

# Multiple QTL mapping in autopolyploids: a random-effect model approach with application in a hexaploid sweetpotato full-sib population

Guilherme da Silva Pereira<sup>\*,†,1,2</sup>, Dorcus C. Gemenet<sup>§,1</sup>, Marcelo Mollinari<sup>\*,†</sup>, Bode A. Oluolu<sup>‡</sup>, Joshua C. Wood<sup>\*\*</sup>, Federico Diaz<sup>††</sup>, Veronica Mosquera<sup>††</sup>, Wolfgang J. Gruneberg<sup>††</sup>, Awais Khan<sup>§§</sup>, C. Robin Buell<sup>\*\*</sup>, G. Craig Yencho<sup>†</sup> and Zhao-Bang Zeng<sup>\*,†</sup>

<sup>\*</sup>Bioinformatics Research Center, North Carolina State University, Raleigh NC 27695, USA, <sup>†</sup>Department of Horticultural Science, North Carolina State University, Raleigh NC 27695, USA, <sup>§</sup>International Potato Center, ILRI Campus, Nairobi 25171-00603, Kenya, <sup>‡</sup>Department of Entomology and Plant Pathology, University of Tennessee, Knoxville TN 37996, USA, <sup>\*\*</sup>Department of Plant Biology, Michigan State University, East Lansing, MI 48824, USA, <sup>††</sup>International Potato Center, Lima 1558, Peru, <sup>§§</sup>Plant Pathology and Plant-Microbe Biology Section, Cornell University, Geneva NY 14456, USA

**ABSTRACT** In developing countries, the sweetpotato, *Ipomoea batatas* (L.) Lam. ( $2n = 6x = 90$ ), is an important autopolyploid species, both socially and economically. However, quantitative trait loci (QTL) mapping has remained limited due to its genetic complexity. Current fixed-effect models can only fit a single QTL and are generally hard to interpret. Here we report the use of a random-effect model approach to map multiple QTL based on score statistics in a sweetpotato bi-parental population ('Beauregard' × 'Tanzania') with 315 full-sibs. Phenotypic data were collected for eight yield component traits in six environments in Peru, and jointly predicted means were obtained using mixed-effect models. An integrated linkage map consisting of 30,684 markers distributed along 15 linkage groups (LGs) was used to obtain the genotype conditional probabilities of putative QTL at every cM position. Multiple interval mapping was performed using our R package QTLPOLY and detected a total of 41 QTL, ranging from one to ten QTL per trait. Some regions, such as those on LGs 3 and 15, were consistently detected among root number and yield traits and provided basis for candidate gene search. In addition, some QTL were found to affect commercial and noncommercial root traits distinctly. Further best linear unbiased predictions allowed us to characterize additive allele effects as well as to compute QTL-based breeding values for selection. Together with quantitative genotyping and its appropriate usage in linkage analyses, this QTL mapping methodology will facilitate the use of genomic tools in sweetpotato breeding as well as in other autopolyploids.

**KEYWORDS** Multiple interval mapping; Polyploid QTL model; Restricted maximum likelihood; Variance components; Yield components; Heritability

Genetic analyses in polyploid species pose extra challenges in comparison to diploid species, in spite of the evolutionary benefits that duplication of whole sets of chromosomes may have brought (Comai 2005; Van De Peer *et al.* 2009). When it comes to molecular markers, a co-dominant, biallelic single nucleotide polymorphism (SNP) directly informs on the genotypes of a diploid locus, but the best it can do alone in a polyploid locus is to inform on its allele dosage. In diploid species, molecular

markers are usually qualitatively scored, and there are several methodologies and tools for performing analysis on genetic linkage (e.g., Stam 1993; Margarido *et al.* 2007) and quantitative trait loci (QTL) mapping (e.g., Broman *et al.* 2003; Da Costa E Silva *et al.* 2012a). In allopolyploid species, such as cotton (Wu *et al.* 2015) and wheat (Hulse-Kemp *et al.* 2015), where preferential pairing dictates meiotic chromosome behavior much like diploids, existing approaches can be readily applied. However, despite many successful studies in diploids and allopolyploids, QTL mapping in autopolyploids remains difficult. In fact, unlike diploid mapping populations, which can have two to four segregating QTL genotypes (in case of inbred or outbred species, respectively), autopolyploid mapping populations can have a

