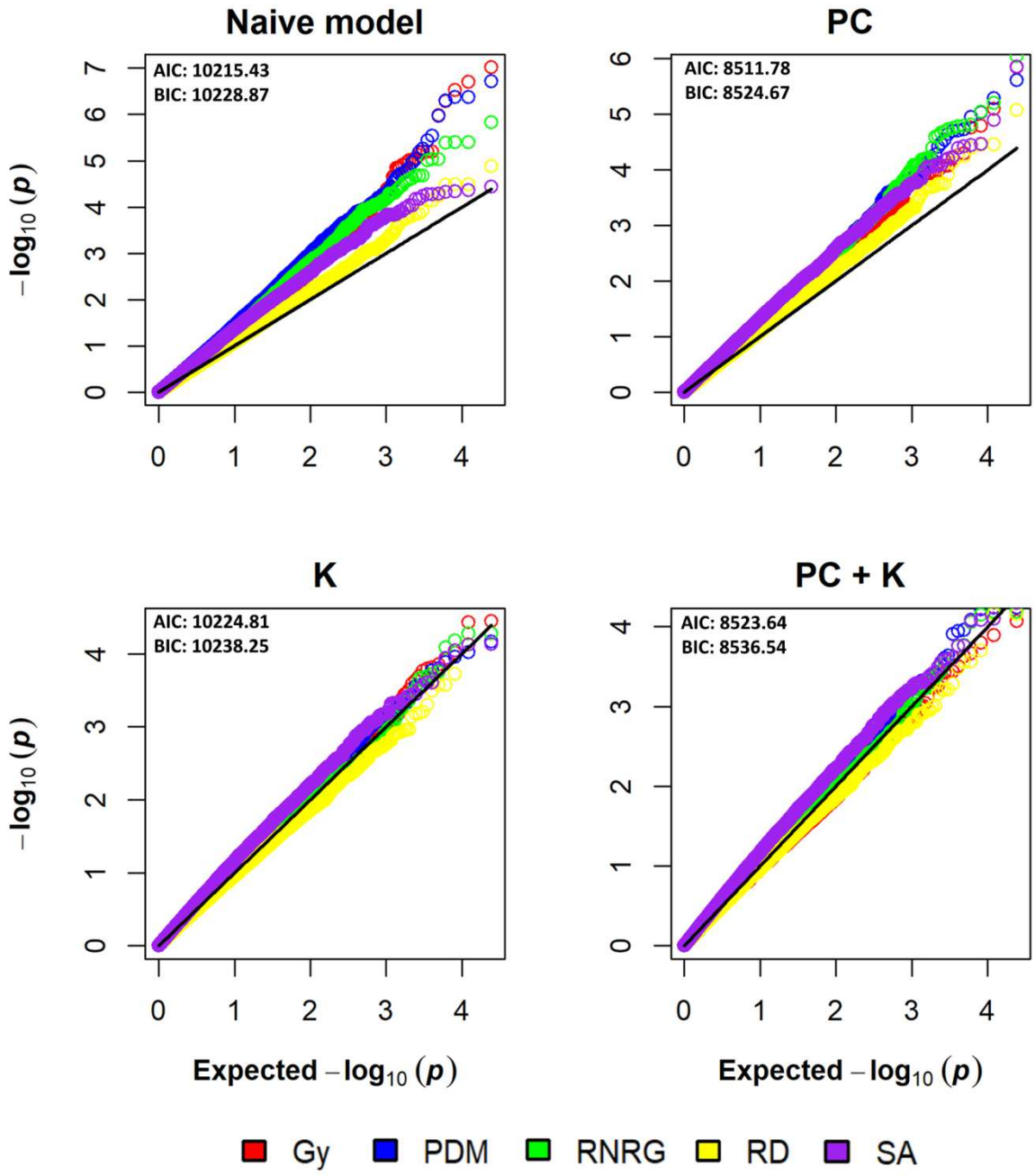


Fig. S2 Quantile-Quantile (Q-Q plots) for different GWAS models. The models are naïve (without population structure or relatedness), PC (with principal component 1 scores as the covariate), K (Kinship, incorporating the relationship matrix), and PC + K. The following traits were used: grain yield (Gy), plant dry matter (PDM); Al tolerance (relative net root growth, RNRG), root diameter (RD), and total surface area (SA). Akaike information criterion (AIC) and Bayesian information criterion (BIC) values are shown.

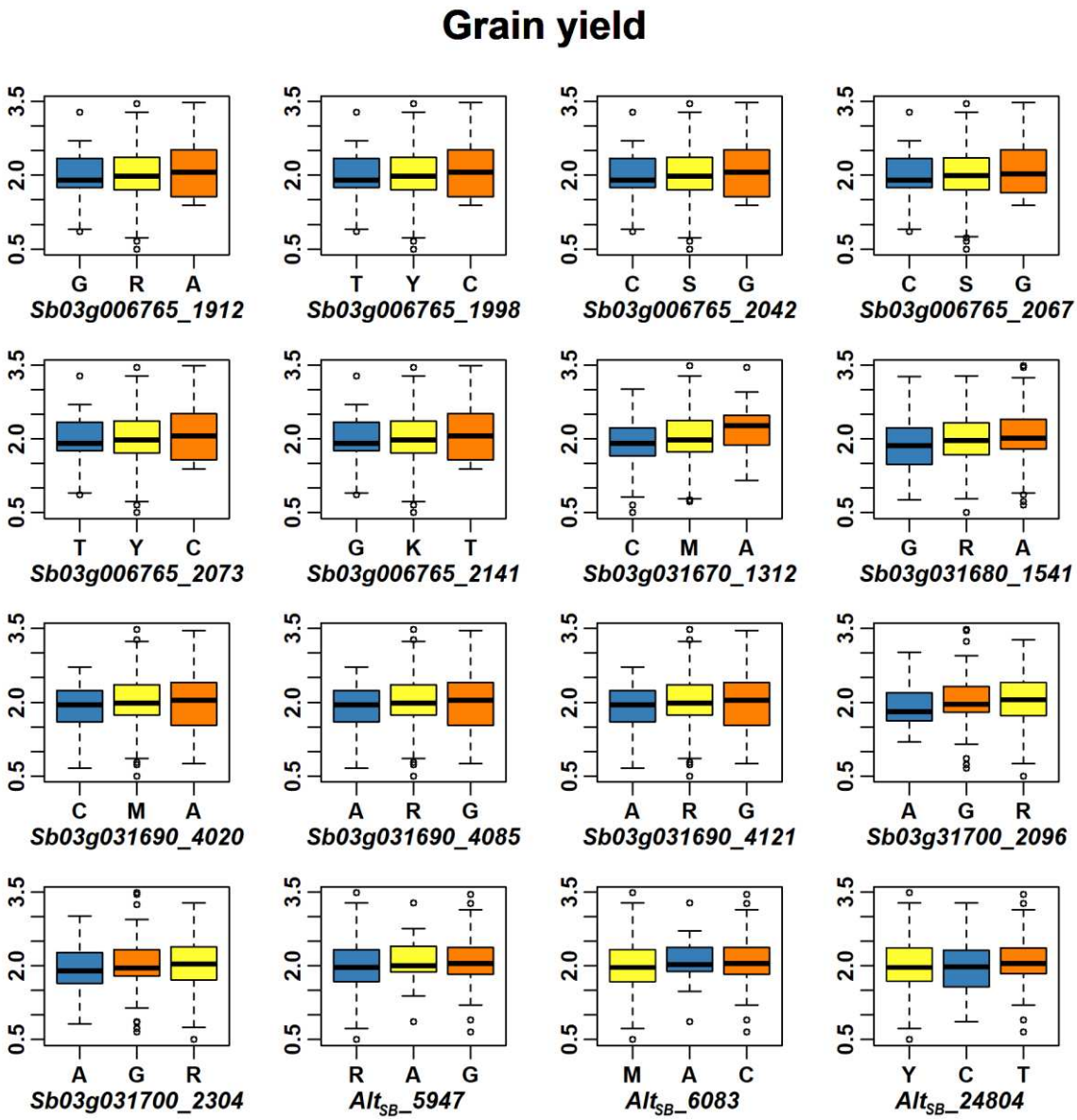


Fig. S3 Grain yield distributions for SNP loci within *SbPSTOL1* genes (*Sb03g006765*, *Sb03g031670*, *Sb03g031680*, *Sb03g031690*, *Sb03g031700*) and within the *Alt$_{SB}$* locus (*Alt$_{SB}$_5947*, *Alt$_{SB}$_6083*, *Alt$_{SB}$_24804*), which were previously associated with grain-yield under low-P (Hufnagel et al. 2014) and Al tolerance (Caniato et al. 2014). The colors blue, yellow and orange indicate the grain yield distribution for BRP13R progeny homozygous for inferior alleles, heterozygous and homozygous for the superior alleles, respectively.

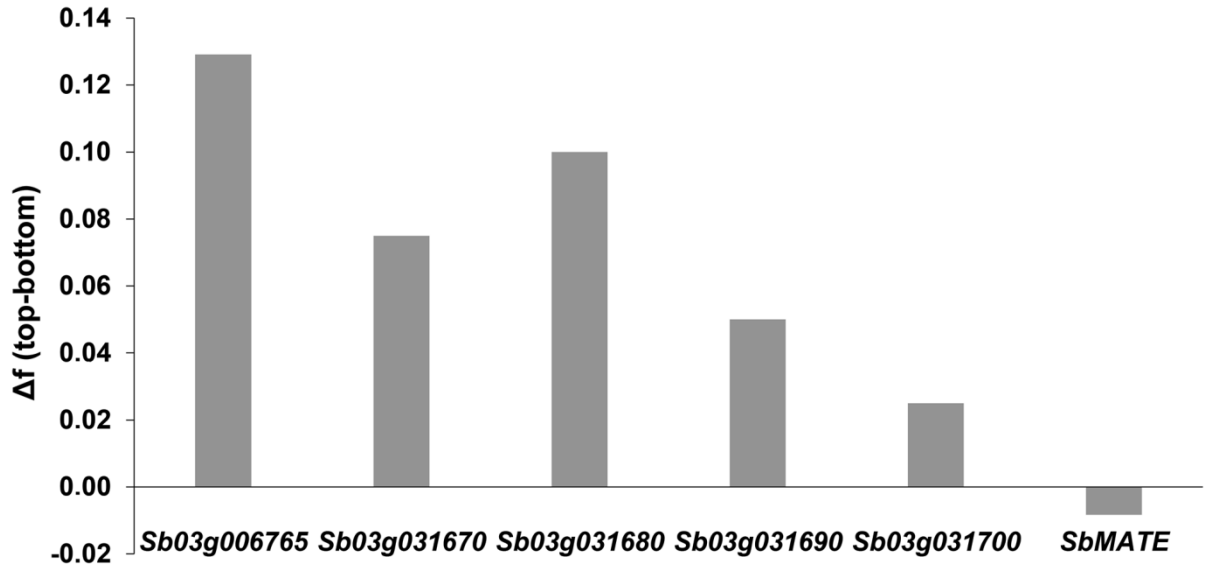


Fig. S4 Difference in favorable allele frequencies between the top and bottom BRP13R progeny (Δf) selected for grain yield. Average Δf per gene was calculated for *SbPSTOL1* genes (*Sb03g006765*, S*b03g031670*, *Sb03g031680*, *Sb03g031690*, *Sb03g031700*) and *SbMATE*, which were previously associated with grain-yield under low-P (Hufnagel et al. 2014) and Al tolerance (Caniato et al. 2014).