Manuscript compiled: Sunday 28<sup>th</sup> April, 2019

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>Corresponding author: North Carolina State University, 1 Lampe Dr, Raleigh, NC 27695-7566. E-mail: [gdasilv@ncsu.edu](mailto:gdasilv@ncsu.edu)

much wider range of possible genotypes per locus. For example, there are up to 36, 400 or 4,900 possible genotypes from crosses between two tetra-, hexa- or octoploid outbred parents, respectively.

Single-dose markers, segregating in 1:1, 3:1 or 1:2:1 fashion, have limited information for building integrated genetic maps in autopolyploids, and can only be used for developing separate parental maps (Shirasawa *et al.* 2017) or rather than the desired integrated linkage groups (Balsalobre *et al.* 2017). In order to make use of multiple-dose markers, the first step is to perform dosage or quantitative SNP calling. Although most methods were designed for tetraploid species (e.g., Voorrips *et al.* 2011; Schmitz Carley *et al.* 2017), additional studies have tackled this problem, and methods including higher ploidy levels are now available (Serang *et al.* 2012; Gerard *et al.* 2018). For building integrated genetic maps, tetraploid species can use the well-established TETRAPLOIDSNPMAP (Hackett *et al.* 2016). For higher ploidy species, MAPPOLY (Mollinari and Garcia 2018) is a better option than POLYMAPR (Bourke *et al.* 2018), because the latter is limited to tetra- and hexaploid species, and lacks the ability to robustly map all multiple-dose markers based on hidden Markov models (HMM). With an integrated map, one can calculate the putative QTL genotype conditional probabilities, ideally using appropriate HMM (Hackett *et al.* 2016; Mollinari and Garcia 2018). Based on a model in Kempthorne (1955), an interval mapping (IM) method has been proposed as a first approach to map QTL in autotetraploids in a form of a regression weighted by the conditional probabilities (Hackett *et al.* 2001), which also turned out to be expanded for an autohexaploid species (van Geest *et al.* 2017).

For a general ploidy level  $m$ , this single-QTL model can be written as

$$y_i = \mu + \sum_{j=2}^m \alpha_j X_{ij} + \sum_{j=m+2}^{2m} \alpha_j X_{ij} + \varepsilon_i \quad (1)$$

where  $y_i$  is the phenotypic value of individual  $i$ ,  $\mu$  is the intercept,  $\alpha_j$  is the additive effect of allele  $j$ ,  $X_{ij}$  is the conditional probability of allele  $j$  in individual  $i$ , and  $\varepsilon_i$  is the residual error. The constraints  $\alpha_1 = 0$  and  $\alpha_{m+1} = 0$  are naturally imposed to satisfy the conditions  $\sum_{j=1}^m X_j = 2$  and  $\sum_{j=m+1}^{2m} X_j = 2$ , so that  $\mu$  is a constant hard to interpret due to these constraints (Hackett *et al.* 2001). Notice that  $2m - 2$  effects need to be estimated, i.e. tetra-, hexa- or octoploid models will have six, ten or 14 main effects, respectively. In order to answer the key question of whether the additive allele effects are different from zero (the null hypothesis), likelihood-ratio tests (LRT) are performed along positions on a genetic map. Commonly, the tests are presented as “logarithm of the odds” (LOD scores), where  $LOD = \frac{LRT}{2 \times \ln(10)}$ . In order to declare a QTL, empirical LOD thresholds are computed for each trait using permutations (Churchill and Doerge 1994). As the only current solution, this approach has been widely used so far (e.g., van Geest *et al.* 2017; Schumann *et al.* 2017; Massa *et al.* 2018). However, limitations in fitting multiple-QTL models have been presented mostly due to the possibility of over-parameterization or the lack of optimized algorithms for model selection (Mengist *et al.* 2018; Klaassen *et al.* 2019).