Supplemental Tables

Table S1 Sorghum founder restorer lines used in the formation of BRP13R

Restorer lines	Other names	IS designation	Parentage	Race	Origin	Cross
BR005R	CMSXS116R, SC326-6	IS3758C	IS375 BC4 - dwarf Martin B	Caudatum	Ethiopia	BRP3R[a] × BR005R
BR012R	CMSXS178R	-	Derived from (SC748-5 × SC326-6)	-	Brazil	BRP3R × BR012R
CMSXS173R	SC748-5	IS3552C	IS3552 BC4 - dwarf Martin B	Caudatum - Guineense	Sudan	BRP3R × CMSXS173R
BR012(SC549)[c]	-	-	-	-	-	BRP3R × BR012(SC549)
BR012	CMSXS178R	-	Derived from (SC748-5 × SC326-6)	-	Brazil	-
SC549[c]	-	IS3625C	IS3625 BC4 - dwarf Martin B	Guinea	Nigeria	-
BR012(5DX61/6/2)	-	-	-	-	-	BRP3R × (BR 012 × 5DX61/6/2)
5DX61/6/2[c]	-	-	-	-	EARC[b] - Uganda	-
CMSXS225R[c]	-	-	Derived from (CMSXS110 × CMSXS153)	-	Brazil	BRP3R × CMSXS 225R
CMSXS153[c]	156-P-5-2-1				EARC[b]	
CMSXS226R[c]	-	-	Derived from (SC283 × SC326-6)	-	Brazil	BRP3R × CMSXS 226R
SC283[c]	IS7173C			Guinea	USA	
CMSXS110R	Tx430	-	Derived from Tx2536 × SC170	-	USA	BRP3R × CMSXS110R
CMSXS106R	Tx2536	-	-	-	USA	BRP3R × CMSXS106R
CMSXS108R	TAM428	-	Derived from (Tx406 × IS12610)	-	USA	BRP3R × CMSXS108R
CMSXS179		-	Derived from (BRP3R × SC326-6)	-	Brazil	BRP3R × CMSXS179
CMSXS180	-	-	Derived from (BRP3R × SC326-6)	-	Brazil	BRP3R × CMSXS180
9929044	-	-	-	-	Brazil	BRP3R × 9929044
9929048	-	-	-	-	Brazil	BRP3R × 9929048
9930002	-	-	-	-	Brazil	BRP3R × 9930002
GR 1-1-1	-	-	Selection greenbug resistance	-	USA	BRP3R × GR 1-1-1
BR501R[c]	Brandes	-	Derived from Collier 706-C and MN1500	-	USA	BRP3R × BR 501 R
QL3	-	IS18757	Queensland 3 – tolerant to sugarcane mosaic virus and Downy Mildew	-	Australia	BRP3R × QL 3
IPA 1011	-	-	-	-	African collection	BRP3R × IPA 1011
156-8-5 Serere[c]	-	-	-	-	EARC[b]	BRP3R × 156-8-5 Serere
SC103	-	-	-	-	USA	BRP3R × SC 103
9929052	-	-	-	-	Brazil	BRP3R × 9929052
CMSXS169	-	-	Derived from TX2536	-	USA	BRP3R × CMSXS 169
CMSXS184	-	-	Derived from SC326 with radiation	-	Brazil	BRP3R × CMSXS 184

[a]BRP3R Brazilian random mating population 3 Restore, corresponding to Purdue population 3R. [b]Eastern Africa Regional Resource Centre. [c]Tolerant to Al toxicity.

Table S2 Heritabilities for traits assessed in low-P conditions (field and hydroponics) and under Al-stress in hydroponics

		Trait	Heritability	Mean	Maximum	Minimum
Field		Grain yield (ton ha^{-1}, Gy)	0.61	2.01	3.48	0.55
		Flowering time (days, FT)	0.81	64.15	79.5	64.15
		Plant height (cm, PH)	0.75	148.71	188.5	118
		Plant phosphorus content (leaves and stem, ton ha^{-1}, Pp)	0.58	2.05	7.33	0.41
		Grain phosphorus content (ton ha^{-1}, Pg)	0.64	4.24	8.92	1.32
		Total phosphorus content (ton ha^{-1}, Pt)	0.55	6.32	14.77	1.96
		Plant dry matter (ton ha^{-1}, PDM)	0.57	2.72	6.48	0.78
		Grain dry matter (ton ha^{-1}, GDM)	0.53	1.77	2.89	0.48
Hydrponic	**P**	Root length (cm, RL)	0.54	346.74	578.31	181.55
		Root diameter (mm, RD)	0.37	0.90	1.08	0.77
		Total root surface area (cm², SA)	0.51	96.89	174.99	48.8
		Surface area of super fine roots (cm² - 0 mm < RD ≤ 1 mm, SA1)	0.54	49.64	89.43	23.72
		Surface area of fine roots (cm² - 1 mm < RD ≤ 2 mm, SA2)	0.43	34.22	55.98	15.85
		Surface area of thicker roots (cm² - 2 mm < RD ≤ 4.5 mm, SA3)	0.47	3.74	10.64	1.19
		Root volume (cm³, RV)	0.49	2.16	3.88	1.02
		Volume of fine roots (cm³ - 1 mm < RD ≤ 2 mm, V2)	0.43	1.15	1.95	0.5
		Shoot dry matter (g, SDM)	0.61	2.01	3.02	1.07
		Root dry matter (g, RDM)	0.64	1.91	3.02	0.76
		Shoot phosphorus content (g, Ps)	0.74	7.58	18.84	4.12
		Root phosphorus content (g, Pr)	0.51	3.97	8.44	1.63
	Al	Relative net root growth (RNRG)	0.73	75.69	326.04	18.17

P: phosphorus, Al: aluminum.

Table S3 Phenotypic correlations among traits evaluated in low-P conditions (field and hydroponics) and Al-stress conditions (hydroponics)

r/*p*-value	Gy	FT	PH	Pp	Pg	Pt	PDM	GDM	RNRG	RL	RD	SA	SA1	SA2	SA3	RV	V2	SDM	RDM	Ps	Pr
Gy		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.35	0.25	0.24	0.22	0.16	0.65	0.23	0.34	0.61	0.56	0.31	0.47	0.86
FT	0.21		0.00	0.51	0.02	0.26	0.00	0.00	0.27	0.76	0.72	0.66	0.81	0.56	0.80	0.98	0.54	0.05	0.14	0.19	0.82
PH	0.30	0.27		0.01	0.00	0.01	0.00	0.00	0.72	0.05	0.62	0.07	0.08	0.13	0.20	0.08	0.12	0.15	0.14	0.06	0.00
Pp	0.35	-0.05	0.17		0.00	0.00	0.00	0.00	0.14	0.93	0.18	0.61	0.96	0.28	0.49	0.54	0.23	0.70	0.96	0.15	0.52
Pg	0.76	0.16	0.26	0.51		0.00	0.00	0.00	0.31	0.85	0.63	0.74	0.82	0.53	0.79	0.84	0.52	0.94	0.99	0.28	0.44
Pt	0.63	0.08	0.18	0.73	0.80		0.00	0.00	0.21	0.79	0.95	0.97	0.82	0.66	0.87	0.94	0.62	0.65	0.57	0.30	0.61
PDM	0.56	0.21	0.43	0.66	0.59	0.66		0.00	0.15	0.30	0.35	0.26	0.16	0.76	0.61	0.36	0.74	0.53	0.93	0.70	0.93
GDM	0.92	0.22	0.29	0.37	0.72	0.63	0.62		0.45	0.33	0.18	0.28	0.19	0.76	0.31	0.47	0.75	0.94	0.53	0.65	0.95
RNRG	-0.07	-0.08	-0.03	-0.11	-0.07	-0.09	-0.10	-0.05		0.06	0.18	0.06	0.09	0.21	0.01	0.10	0.17	0.00	0.03	0.89	0.79
RL	0.08	-0.02	0.14	0.01	0.01	-0.02	0.07	0.07	0.13		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RD	-0.08	-0.03	-0.03	0.09	-0.03	0.00	-0.07	-0.10	-0.10	-0.28		0.40	0.00	0.00	0.00	0.04	0.00	0.49	0.25	0.01	0.00
SA	0.09	-0.03	0.13	0.04	0.02	0.00	0.08	0.08	0.13	0.96	-0.06		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SA1	0.10	-0.02	0.13	0.00	0.02	-0.02	0.10	0.09	0.12	0.98	-0.36	0.93		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SA2	0.03	-0.04	0.11	0.08	0.04	0.03	0.02	0.02	0.09	0.70	0.36	0.84	0.61		0.00	0.00	0.00	0.00	0.00	0.00	0.00
SA3	0.09	0.02	0.09	0.05	0.02	0.01	0.04	0.07	0.17	0.50	0.35	0.62	0.42	0.73		0.00	0.00	0.00	0.00	0.00	0.00
RV	0.07	0.00	0.12	0.04	0.01	0.01	0.07	0.05	0.12	0.86	0.14	0.95	0.82	0.91	0.7		0.00	0.00	0.00	0.00	0.00
V2	0.04	-0.04	0.11	0.09	0.05	0.04	0.02	0.02	0.10	0.68	0.39	0.82	0.59	1.00	0.74	0.90		0.00	0.00	0.00	0.00
SDM	0.04	-0.14	0.10	0.03	-0.01	-0.03	-0.04	0.01	0.21	0.78	0.05	0.81	0.71	0.75	0.64	0.80	0.74		0.00	0.00	0.00
RDM	0.07	-0.10	0.10	0.00	0.00	-0.04	-0.01	0.04	0.15	0.75	0.08	0.79	0.67	0.76	0.71	0.81	0.76	0.82		0.00	0.00
Ps	0.05	-0.09	0.13	0.10	0.08	0.07	0.03	0.03	-0.01	0.49	0.18	0.56	0.44	0.60	0.50	0.59	0.59	0.67	0.59		0.00
Pr	0.01	-0.02	0.21	0.05	0.05	0.04	0.01	0.00	-0.02	0.33	0.33	0.42	0.27	0.54	0.48	0.50	0.55	0.46	0.62	0.60	

Traits assessed in a low-P soil are: Gy - grain yield (ton ha^{-1}); FT - flowering time (days); PH - plant height (cm); Pp - plant phosphorus content (leaves and stem - ton ha^{-1}); Pg - grain phosphorus content (ton ha^{-1}); Pt - total phosphorus content (ton ha^{-1}); PDM - plant dry matter (ton ha^{-1}); GDM - grain dry matter (ton ha^{-1}); Root morphology traits assessed in a low-P nutrient solution are: RL - root length (cm); RD - root diameter (mm); SA - total root surface area (cm^2); SA1 - surface area of super fine roots (cm^2 - 0 mm < RD ≤ 1 mm); SA2 - surface area of fine roots (cm^2 - 1 mm < RD ≤ 2 mm); SA3 - surface area of thicker roots (cm^2 - 2 mm < RD ≤ 4.5 mm); RV - root volume (cm^3); V2 - volume of fine roots (cm^3 - 1 mm < RD ≤ 2 mm); SDM - shoot dry matter (g); RDM - root dry matter (g); Ps - shoot phosphorus content (g); Pr - root phosphorus content (g); Al tolerance assessed in hydroponics: RNRG: relative net root growth.
r: correlation.

Table S4 Imputation accuracies per genotypic class with varying imputation windows

Genotypic class	Window size							
	10kb	**50kb**	**100kb**	**150kb**	**500kb**	**1Mb**	**5Mb**	**10Mb**
Homozygous - major allele	99.63	99.67	99.69	99.69	99.75	99.77	99.78	99.78
Heterozygous	38.87	43.06	44.74	45.32	45.95	45.43	45.55	45.49
Homozygous - minor allele	31.24	41.33	47.37	49.07	54.92	55.67	56.59	57.39
Global	97.33	97.54	97.64	97.67	97.76	97.76	97.78	97.79

The original dataset had 50% or less missing genotypes per site and read depth ≥ 6. kb: kilobase pairs; Mb: megabase pairs.

Table S5 Summary of association mapping results including marker effects, p-values, coefficient of determination (R^2) and favorable alleles

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
FT	S1_60416633	1	60416633	0.83	0.67	-2.70 *	4.16	3.78	7.43	G
FT	S1_69085561	1	69085561	2.53	1.76	-2.54	3.41	3.96	7.77	C
FT	S1_161097084	3	9674125	1.09 **A+D**	3.92	-	-	3.92	6.32	A
FT	S1_206154035	3	54731076	1.84	0.64	-2.71 *	4.40	3.99	7.82	C
FT	S1_242085844	4	16254388	1.86	0.11	-2.94 *	4.37	3.66	7.20	T
FT	S1_251051876	4	25220420	1.07 **A+D**	3.87	-	-	3.87	6.26	C
FT	S1_281145790	4	55314334	1.82 *	4.18	-0.12	0.08	3.46	6.81	G
FT	S1_298536315	5	4738000	1.35 ***A+D**	4.24	-	-	4.24	6.90	A
FT	S1_307309561	5	13511246	1.00 *	3.90	1.29	1.25	3.98	7.79	T
FT	S1_326171205	5	32372890	1.16 ***A+D**	4.27	-	-	4.27	6.95	C
FT	**S1_340669463**	5	46871148	1.13 ***A+D**	4.05	-	-	4.05	6.56	G
FT	S1_355312210	5	61513895	1.62 *	3.88	0.38	0.32	3.28	6.48	T
FT	S1_362404293	6	6362373	3.24 *	7.05	-1.07	0.86	6.67	12.68	G
FT	S1_366601212	6	10559292	2.78 *	7.42	-0.84	0.82	7.00	13.24	A
FT	S1_370257748	6	14215828	1.59 ***A+D**	6.22	-	-	6.22	10.47	A
FT	S1_374059372	6	18017452	1.12	3.26	1.55	2.26	4.27	8.33	C
FT	S1_378510936	6	22469016	1.60 ***A+D**	6.38	-	-	6.38	10.67	C
FT	S1_382702036	6	26660116	1.51 ***A+D**	6.00	-	-	6.00	11.02	A
FT	S1_388269676	6	32227756	2.38 *	4.60	-0.61	0.41	4.00	7.83	C
FT	S1_391907484	6	35865564	**1.71** *	4.97	0.70	0.59	4.48	8.72	A
FT	S1_395217997	6	39176077	2.13 ***A+D**	5.25	0.10	0.06	4.47	8.70	A
FT	S1_543341880	9	5489179	1.06	3.81	-	-	3.81	6.14	A
GDM	**S1_21408227**	1	21408227	130.59 ***A+D**	4.20	-	-	4.20	7.81	A
GDM	S1_54059972	1	54059972	122.77 ***A+D**	3.75	-	-	3.75	6.72	T

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
GDM	**S1_146747737**	2	73019702	186.71 *A+D	0.03	358.60 *	4.71	3.97	8.52	G
GDM	S1_160338920	3	8915961	135.83	4.30	-	-	4.30	7.66	T
GDM	S1_225055858	3	73632899	222.44 *	4.51	-2.76	0.01	3.76	8.08	C
GDM	**S1_296618739**	5	2820424	133.16 *	3.98	148.26	1.31	4.11	8.78	C
GDM	S1_350329163	5	56530848	111.66	0.37	275.86 *	4.45	3.86	8.28	A
GDM	S1_391587780	6	35545860	135.26 *A+D	3.75	-	-	3.75	6.58	C
GDM	**S1_396798101**	6	40756181	47.17	3.50	295.35 *	3.97	6.07	12.70	G
GDM	S1_410562791	6	54520871	96.46	3.19	202.36	2.31	4.25	9.08	T
GDM	S1_481615399	7	63381362	136.63 *A+D	3.91	-	-	3.91	6.99	T
GDM	S1_521046290	8	38548245	127.02 *A+D	4.29	-	-	4.29	7.84	C
GDM	**S1_533287952**	8	50789907	136.11 *A+D	4.66	-	-	4.66	8.55	G
GDM	S1_537344960	8	54846915	241.67	0.66	330.76 *	4.17	3.78	8.13	C
GDM	**S1_587216909**	9	49364208	96.32	3.65	226.02	2.76	5.09	10.77	A
GDM	S1_594353037	9	56500336	127.66 *A+D	4.21	-	-	4.21	7.53	A
GDM	S1_655299412	10	57992365	161.69	0.01	318.15 *	4.11	3.40	7.33	A
Gy	S1_21158289	1	21158289	0.01	1.65	0.31	3.41	3.86	8.39	T
Gy	S1_46421958	1	46421958	0.28	3.67	0.17	1.48	3.96	8.59	C
Gy	S1_78121019	2	4392984	0.15	0.25	0.36 *	3.82	3.19	6.99	G
Gy	S1_208828314	3	57405355	0.19	0.25	0.38 *	3.90	3.28	7.17	G
Gy	S1_289681597	4	63850141	0.25	0.85	0.35 *	4.00	3.77	8.19	G
Gy	S1_295999524	5	2201209	0.07	2.26	0.26	2.72	3.76	8.18	G
Gy	**S1_340669463**	5	46871148	0.15 *A+D A+D	3.80	-	-	3.80	6.78	G
Gy	S1_347549632	5	53751317	0.15	4.08	-	-	4.08	7.60	A
Gy	S1_350323609	5	56525294	0.29	0.23	0.43 *	4.48	3.81	8.29	A

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
Gy	S1_381970924	6	25929004	0.16 A+D	3.83	-	-	3.83	7.07	A
Gy	**S1_396798101**	6	40756181	0.11 *	4.26	0.29	2.96	5.85	12.41	G
Gy	S1_521046290	8	38548245	0.14 A+D	3.75	-	-	3.75	6.84	C
Gy	**S1_533287952**	8	50789907	0.15 A+D	4.26	-	-	4.26	7.83	G
Gy	**S1_582556694**	9	44703993	0.03	0.08	0.30 *	4.17	3.47	7.57	C
Gy	**S1_587216909**	9	49364208	0.09	2.82	0.24	2.36	3.95	8.57	A
Gy	**S1_613208087**	10	15901040	0.09	0.19	0.31 *	4.13	3.46	7.56	C
Gy	S1_641860550	10	44553503	0.14	0.04	0.33 *	4.13	3.42	7.47	T
Gy	S1_650246331	10	52939284	0.15 A+D	3.95	-	-	3.95	7.17	C
PDM	S1_15151700	1	15151700	249.97 A+D	4.54	-	-	4.54	7.71	C
PDM	S1_31357280	1	31357280	421.53 A+D	3.82	-	-	3.82	6.46	T
PDM	S1_38512761	1	38512761	377.19 A+D	4.47	-	-	4.47	7.61	G
PDM	S1_42063798	1	42063798	282.22 *	3.76	139.87	0.46	3.25	6.66	T
PDM	S1_54073570	1	54073570	254.03 A+D	3.82	-	-	3.82	6.46	A
PDM	S1_57841084	1	57841084	245.57 A+D	3.85	-	-	3.85	6.49	T
PDM	S1_65749085	1	65749085	407.87 A+D	4.07	-	-	4.07	6.84	G
PDM	S1_80582145	2	6854110	277.56 A+D	4.73	-	-	4.73	8.40	A
PDM	S1_146458740	2	72730705	337.97 *	4.07	105.71	0.46	3.54	7.23	C
PDM	**S1_296618739**	5	2820424	321.20 *	4.23	96.30	0.33	3.61	3.37	C
PDM	**S1_341450513**	5	47652198	247.73 A+D	4.71	-	-	4.71	8.06	A
PDM	**S1_396798101**	6	40756181	117.30	3.21	418.87	2.73	4.65	9.37	G
PDM	S1_428860796	7	10626759	222.21 A+D	4.14	-	-	4.14	7.07	G
PDM	**S1_461864650**	7	43630613	547.56 A+D	4.95	-	-	4.95	8.46	C
PDM	S1_476131785	7	57897748	253.84 A+D	3.83	-	-	3.83	6.40	T

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
PDM	S1_481359002	7	63124965	415.54	3.57	304.02 *	1.95	4.28	8.66	G
PDM	S1_485032308	8	2534263	72.86	0.16	438.51	3.79	3.13	6.43	T
PDM	S1_587216883	9	49364182	327.52 *	4.39	104.06	0.39	3.80	7.73	A
PDM	S1_591360155	9	53507454	450.76	3.02	-329.57	2.02	3.82	7.78	C
PDM	S1_594231567	9	56378866	285.99 *	4.68	191.72	0.87	4.41	8.92	G
PDM	S1_609075978	10	11768931	288.19 A+D	4.69	-	-	4.69	8.12	C
PDM	S1_613596988	10	16289941	273.15 A+D	3.99	-	-	3.99	6.98	T
Pg	S1_14530781	1	14530781	0.44 A+D	5.27	-	-	5.27	9.69	A
Pg	**S1_21408227**	1	21408227	0.42 A+D	4.74	-	-	4.74	8.99	A
Pg	**S1_54039814**	1	54039814	0.39 A+D	4.19	-	-	4.19	7.73	C
Pg	**S1_86668393**	2	12940358	0.29	0.11	0.79 *	3.89	3.21	7.00	A
Pg	**S1_146747737**	2	73019702	0.75	0.57	1.03 *	4.28	3.83	8.30	G
Pg	S1_151489093	3	66134	0.39 A+D	4.13	-	-	4.13	7.67	C
Pg	S1_229965036	4	4133580	0.39 A+D	4.02	-	-	4.02	7.32	A
Pg	S1_295736381	5	1938066	0.92 A+D	5.11	-	-	5.11	9.32	C
Pg	S1_336749746	5	42951431	0.76 *	4.40	-	-	4.40	7.99	T
Pg	S1_409390332	6	53348412	0.84 A+D	4.56	-0.28	0.57	4.07	8.79	C
Pg	S1_424140922	7	5906885	0.56 *	3.76	-	-	3.76	6.71	G
Pg	S1_535505137	8	53007092	0.66	3.89	0.40	1.41 *	4.11	8.87	C
Pg	**S1_613208087**	10	15901040	0.44	1.03	0.79	3.98	3.88	8.40	C
PH	S1_25624633	1	25624633	4.49 A+D	4.23	-	- *	4.23	7.60	T
PH	S1_135151350	2	61423315	0.84	0.67	8.51	4.73	4.32	9.18	T
PH	**S1_145105735**	2	71377700	6.53 A+D	4.54	-	- *	4.54	8.11	G
PH	S1_148465787	2	74737752	5.18	0.35	9.50	4.29	3.69	7.90	A