Variance component methods have been used for performing QTL mapping in related individuals of complex population structures or families in humans (Lippert *et al.* 2014), animals (Druet *et al.* 2008) and plants (Crepieux *et al.* 2005). In common, these approaches take into account the flexibility of mixed

models in dealing with the correlated QTL effects among individuals due to shared alleles identical-by-descent (IBD) by each relative pair at a particular location in the genome. For bi-parental populations of autopolyploids, although allele effects are usually regarded as fixed, the QTL genotype effects can be treated as random. Since a higher ploidy level leads to a much larger number of allele combinations, genotypic effects may be very hard to assess from the small population sample sizes usually available. In this case, the integrated genetic map provides key information on the inheritance of chromosomal segments from parents to progeny (Mollinari and Garcia 2018), making up the basis for IBD allele sharing estimations. If a locus is linked to a region underlying the variation of a trait of interest, higher IBD allele sharing for that locus is expected among individuals with similar phenotypic values (Almasy and Blangero 2010). Thus, the key parameter in this model are the variance components attributable to putative QTL, that determines the presence of linkage. Because only one parameter per QTL (the variance component) needs to be estimated, one could try to build a multiple-QTL model for polyploids, inspired by the corresponding multiple interval mapping (MIM) for diploid mapping populations (Kao *et al.* 1999), without the risk of model over-parameterization.

A multiple-QTL mapping approach may benefit several autopolyploid horticultural (e.g. potato, blueberry, kiwifruit, strawberry), ornamental (e.g., rose, chrysanthemum), forage (e.g., alfalfa, guinea grass) and field (e.g., sugarcane) crops. The sweetpotato [*Ipomoea batatas* (L.) Lam. ( $2n = 6x = 90$ )] is a staple food in several developing countries, with a production of 112 million tons worldwide in 2017 (FAO 2019). Particularly, it has attracted growing interest due to its characteristics for food and nutrition security (Mwanga *et al.* 2017). In addition to carbohydrates, dietary fiber, vitamins and minerals, orange-fleshed sweetpotatoes provide high levels of  $\beta$ -carotene to fight vitamin A deficiency in vulnerable populations, such as those in sub-Saharan Africa (Low *et al.* 2017). In order to increase production and meet farmer’s and market needs, it is imperative to make molecular-assisted selection an effective part of sweetpotato breeding programs. Toward this end, one of the first steps is characterizing the genetic architecture of traits of interest, such as those related to storage root yield and quality, and resistance to biotic and abiotic stresses (Khan *et al.* 2016). In spite of being considered an “orphan” crop, there have been recent advances in building genome references from its wild diploid relatives (Wu *et al.* 2018), optimizing a genotyping-by-sequencing protocol (GBSpoly) for high-throughput SNP genotyping (Wadl *et al.* 2018), and building a high-density integrated genetic map (Mollinari *et al.* 2019, in preparation). In this paper, we introduce a random-effect multiple interval mapping (REMIM) model for autopolyploids. Using a genome-assisted, GBSpoly-based integrated genetic map from a sweetpotato bi-parental population, we map QTL for yield-related traits with our open-source software, QTLPOLY.

## Materials and Methods

### Full-sib Population

A bi-parental mapping population (named BT) comprising 315 individuals was developed by crossing an orange-fleshed American variety, ‘Beauregard’ (CIP440132), and a non-orange-fleshed African landrace, ‘Tanzania’ (CIP440166), as male and female parents, respectively. The parents show contrasting phenotypes for several traits such as dry matter,  $\beta$ -carotene and sugar con-

155 tent, and susceptibility to biotic (e.g., virus disease) and abiotic  
 156 (e.g., drought) stresses. 'Beauregard' is known as to have higher  
 157 yield than 'Tanzania', and the current QTL mapping study will  
 158 focus on the yield components.