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
PH	S1_211122560	3	59699601	4.06	0.21	10.19 *	4.64	3.96	8.45	A
PH	S1_222068214	3	70645255	6.84 *	2.88	6.18	2.50	4.13	8.79	C
PH	S1_229856008	4	4024552	6.24	4.15	1.70	0.41	3.58	7.68	C
PH	S1_271732607	4	45901151	2.08	0.94	8.14 *	4.33	4.15	8.84	T
PH	S1_276390282	4	50558826	0.35	0.04	8.34 *	4.57	3.84	8.20	C
PH	S1_297471093	5	3672778	4.89	0.75	12.88	6.13	5.73	11.98	T
PH	S1_305201339	5	11403024	7.32 A+D	4.75	-	-	4.75	8.51	G
PH	S1_316639110	5	22840795	4.72 A+D	3.98	-	-	3.98	7.04	T
PH	**S1_341450513**	5	47652198	4.39 A+D	4.59	-	-	4.59	8.23	A
PH	S1_355732521	5	61934206	4.06 A+D	3.85	-	-	3.85	6.97	C
PH	S1_359318471	6	3276551	6.30 *	4.15	-	-	4.15	7.42	C
PH	S1_394180948	6	38139028	6.58 A+D	4.62	1.88	0.33	3.98	8.50	T
PH	S1_398344107	6	42302187	8.44	6.79	-	-	6.79	12.40	C
PH	S1_408964333	6	52922413	2.45 A+D	3.11	6.50	2.29	4.16	8.85	G
PH	S1_471675028	7	53440991	5.40 A+D	4.30	-	-	4.30	7.79	T
PH	S1_477872980	7	59638943	4.31 A+D	3.77	-	-	3.77	6.61	A
PH	S1_482800222	8	302177	5.47 A+D	3.83	-	-	3.83	6.72	T
PH	S1_486706783	8	4208738	5.15 A+D	5.56	-	-	5.56	10.23	T
PH	S1_540168834	9	2316133	6.91	4.60	-	-	4.60	8.22	G
PH	S1_549529562	9	11676861	1.58 A+D	1.25	9.75 *	4.24	4.31	9.17	T
PH	S1_588619886	9	50767185	5.29 A+D	4.72	-	-	4.72	8.50	G
PH	S1_595408844	9	57556143	7.62 A+D	5.06	-	-	5.06	9.16	A
PH	S1_597420195	10	113148	4.09	3.82	-	-	3.82	7.02	G
PH	S1_602459440	10	5152393	2.36	0.06	9.46 *	5.27	4.52	9.58	T

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
PH	S1_606683736	10	9376689	0.55	1.95	8.35	3.62	4.32	9.19	C
PH	S1_651965662	10	54658615	4.21 A+D	3.96	-	-	3.96	7.23	A
Pp	S1_51011831	1	51011831	0.06	1.41	0.57 *	4.22	4.43	9.67	C
Pp	S1_54531403	1	54531403	0.24 A+D	3.94	-	-	3.94	7.20	G
Pp	S1_67345901	1	67345901	0.32 A+D	4.86	-	-	4.86	9.12	C
Pp	S1_233408177	4	7576721	0.29 A+D	4.29	-	-	4.29	7.90	A
Pp	S1_236144144	4	10312688	0.51 A+D	4.17	-	-	4.17	7.61	A
Pp	S1_298850937	5	5052622	0.28 A+D	3.77	-	-	3.77	6.80	A
Pp	S1_308851706	5	15053391	0.47 A+D	3.96	-	-	3.96	7.18	T
Pp	**S1_461864650**	7	43630613	0.58 A+D	5.37	-	-	5.37	9.98	C
Pp	S1_477930623	7	59696586	0.28 A+D	4.62	-	-	4.62	8.62	G
Pp	S1_594903194	9	57050493	0.23 A+D	4.17	-	-	4.17	7.94	C
Pp	S1_613253980	10	15946933	0.30 A+D	3.84	-	-	3.84	6.93	G
Pp	S1_642846562	10	45539515	0.19 A+D	3.84	-	-	3.84	7.22	G
Pr	S1_10974423	1	10974423	0.46 A+D	3.87	-	-	3.87	7.00	T
Pr	S1_42953217	1	42953217	0.45 A+D	4.68	-	-	4.68	8.67	T
Pr	S1_53580556	1	53580556	0.39 A+D	4.51	-	-	4.51	8.35	G
Pr	S1_59260003	1	59260003	0.50 A+D *	3.81	-	-	3.81	6.87	A
Pr	S1_65205800	1	65205800	0.49	4.37	0.07	0.16	3.67	8.07	G
Pr	S1_81943538	2	8215503	0.09	0.21	0.66 *	3.98	3.33	7.36	A
Pr	S1_85937831	2	12209796	0.32 A+D	4.08	-	-	4.08	7.75	A
Pr	S1_127959975	2	54231940	0.31 A+D	3.92	-	-	3.92	7.10	G
Pr	**S1_145193535**	2	71465500	0.70 A+D	3.85	-	-	3.85	6.94	T
Pr	S1_162424657	3	11001698	0.29	0.20	0.74 *	3.83	3.18	7.04	G

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
Pr	S1_281200057	4	55368601	0.52 * A+D	3.89	0.00	0.01	3.18	7.03	C
Pr	S1_296360850	5	2562535	0.35	4.19	-	-	4.19	7.67	C
Pr	S1_406518953	6	50477033	0.48	0.25	0.78 *	3.96	3.33	7.37	C
Pr	S1_531535379	8	49037334	0.25	0.81	0.96 *	4.11	3.84	8.43	T
Pr	S1_562908720	9	25056019	0.57 *	5.81	-0.13	0.28	5.09	11.03	C
Pr	S1_595234637	9	57381936	0.20	0.47	0.75 *	4.00	3.49	7.70	G
Pr	S1_600779329	10	3472282	0.42 A+D	4.66	-	-	4.66	8.70	A
Pr	S1_610048221	10	12741174	0.40 A+D	3.79	-	-	3.79	6.94	A
Pr	S1_657395792	10	60088745	0.68 A+D	4.35	-	-	4.35	7.94	G
Ps	S1_5028801	1	5028801	0.48 A+D	3.87	-	-	3.87	7.09	C
Ps	S1_11292709	1	11292709	0.64 A+D	3.85	-	-	3.85	7.24	T
Ps	S1_54415741	1	54415741	0.98 A+D	3.82	-	-	3.82	6.99	G
Ps	S1_64713799	1	64713799	0.72 A+D	4.11	-	-	4.11	7.66	T
Ps	S1_70027560	1	70027560	0.71 A+D	3.76	-	-	3.76	6.87	C
Ps	S1_86315887	2	12587852	0.35	0.25	0.96 *	3.82	3.20	7.19	T
Ps	S1_134694342	2	60966307	0.51 A+D	4.48	-	-	4.48	8.61	A
Ps	S1_145529973	2	71801938	0.70 A+D	5.41	-	-	5.41	10.30	T
Ps	S1_150381870	2	76653835	0.23	0.32	0.92 *	3.76	3.18	7.15	G
Ps	S1_221290834	3	69867875	0.16	0.11	1.10 *	4.82	4.09	9.11	G
Ps	S1_279535720	4	53704264	1.31 *	5.01	0.61	1.45	5.20	11.42	G
Ps	S1_296360817	5	2562502	0.52 A+D	4.40	-	-	4.40	8.19	T
Ps	S1_371383342	6	15341422	0.56 A+D	3.83	-	-	3.83	7.02	A
Ps	S1_389761352	6	33719432	0.60 A+D	4.78	-	-	4.78	9.06	T
Ps	**S1_408281673**	6	52239753	0.03	0.50	1.15 *	5.08	4.54	10.05	C

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
Ps	S1_416258408	6	60216488	0.54 A+D	4.89	-	-	4.89	9.52	C
Ps	S1_487635842	8	5137797	0.65 A+D	5.45	-	-	5.45	10.65	A
Ps	S1_534165292	8	51667247	0.03	1.61	1.05	3.57	3.98	8.88	C
Ps	S1_560220745	9	22368044	0.49 A+D	4.14	-	-	4.14	7.76	C
Ps	S1_564439482	9	26586781	0.48 A+D	3.91	-	-	3.91	7.44	A
Ps	S1_596584560	9	58731859	0.18	0.34	1.07 *	4.18	3.58	8.03	C
Ps	S1_605391727	10	8084680	0.97 A+D	3.99	-	-	3.99	7.35	G
Ps	S1_650391392	10	53084345	1.06 A+D	4.68	-	-	4.68	8.74	C
Ps	S1_655048088	10	57741041	0.30	0.22	0.99 *	3.81	3.17	7.14	T
Pt	S1_15605312	1	15605312	0.41	0.44	1.34 *	4.48	3.93	8.46	T
Pt	**S1_54039814**	1	54039814	0.61 A+D	5.02	-	-	5.02	9.35	C
Pt	S1_60441337	1	60441337	0.50 A+D	3.81	-	-	3.81	7.01	T
Pt	**S1_86668393**	2	12940358	0.32	0.09	1.20 *	4.36	3.65	7.87	A
Pt	S1_288025227	4	62193771	0.65	1.20	1.12 *	4.28	4.31	9.24	G
Pt	S1_357935805	6	1893885	0.32	1.48	1.38	3.65	3.95	8.50	C
Pt	S1_529466579	8	46968534	1.37 A+D	5.09	-	-	5.09	9.24	A
Pt	**S1_582556694**	9	44703993	0.01	0.17	1.11 *	4.33	3.64	7.87	G
Pt	**S1_587216909**	9	49364208	0.33	3.01	0.96	2.71	4.44	9.50	A
Pt	S1_594903194	9	57050493	0.54 A+D	4.01	-	-	4.01	7.43	C
Pt	**S1_613208087**	10	15901040	0.50	0.59	1.15 *	4.16	3.72	8.04	C
RD	S1_21134033	1	21134033	0.02 A+D	3.74	-	-	3.74	6.98	T
RD	S1_55233715	1	55233715	0.02 *	3.02	-0.02	1.93	3.75	8.41	G
RD	S1_156114939	3	4691980	0.03 *	4.13	-0.01	0.59	3.69	8.28	G
RD	S1_279112255	4	53280799	0.03	4.95	-0.01	0.58	4.45	9.92	A

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
RD	S1_404831261	6	48789341	0.02 A+D*	4.25	-	-	4.25	7.98	A
RD	S1_409672785	6	53630865	0.02 *	4.72	-0.01	0.80	4.39	9.78	C
RD	S1_647982235	10	50675188	0.02 *	3.92	0.01	1.16	3.92	8.78	C
RDM	S1_50871588	1	50871588	0.24 A+D	3.87	-	-	3.87	7.13	A
RDM	S1_55364790	1	55364790	0.14 A+D	3.95	-	-	3.95	7.33	A
RDM	S1_72149849	1	72149849	0.25 *	3.09	0.18	2.07	3.93	8.81	C
RDM	S1_75266067	2	1538032	0.15 *	3.78	0.10	1.10	3.74	8.40	C
RDM	**S1_141514225**	2	67786190	0.17	3.96	0.04 *	0.36	3.37	7.60	C
RDM	**S1_153679712**	3	2256753	0.13	0.64	0.23	3.86	3.47	7.82	C
RDM	S1_156727829	3	5304870	0.13 A+D	4.08	-	-	4.08	7.60	C
RDM	S1_162220640	3	10797681	0.13 A+D	3.78	-	-	3.78	6.95	T
RDM	S1_215075130	3	63652171	0.16 A+D	4.20	-	-	4.20	7.80	C
RDM	**S1_216344190**	3	64921231	0.15 A+D	3.94	-	-	3.94	7.27	G
RDM	**S1_267951018**	4	42119562	0.18 A+D	4.85	-	-	4.85	9.19	A
RDM	**S1_358161040**	6	2119120	0.14	4.51	-	-	4.51	8.43	C
RDM	**S1_408281673**	6	52239753	0.07	0.12	0.28 *	5.28	4.53	10.08	C
RDM	S1_414219495	6	58177575	0.00	0.01	0.24 *	4.69	3.95	8.85	C
RDM	S1_419264560	7	1030523	0.13 A+D	4.52	-	-	4.52	8.50	A
RDM	S1_528272391	8	45774346	0.18 *	3.82	-0.03	0.26	3.19	7.20	T
RDM	S1_595346408	9	57493707	0.14 A+D	3.88	-	-	3.88	7.20	G
RL	**S1_46958948**	1	46958948	26.85 A+D	4.07	-	-	4.07	7.42	C
RL	**S1_56164573**	1	56164573	21.55 *	0.85	45.24 *	3.88	3.64	7.99	A
RL	**S1_141514225**	2	67786190	33.35 *	4.01	21.62	1.20	4.04	8.82	C
RL	**S1_204963630**	3	53540671	40.39	5.34	11.37	0.51	4.77	10.33	T

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
RL	**S1_222076254**	3	70653295	30.46 A+D	5.60	-	-	5.60	10.75	A
RL	**S1_223572404**	3	72149445	22.30 A+D	3.86	-	-	3.86	7.17	A
RL	**S1_292536180**	4	66704724	19.22	0.91	47.27 *	4.46	4.25	9.25	C
RL	S1_294169656	5	371341	25.48 A+D	3.83	-	-	3.83	6.99	A
RL	S1_355673893	5	61875578	23.29 A+D	3.86	-	-	3.86	7.01	T
RL	**S1_403442858**	6	47400938	24.99	0.78	45.65 *	3.78	3.50	7.69	C
RL	**S1_413906486**	6	57864566	23.63 A+D	4.21	-	-	4.21	7.87	A
RL	S1_420122107	7	1888070	26.89 *	3.90	19.75	0.83	3.64	7.98	T
RL	S1_597422431	10	115384	45.24 A+D	3.90	-	-	3.90	7.04	A
RL	**S1_602476018**	10	5168971	4.61	0.32	49.83 *	4.50	3.88	8.49	C
RL	**S1_649351290**	10	52044243	24.64 A+D	3.79	-	-	3.79	6.84	T
RNRG	S1_4220530	1	4220530	0.07 A+D	4.73	-	-	4.73	8.65	C
RNRG	S1_50020081	1	50020081	0.06 *	4.69	-	-	4.69	8.47	C
RNRG	S1_129120842	2	55392807	0.12	4.76	-0.03	0.46	4.19	8.76	T
RNRG	S1_131345126	2	57617091	0.07 *	0.06	0.14 *	3.89	3.19	6.76	A
RNRG	S1_133646267	2	59918232	0.11 A+D	4.71	-0.03	0.51	4.18	8.74	C
RNRG	S1_137616146	2	63888111	0.11 *	3.97	-	-	3.97	6.87	G
RNRG	S1_141656269	2	67928234	0.09 *	3.75	0.03	0.40	3.20	6.76	T
RNRG	S1_222514387	3	71091428	0.13 A+D	6.60	0.03	0.60	6.05	12.38	C
RNRG	S1_352445350	5	58647035	0.13 A+D	4.11	-	-	4.11	7.15	A
RNRG	S1_454873509	7	36639472	0.06 A+D	4.62	-	-	4.62	8.11	A
RNRG	S1_491884258	8	9386213	0.06 A+D	5.05	-	-	5.05	9.24	A
RNRG	S1_520312530	8	37814485	0.09 A+D	4.09	-	-	4.09	7.16	G
RNRG	S1_568571433	9	30718732	0.06	3.81	-	-	3.81	6.56	G

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
RNRG	S1_595947603	9	58094902	0.12 *	2.10	0.12	3.12	4.00	8.38	T
RNRG	S1_618413863	10	21106816	0.13 A+D	4.35	-0.04	0.62	3.91	8.20	T
RNRG	S1_632715095	10	35408048	0.06 *	3.88	-	-	3.88	6.73	T
RNRG	S1_640280164	10	42973117	0.14	4.98	0.05	0.75	4.59	9.56	C
RV	**S1_46614527**	1	46614527	0.25	0.37	0.36 *	4.13 *	3.56	7.86	C
RV	**S1_56164573**	1	56164573	0.08	0.19	0.32	4.77	4.08	8.96	A
RV	**S1_140541779**	2	66813744	0.11 A+D	2.49	0.24	3.07	4.30	9.42	A
RV	**S1_201976881**	3	50553922	0.23 A+D	4.49	-	-	4.49	8.25	C
RV	**S1_222076254**	3	70653295	0.17	4.53	-	-	4.53	8.63	A
RV	**S1_293011871**	4	67180415	0.00	0.80	0.29 *	3.80	3.54	7.82	T
RV	S1_295378092	5	1579777	0.00	2.52	0.31	2.88	4.16	9.13	A
RV	**S1_346109042**	5	52310727	0.11	2.14	0.24	3.13	4.04	8.88	C
RV	**S1_408281673**	6	52239753	0.03	0.24	0.29 *	3.82 *	3.19	7.08	C
RV	**S1_599253388**	10	1946341	0.01	0.08	0.28	3.99	3.29	7.30	G
RV	**S1_602476018**	10	5168971	0.02 A+D	0.93	0.30 *	4.24	4.05	8.91	A
RV	S1_608399856	10	11092809	0.19	3.80	-	-	3.80	6.90	A
RV	**S1_647980048**	10	50673001	0.13	1.15	0.27 *	3.77	3.78	8.34	G
SA	**S1_46614527**	1	46614527	10.97	0.32	16.39 *	4.36 *	3.75	8.26	C
SA	**S1_56164573**	1	56164573	5.05	0.52	13.86 *	4.74 *	4.23	9.28	A
SA	S1_95923303	2	22195268	4.12 *	0.35	14.98	3.91	3.34	7.40	G
SA	**S1_141514225**	2	67786190	9.77 *	4.38	4.56	0.82	4.09	8.98	C
SA	**S1_204963630**	3	53540671	9.72 A+D	4.26	3.32	0.54	3.77	8.31	T
SA	**S1_222076254**	3	70653295	8.51	5.85	-	-	5.85	11.34	A
SA	**S1_293011871**	4	67180415	0.09	0.92	14.10 *	4.64	4.43	9.68	T

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
SA	**S1_346109042**	5	52310727	5.64	2.64	10.51 *	3.10	4.46	9.76	C
SA	**S1_403442858**	6	47400938	6.40	0.58	13.40	4.28	3.83	8.44	C
SA	**S1_413906486**	6	57864566	6.41 A+D	4.08	-	-	4.08	7.69	A
SA	**S1_493788836**	8	11290791	6.18 A+D	4.05	-	-	4.05	7.52	T
SA	**S1_599253388**	10	1946341	0.93	0.02	12.25 *	4.06	3.35	7.42	G
SA	**S1_602476018**	10	5168971	0.28	0.68	13.12 *	4.22	3.85	8.47	A
SA	**S1_649351290**	10	52044243	6.72 A+D	3.77	-	-	3.77	6.83	T
SA1	S1_2709945	1	2709945	7.21 A+D	3.76	-	-	3.76	6.76	C
SA1	**S1_46958948**	1	46958948	4.47 A+D	4.67	-	-	4.67	8.64	C
SA1	S1_56777857	1	56777857	5.71 A+D	5.19	-	-	5.19	9.59	A
SA1	S1_71971026	1	71971026	3.46	3.96	-	-	3.96	7.40	C
SA1	**S1_140541779**	2	66813744	1.81 *	1.65	6.47	3.49	3.95	8.65	A
SA1	**S1_204963630**	3	53540671	6.76 A+D	6.11	1.47	0.41	5.44	11.72	T
SA1	**S1_222076254**	3	70653295	4.93	6.10	-	-	6.10	11.75	A
SA1	**S1_293011871**	4	67180415	0.30	1.18	8.17 *	4.88	4.87	10.55	C
SA1	S1_355673893	5	61875578	3.80 A+D	4.24	-	-	4.24	7.80	T
SA1	**S1_413906486**	6	57864566	3.81 A+D	4.53	-	-	4.53	8.53	A
SA1	**S1_602476018**	10	5168971	0.04	0.68	7.85 *	4.68	4.28	9.33	A
SA2	**S1_19661942**	1	19661942	1.23	0.61	5.49 *	4.10	3.68	8.20	T
SA2	**S1_47891922**	1	47891922	3.06	0.00	5.93 *	4.10	3.39	7.59	G
SA2	**S1_56164573**	1	56164573	0.54	0.15	5.24	4.69	3.98	8.84	A
SA2	**S1_63521031**	1	63521031	3.28 A+D	3.79	-	-	3.79	6.91	C
SA2	**S1_141514225**	2	67786190	3.67 *	4.17	1.15	0.47	3.64	8.12	C
SA2	**S1_201976881**	3	50553922	3.97 A+D	4.64	-	-	4.64	8.64	C