### 159 Phenotypic Analyses

160 **Field trials:** In addition to the 315 full-sibs, parents (each repli-  
 161 cated twice) and another variety, 'Daga' (CIP199062.1), were  
 162 used as checks in order to make up a total of 320 individuals  
 163 per replication in an  $80 \times 4$  alpha-lattice design. Virus-free plant-  
 164 ing material derived from tissue culture was obtained from the  
 165 CIP-Peru Genebank in La Molina. The clones were grown in a  
 166 screen house in CIP sub-station San Ramon, and the planting  
 167 material multiplied under low-disease pressure field conditions  
 168 in Satipo, where cuttings for the six experiments were obtained.  
 169 Four experiments were conducted in Ica ( $14^{\circ}01' S$  and  $75^{\circ}44' W$ ,  
 170 420 m), with two independent trials over two seasons, and one  
 171 experiment each was conducted in San Ramon ( $11^{\circ}07' S$  and  
 172  $75^{\circ}21' W$ , 828 m) and Pucallpa ( $8^{\circ}23' S$  and  $74^{\circ}31' W$ , 154 m). The  
 173 number of replications were two at Ica and three at San Ramon  
 174 and Pucallpa. In all trials, 1 m and 0.3 m of inter- and intra-row  
 175 spacing was used, respectively. In the first season at Ica (from  
 176 February 25 to June 29, 2016), the plot size was  $6 \text{ m}^2$  of 16 plants  
 177 arranged in four rows (four plants per row) with one empty row  
 178 between plots. In the second season at Ica (from November 15  
 179 2016 to March 17, 2017), the plot size was  $4.8 \text{ m}^2$  of 16 plants  
 180 arranged in two rows (eight plants per row) with no empty row  
 181 between plots. In San Ramon (from May 14 to September 15,  
 182 2016) and Pucallpa (from July 1 to November 4, 2016), the plot  
 183 size was  $9 \text{ m}^2$  of 30 plants arranged in three rows (ten plants per  
 184 row) with no empty row between plots.

185 **Phenotypic data:** Eight yield-related phenotypes were collected  
 186 per plot at harvest,  $\sim 120$  days after transplanting (see File S1).  
 187 For analysis purposes, foliage and root yield data were standard-  
 188 ized by plot size (relative to the largest) and converted to tons  
 189 per hectare ( $\text{t} \cdot \text{ha}^{-1}$ ) to allow comparisons across trials. Number  
 190 of roots was divided by the number of plants in the plot. The  
 191 total number of storage roots per plant (TNR) and total root  
 192 yield (RYTHA) considered all storage roots from the whole plot  
 193 regardless of their individual weight. Number of commercial  
 194 roots per plant (NOCR) and commercial root yield (CYTHA) con-  
 195 sidered only storage roots of marketable size ( $\geq 100 \text{ g}$  for African  
 196 market). Number of noncommercial roots per plant (NONC)  
 197 and noncommercial root weight (NCYTHA) were obtained from  
 198 the difference between total and commercial roots. Foliage yield  
 199 (FYTHA) was measured by weighing all above-ground biomass  
 200 per plot. Finally, commercial index (CI) was calculated as the  
 201 ratio between CYTHA and total biomass (i.e. the sum of RYTHA  
 202 and FYTHA).

**Multi-environment phenotypic model:** We considered each one  
 of the six field trials as an environment. Jointly predicted means  
 for each full-sib were obtained by using the following mixed-  
 effect model

$$y_{ijkl} = \mu + e_l + r_{k(l)} + b_{j(kl)} + t_i + te_{il} + \varepsilon_{ijkl}$$

where  $y_{ijkl}$  is the phenotype of the  $i^{\text{th}}$  treatment in the  $j^{\text{th}}$  block  
 within the  $k^{\text{th}}$  replicate at the  $l^{\text{th}}$  environment,  $\mu$  is the overall  
 mean,  $e_l$  is the fixed effect of the  $l^{\text{th}}$  environment ( $l = 1, \dots, L$ ;  
 $L = 6$ ),  $r_{k(l)}$  is the fixed effect of the  $k^{\text{th}}$  replicate ( $k = 1, \dots, K$ ;