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
SA2	**S1_210059602**	3	58636643	2.79 A+D	4.48	-	-	4.48	8.44	T
SA2	**S1_212830442**	3	61407483	0.75	2.73	5.03 A+D	3.32	4.75	10.46	T
SA2	**S1_223572404**	3	72149445	2.81	5.51	-	-	5.51	10.71	A
SA2	**S1_346109042**	5	52310727	1.29	1.65	4.86 *	4.33	4.74	10.43	C
SA2	**S1_408281673**	6	52239753	0.82	0.13	5.49 *	4.70	3.98	8.85	C
SA2	**S1_416720116**	6	60678196	2.50 A+D	3.91	-	-	3.91	7.19	G
SA2	**S1_493788836**	8	11290791	2.24 A+D	3.84	-	-	3.84	7.17	T
SA2	**S1_647980048**	10	50673001	1.72	0.76	5.02 *	4.56	4.22	9.36	G
SA3	S1_52623530	1	52623530	0.62 *	3.77	0.02	0.03	3.06	6.92	G
SA3	S1_163391335	3	11968376	0.59 A+D	3.88	-	-	3.88	7.17	C
SA3	**S1_201976881**	3	50553922	0.60 A+D	4.18	-	-	4.18	7.76	C
SA3	S1_229103730	4	3272274	0.53 A+D	4.31	-	-	4.31	8.17	G
SA3	S1_247669507	4	21838051	0.62 *	4.75	0.19	0.49	4.20	9.37	A
SA3	**S1_267951018**	4	42119562	0.53 A+D	3.78	-	-	3.78	6.98	A
SA3	S1_283220988	4	57389532	0.59 A+D	4.24	-	-	4.24	7.88	C
SA3	S1_287316327	4	61484871	0.59 *	3.87	-0.06	0.12	3.18	7.17	C
SA3	S1_339404371	5	45606056	0.53 A+D	4.39	-	-	4.39	8.41	G
SA3	S1_344799018	5	51000703	0.67 A+D	4.16	-	-	4.16	7.72	C
SA3	**S1_408281673**	6	52239753	0.07	0.33	0.90 *	4.90	4.26	9.51	C
SDM	S1_8560472	1	8560472	0.15 A+D	4.55	-	-	4.55	8.66	G
SDM	**S1_46614527**	1	46614527	0.16	0.04	0.31 *	4.11	3.41	7.68	C
SDM	**S1_88403727**	2	14675692	0.10	0.20	0.29 *	3.97	3.32	7.49	T
SDM	**S1_141514225**	2	67786190	0.18 *	3.93	0.11	1.08	3.87	8.68	C
SDM	**S1_153679712**	3	2256753	0.09	0.07	0.28 *	4.37	3.65	8.22	C

Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
SDM	S1_158842215	3	7419256	0.03	0.74	0.26 *	3.87	3.56	8.01	C
SDM	S1_211354668	3	59931709	0.15 A+D	4.33	-	-	4.33	8.21	T
SDM	**S1_216344190**	3	64921231	0.16 A+D	3.78	-	-	3.78	6.96	G
SDM	S1_227215798	4	1384342	0.16	3.53	0.14	1.68	4.00	8.96	G
SDM	S1_243255643	4	17424187	0.12	0.72	0.32 *	4.20	3.86	8.66	T
SDM	**S1_267951018**	4	42119562	0.18 A+D	4.02	-	-	4.02	7.48	A
SDM	**S1_292536180**	4	66704724	0.09	0.62	0.27 *	4.96	4.51	10.04	C
SDM	**S1_358161040**	6	2119120	0.16 A+D	4.49	-	-	4.49	8.41	C
SDM	**S1_408281673**	6	52239753	0.05	0.12	0.30 *	5.05	4.32	9.64	C
SDM	S1_419260038	7	1026001	0.13 A+D	4.01	-	-	4.01	7.49	G
SDM	S1_509389039	8	26890994	0.14 A+D *	4.13	-	-	4.13	7.78	A
SDM	S1_544465783	9	6613082	0.18	4.20	0.06	0.46	3.66	8.24	C
SDM	S1_599123384	10	1816337	0.17	0.14	0.29 *	3.97	3.29	7.43	A
V2	S1_10850288	1	10850288	0.18 A+D	3.82	-	-	3.82	6.98	G
V2	**S1_19661942**	1	19661942	0.05	0.55	0.20 *	4.14	3.68	8.22	T
V2	**S1_47891922**	1	47891922	0.11	0.00	0.21 *	4.21	3.49	7.82	G
V2	**S1_56164573**	1	56164573	0.01	0.22	0.18	4.62	3.94	8.78	A
V2	**S1_63521031**	1	63521031	0.12 A+D	3.97	-	-	3.97	7.30	G
V2	**S1_88403727**	2	14675692	0.07 *	0.16	0.19	3.78	3.12	7.02	T
V2	**S1_141514225**	2	67786190	0.13 A+D	4.02	0.04	0.51	3.52	7.88	C
V2	**S1_201976881**	3	50553922	0.14 A+D	4.71	-	-	4.71	8.79	C
V2	**S1_210059602**	3	58636643	0.10	4.44	-	-	4.44	8.38	T
V2	**S1_212830442**	3	61407483	0.03	2.90	0.18	3.32	4.90	10.79	T
V2	S1_218482222	3	67059263	0.09 A+D	3.81	-	-	3.81	7.05	G

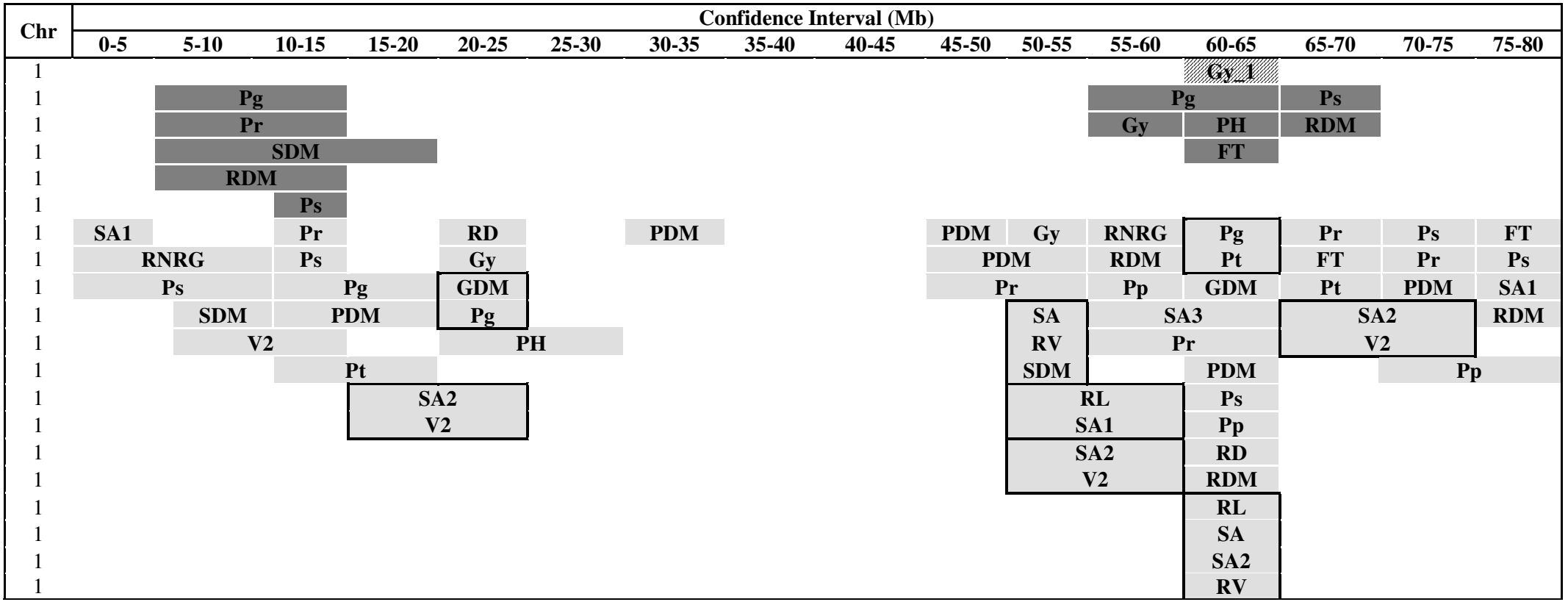
Trait	Marker	Chr	Pos (pb)	Additive		Dominance		Global -log10(*p*)	R^2 (%)	Favorable allele
				Effect	-log10(*p*)	Effect	-log10(*p*)			
V2	S1_223906763	3	72483804	0.02	0.01	0.17 *	3.79	3.10	6.96	G
V2	**S1_346109042**	5	52310727	0.04	1.49	0.18 *	4.48	4.75	10.48	C
V2	**S1_408281673**	6	52239753	0.03	0.13	0.20 *	4.79	4.07	9.06	C
V2	**S1_416720116**	6	60678196	0.09 *A+D	4.01	-	-	4.01	7.42	G
V2	**S1_647980048**	10	50673001	0.06	0.80	0.18 *	4.54	4.24	9.41	G

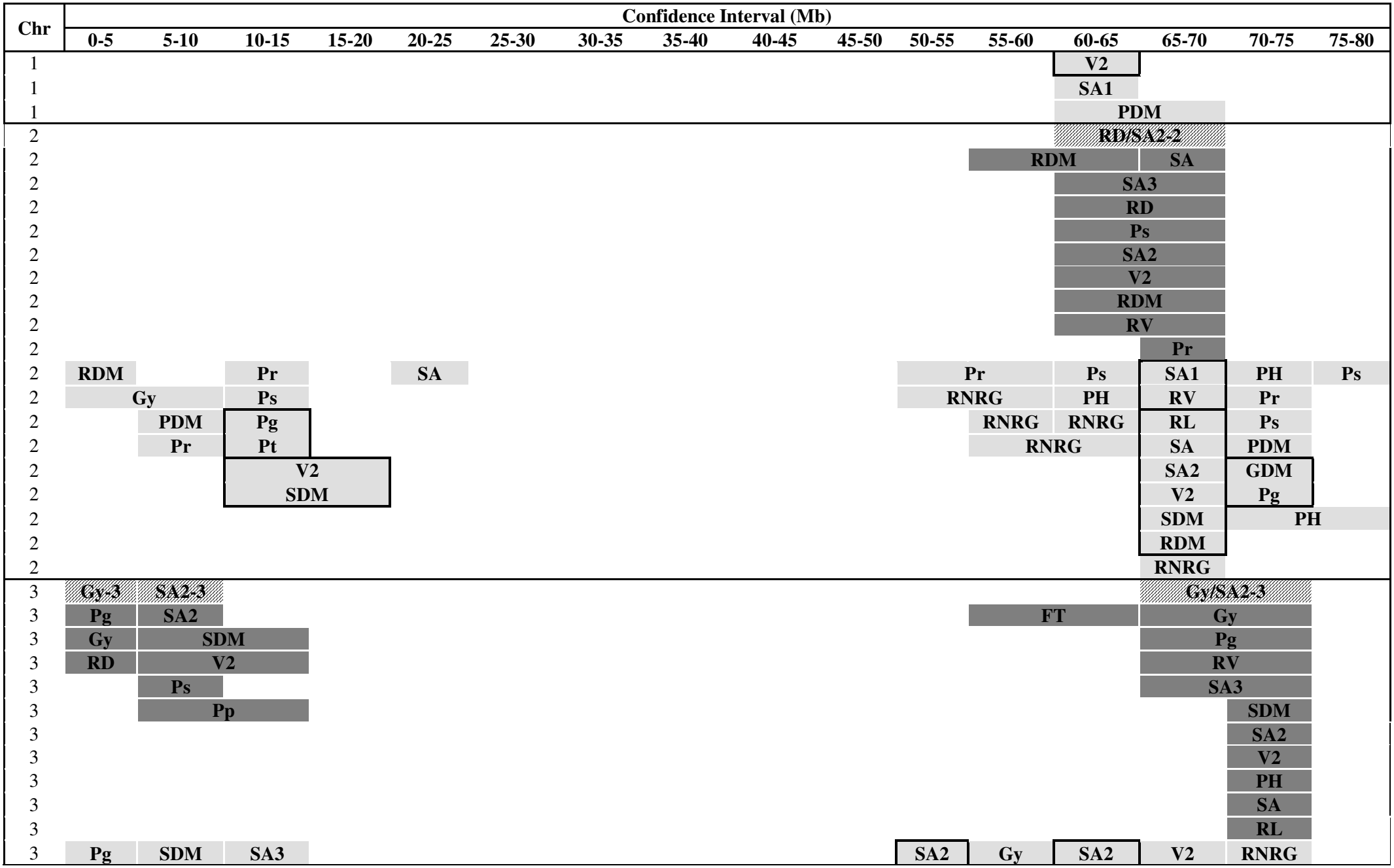
Chr: chromosome; Pos (pb): position in par bases; p: p-value; R^2: coefficient of determination. Gy: grain yield (ton ha-1); FT: flowering time (days); PH: plant height (cm); Pp: plant phosphorus content (leaves and stem - ton ha-1); Pg: grain phosphorus content (ton ha-1); Pt: total phosphorus content (ton ha-1); PDM: plant dry matter (ton ha-1); GDM: grain dry matter (ton ha-1); RL: root length (cm); RD: root diameter (mm); SA: total root surface area (cm^2); SA1: surface area of super fine roots (cm^2 - 0 mm < RD ≤ 1 mm); SA2: surface area of fine roots (cm^2 - 1 mm < RD ≤ 2 mm); SA3: surface area of thicker roots (cm^2 - 2 mm < RD ≤ 4.5 mm); RV: root volume (cm^3); V2: volume of fine roots (cm^3 - 1 mm < RD ≤ 2 mm); SDM: shoot dry matter (g); RDM: root dry matter (g); Ps: shoot phosphorus content (g); Pr: root phosphorus content (g); RNRG: relative net root growth. $^{A+D}$Indicates cases where additive and dominance effects are confounded due to the lack of the three possible genotypic classes. *Indicates significant effects ($p < 0.05$).In Bold are markers associated with multiple traits.*Indicates significant effects ($p < 0.05$). In Bold are markers associated with multiple traits.

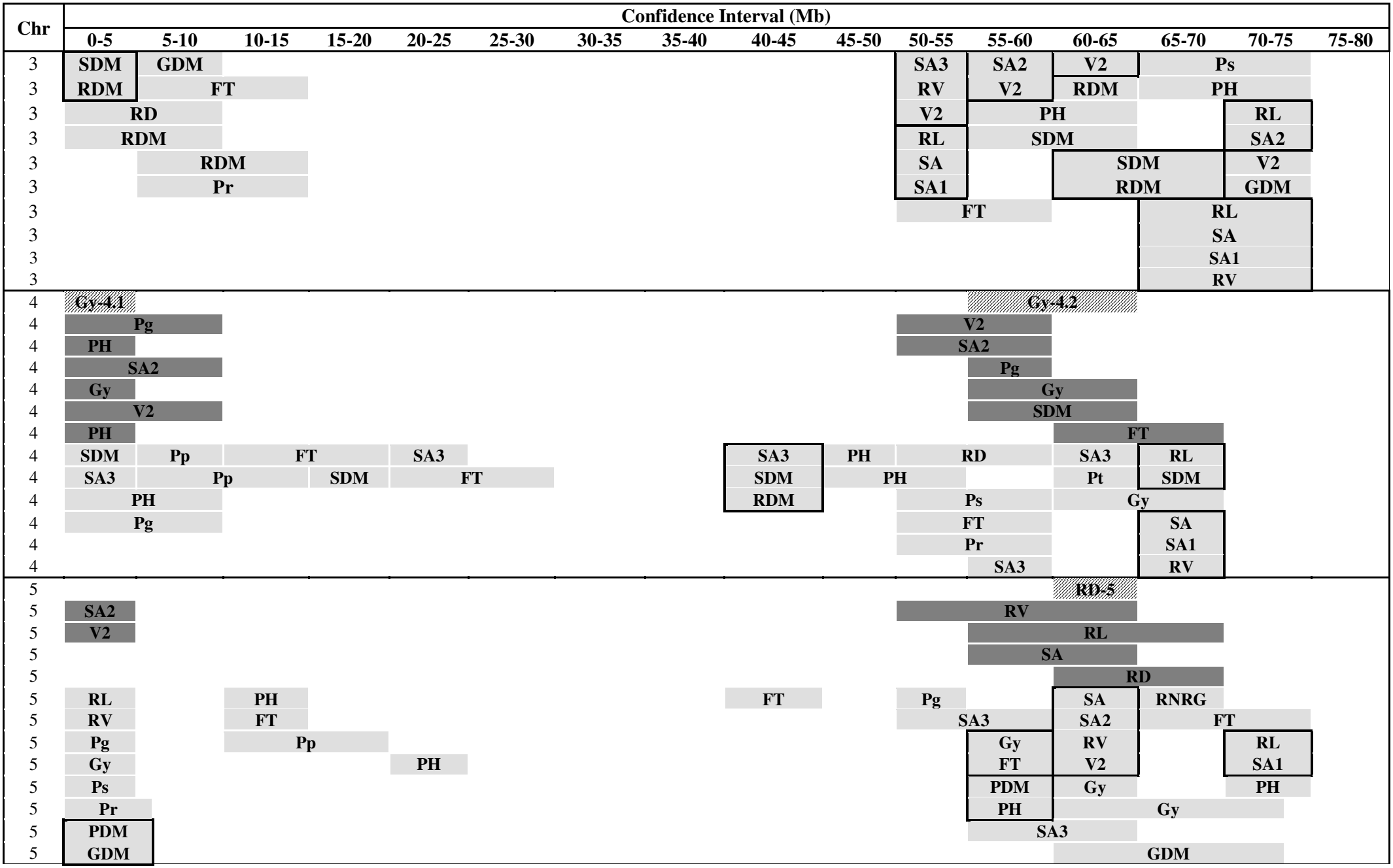
Table S6 Confidence interval for QTLs detected by single-trait and multi-trait mapping (Bernardino et al. 2019), and association mapping on BRP13R (3.1 version genome sorghum)

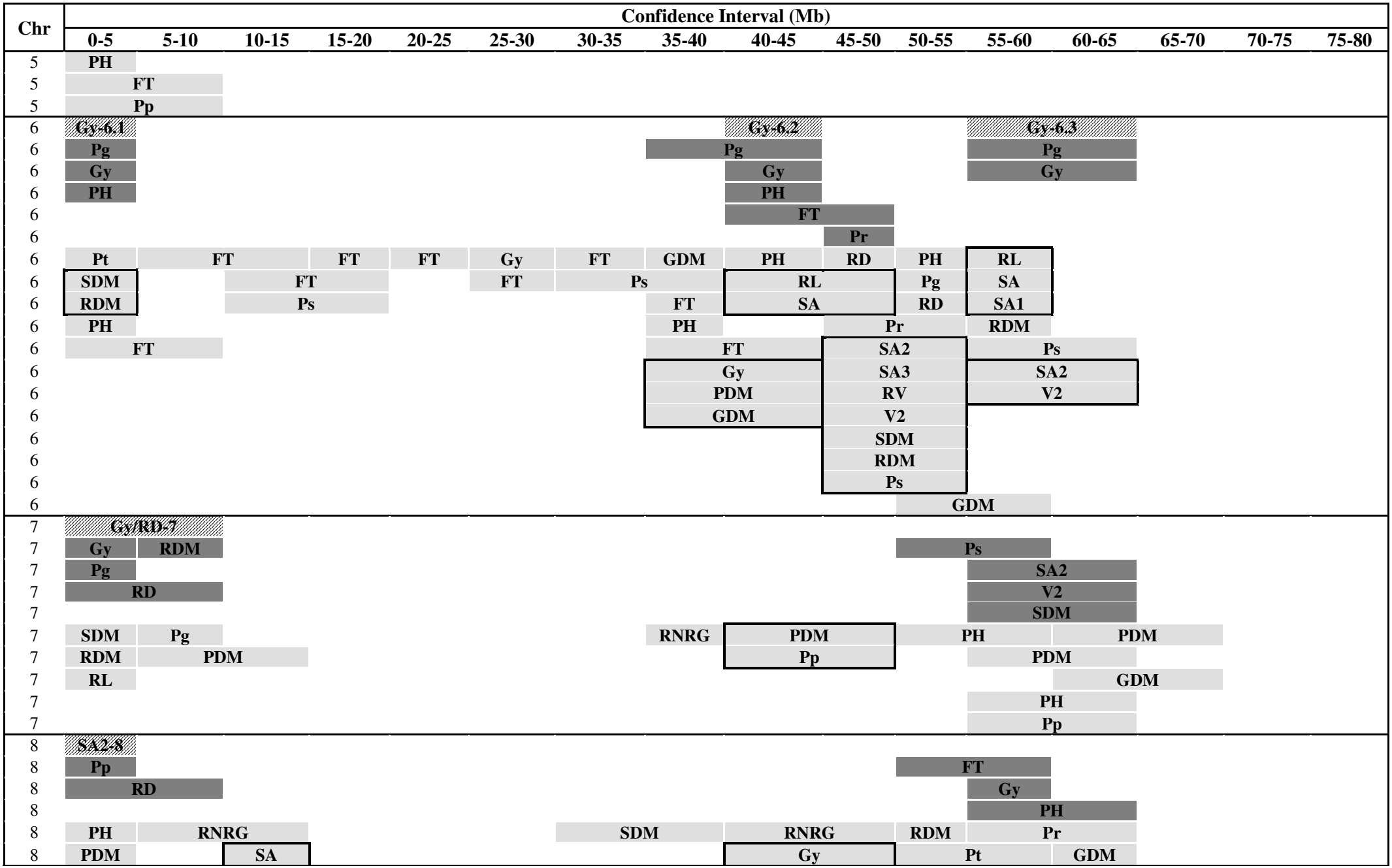
FIELD TRAITS		NUTRIENT SOLUTION TRAITS			
GDM	Grain dry matter (kg ha^{-1} or ton ha^{-1})	**RD**	Root diameter (mm)	**V2**	Volume of fine roots (cm³ - 1 mm < RD ≤ 2 mm)
Gy	Grain yield (kg ha^{-1} or ton ha^{-1})	**RL**	Root length (cm)	**SDM**	Shoot dry matter (g)
FT	Flowering time (days)	**RNRG**	Relative net root growth (%)	**RDM**	Root dry matter (g)
PH	Plant height (cm)	**SA**	Total root surface area (cm²)	**Ps**	Shoot phosphorus content (g)
PMD	Plant dry matter (kg ha^{-1} or ton ha^{-1})	**SA1**	Surface area of fine roots (cm² - 0 mm < RD ≤ 1 mm)	**Pr**	Root phosphorus content (g)
Pp	Plant P content (leaves-stem, kg ha^{-1} or ton ha^{-1})	**SA2**	Surface area of fine roots (cm² - 1 mm < RD ≤ 2 mm)		
Pg	Grain phosphorus content (kg ha^{-1} or ton ha^{-1})	**SA3**	Surface area of thick roots (cm² - 2 mm < RD ≤ 4.5 mm)		
Pt	Total phosphorus content (kg ha^{-1} or ton ha^{-1})	**RV**	Root volume (cm³)		

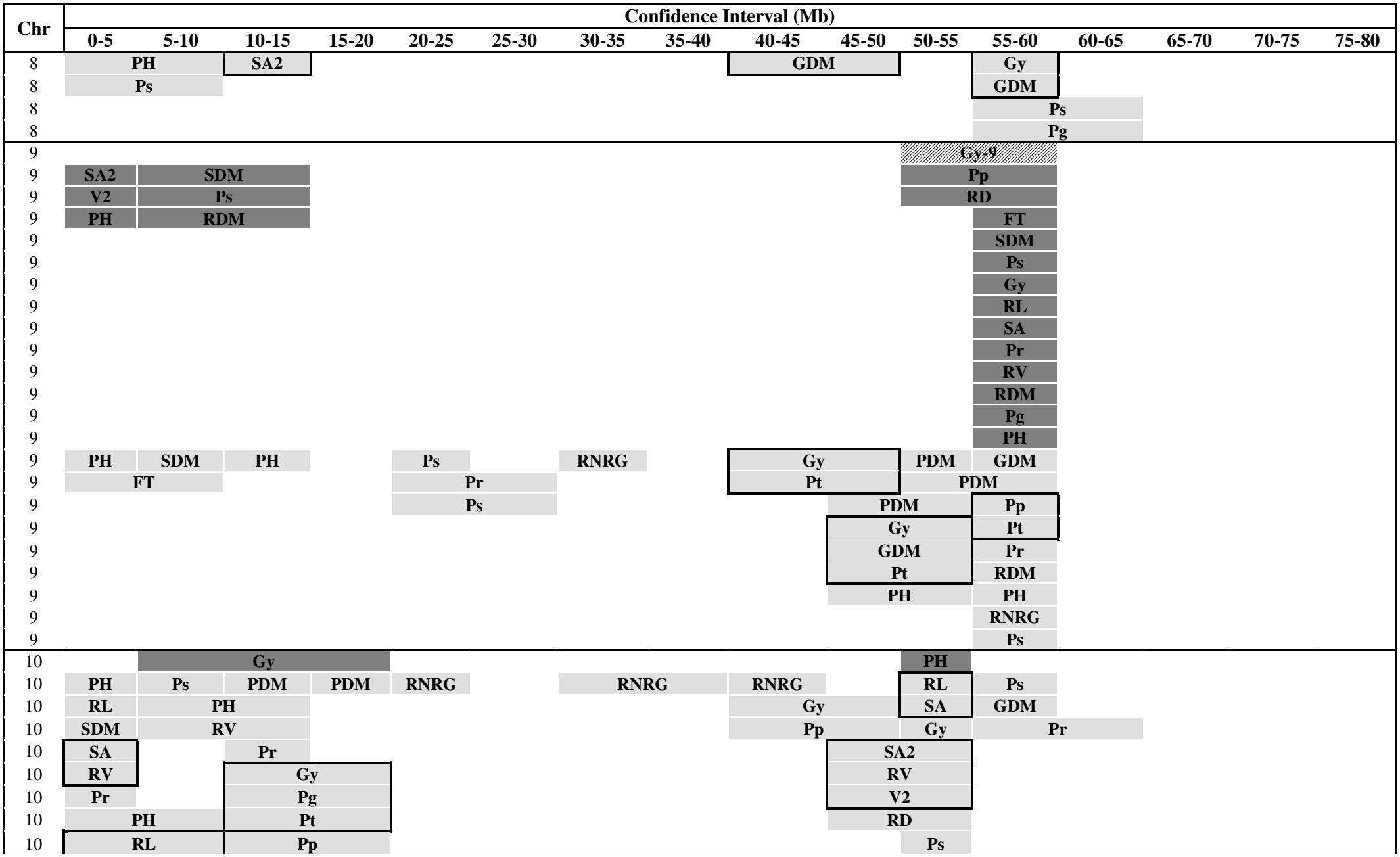
QTL (RILs) - Multi trait (Bernardino et al. 2019) **QTL (RILs) - Single trait** (Bernardino et al. 2019) **GWAS - BRP13R** **Same SNP for different traits - BRP13R**




Chr
Confidence Interval (Mb)
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
1
V2
SA1
PDM
2
RD/SA2-2
RDM
SA
SA3
RD
Ps
SA2
V2
RDM
RV
Pr
RDM
Pr
SA
Pr
Ps
SA1
PH
Ps
Gy
Ps
RNRG
PH
RV
Pr
PDM
Pg
RNRG
RNRG
RL
Ps
Pr
Pt
RNRG
SA
PDM
V2
SA2
GDM
SDM
V2
Pg
SDM
PH
RDM
RNRG
3
Gy-3
SA2-3
Gy/SA2-3
Pg
SA2
FT
Gy
Gy
SDM
Pg
RD
V2
RV
Ps
SA3
Pp
SDM
SA2
V2
PH
SA
RL
Pg
SDM
SA3
SA2
Gy
SA2
V2
RNRG


Chr
Confidence Interval (Mb)
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
3 SDM RDM GDM FT RD RDM RDM Pr SA3 RV V2 RL SA SA1 SA2 V2 PH SDM RDM FT Ps PH RL SA2 V2 GDM RL SA SA1 RV
4 Gy-4.1 Pg PH SA2 Gy V2 PH SDM Pp FT SA3 SA3 Pp SDM FT PH Pg SA3 PH SDM RDM PH RD Ps FT Pr SA3 Gy-4.2 V2 SA2 Pg Gy SDM FT SA3 PH RD SA3 RL Pt SDM Gy SA SA1 RV
5 SA2 V2 RL RV PH FT Pp PH RD-5 RV RL SA RD FT Pg SA3 SA Gy FT PDM PH SA2 RV V2 Gy Gy SA3 RNRG FT RL SA1 PH GDM


Chr
Confidence Interval (Mb)
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
5 PH
5 FT
5 Pp
6 Gy-6.1
6 Pg
6 Gy
6 PH
6 Pt
6 SDM
6 RDM
6 PH
6 FT
6 Gy
6 PDM
6 GDM
6 Gy-6.2
6 Pg
6 Gy
6 PH
6 FT
6 Pr
6 FT
6 FT
6 FT
6 FT
6 Ps
6 GDM
6 PH
6 RL
6 Pg
6 RD
6 FT
6 Ps
6 PH
6 FT
6 RL
6 SA
6 PH
6 SA2
6 SA3
6 RV
6 V2
6 SDM
6 RDM
6 Ps
6 GDM
6 Gy-6.3
6 Pg
6 Gy
6 RL
6 SA
6 SA1
6 RDM
6 Ps
6 SA2
6 V2
7 Gy/RD-7
7 Gy
7 RDM
7 Pg
7 RD
7 SDM
7 RDM
7 RL
7 Pg
7 PDM
7 RNRG
7 PDM
7 Pp
7 Ps
7 PH
7 PDM
7 SA2
7 V2
7 SDM
7 PDM
7 GDM
7 PH
7 Pp
8 SA2-8
8 Pp
8 RD
8 PH
8 PDM
8 RNRG
8 SA
8 SDM
8 RNRG
8 Gy
8 FT
8 RDM
8 Gy
8 PH
8 Pr
8 Pt
8 GDM


Chr
Confidence Interval (Mb)
0-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50
50-55
55-60
60-65
65-70
70-75
75-80
8 PH SA2 GDM Gy
8 Ps GDM
8 Ps
8 Pg
9 Gy-9
9 SA2 SDM Pp
9 V2 Ps RD
9 PH RDM FT
9 SDM
9 Ps
9 Gy
9 RL
9 SA
9 Pr
9 RV
9 RDM
9 Pg
9 PH
9 PH SDM PH Ps RNRG Gy PDM GDM
9 FT Pr Pt PDM
9 Ps PDM Pp
9 Gy Pt
9 GDM Pr
9 Pt RDM
9 PH PH
9 RNRG
9 Ps
10 Gy PH
10 PH Ps PDM PDM RNRG RNRG RNRG RL Ps
10 RL PH Gy SA GDM
10 SDM RV Pp Gy Pr
10 SA Pr SA2
10 RV Gy RV
10 Pr Pg V2
10 PH Pt RD
10 RL Pp Ps

Chr	Confidence Interval (Mb)															
	0-5	5-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60	60-65	65-70	70-75	75-80
10	SA										PH					
10	SA1															
10	RV															

Table S7 Mean prediction accuracy for different genomic selection models

Trait	Model[a]	Mean accuracy	Standard deviation
Gy	GBLUP-A	0.22	0.08
Gy	GBLUP-AD	0.24	0.08
Gy	GF-GBLUP-A	0.20	0.09
Gy	GF-GBLUP-AD	0.21	0.09
Gy	GWAS-GBLUP-A	0.24	0.10
Gy	GWAS-GBLUP-AD	0.28	0.11
PH	GBLUP-A	0.29	0.06
PH	GBLUP-AD	0.28	0.07
PH	GF-GBLUP-A	0.22	0.07
PH	GF-GBLUP-AD	0.18	0.07
PH	GWAS -GBLUP-A	0.40	0.06
PH	GWAS -GBLUP-AD	0.53	0.05
PDM	GBLUP-A	0.27	0.08
PDM	GBLUP-AD	0.28	0.08
PDM	GF-GBLUP-A	0.24	0.08
PDM	GF-GBLUP-AD	0.23	0.09
PDM	GWAS -GBLUP-A	0.28	0.09
PDM	GWAS -GBLUP-AD	0.26	0.09
RNRG	GBLUP-A	0.35	0.06
RNRG	GBLUP-AD	0.34	0.06
RNRG	GF-GBLUP-A	0.37	0.06
RNRG	GF-GBLUP-AD	0.33	0.08
RNRG	GWAS -GBLUP-A	0.42	0.07
RNRG	GWAS-GBLUP-AD	0.45	0.09

[a]Model GBLUP-A: GBLUP with random additive matrix; Model GBLUP-AD: GBLUP with random additive matrix and random dominance matrix; Model GF-GBLUP-A: GBLUP with random additive matrix and *SbPSTOL1* and *SbMATE* markers as fixed-effect covariates; Model GF-GBLUP-AD: GBLUP with random additive matrix and random dominance matrix, and *SbPSTOL1* and *SbMATE* genes as fixed-effect covariates; Model GWAS-GBLUP-A: GBLUP with random additive matrix and markers selected via GWAS as fixed-effect covariates; Model GWAS-GBLUP-AD: GBLUP with random additive matrix and random dominance matrix, and including GWAS markers as fixed-effect covariates. Gy: grain yield (ton ha^{-1}); PH: plant height (cm); PDM: plant dry matter (ton ha^{-1}); RNRG: relative net root growth.

CONCLUSÕES GERAIS

A eficiência na aquisição de fósforo é o principal componente na eficiência do uso de fósforo em sorgo; os genes *SbPSTOL1* podem ser utilizados como uma ferramenta no desenvolvimento de cultivares superiores em condições de baixo P; SNPs gene-específicos, como os estudados para os genes *SbPSTOL1*, podem ser utilizados na fase de pré-melhoramento, possibilitando, por exemplo, um screening inicial dos materiais do banco de germoplasma; populações como, a BRP13R, podem ser utilizadas como recurso para a descoberta de genes, os quais podem auxiliar no desenvolvimento de cultivares superiores para diferentes condições de estresse; as técnicas de mapeamento de QTLs, mapeamento associativo e seleção genômica ao serem empregadas de forma complementar geram resultados que, provavelmente, não seriam identificados se utilizadas isoladamente.