$K = 2$  or  $3$  depending on the environment) at the  $l^{\text{th}}$  environ-  
 ment,  $b_{j(kl)}$  is the random effect of the  $j^{\text{th}}$  block ( $j = 1, \dots, J$ ;  
 $J = 80$ ) within the  $k^{\text{th}}$  replicate at the  $l^{\text{th}}$  environment with  
 $b_{j(kl)} \sim \mathcal{N}(0, \sigma_b^2)$ ,  $t_i$  is the effect of the  $i^{\text{th}}$  treatment ( $i = 1, \dots, I$ ;  
 $I = 318$ ),  $te_{il}$  is the effect of treatment by environment interac-  
 tion, and  $\varepsilon_{ijkl}$  is the random residual error with  $\varepsilon_{ijkl} \sim \mathcal{N}(0, \sigma^2)$ .  
 The treatment effect ( $t_i$ ) was separated into two groups, in which  
 $g_i$  is the random effect of the  $i^{\text{th}}$  genotype ( $i = 1, \dots, I_g$ ;  $I_g = 315$ )  
 with  $g_i \sim \mathcal{N}(0, \sigma_g^2)$ , and  $c_i$  is the fixed effect of the  $i^{\text{th}}$  check  
 ( $i = I_g + 1, \dots, I_g + I_c$ ;  $I_c = 3$ ). Similarly, treatment by environ-  
 ment interaction ( $te_{il}$ ) was separated into the random effect of  
 genotype by environment interaction ( $ge_{il}$ ) with  $ge_{il} \sim \mathcal{N}(0, \sigma_{ge}^2)$ ,  
 and the fixed effect of check by environment interaction ( $ce_{il}$ ).  
 We removed the check by environment interaction effect ( $ce_{il}$ )  
 from the model if Wald's test was not significant ( $p < 0.01$ ). Vari-  
 ance components were estimated by restricted maximum like-  
 likelihood (REML) using GENSTAT (v16; VSN International 2014).  
 Mean-basis broad-sense heritabilities ( $H^2$ ) were calculated as  
 the ratio between genotypic and phenotypic variances as

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{\bar{K}} + \frac{\sigma^2}{\bar{KL}}}$$

203 where  $\bar{K} = 2.25$  is the harmonic mean of the number of replicates  
 204 across environments. The R package PSYCH (v1.8.10; Revelle  
 205 2018) was used to calculate and plot Pearson's correlations (sig-  
 206 nificance  $*p < 0.05$ ,  $**p < 0.01$  and  $***p < 0.001$ ) among the  
 207 individual predicted means.

### 208 Genotypic Analyses

209 **GBSpoly and dosage calling:** A modified GBS protocol called  
 210 GBSpoly was carried out according to Wadl *et al.* (2018)  
 211 and described in detail for BT population by (Mollinari *et al.* 2019,  
 212 in preparation). In brief, total DNA was extracted and double  
 213 restricted using *CviAII-TseI* enzyme combination for all full-  
 214 sibs and parents (each parent replicated 10 times). Restriction  
 215 fragments were ligated to adapters, size selected and ampli-  
 216 fied. Adapters contained an 8-bp buffer sequence in addition to  
 217 sample-specific variable length barcodes (6-9 bp). Each 64-plex  
 218 library was sequenced using eight lanes of Illumina HiSeq 2500  
 219 system in order to ensure optimal read depth for dosage calling.  
 220 We trimmed the 8-bp buffer sequence from the reads using the  
 221 FASTX-TOOLKIT (available at [hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/)).  
 222 A modified version of TASSEL-GBS pipeline (v4.3.8), called  
 223 TASSEL4-POLY (Pereira *et al.* 2018, available at [https://github.com/](https://github.com/gramarga/tassel4-poly)  
 224 [gramarga/tassel4-poly](https://github.com/gramarga/tassel4-poly)) was used to demultiplex and to count  
 225 and store the actual read depth for all loci in variant call for-  
 226 mat (VCF) files (see File S2). We used BOWTIE2 (Langmead and  
 227 Salzberg 2013) to align 64-bp tags against the *I. trifida* and *I. triloba*  
 228 genomes, two sweetpotato wild relative diploid species (Wu  
 229 *et al.* 2018, available at <http://sweetpotato.plantbiology.msu.edu>).  
 230 Finally, the software SUPERMASSA (Serang *et al.* 2012, avail-  
 231 able at <https://bitbucket.org/orserang/supermassa>) was used to  
 232 perform multi-threading dosage call through a wrapper func-  
 233 tion named VCF2SM (Pereira *et al.* 2018, available at [https://](https://github.com/gramarga/vcf2sm)  
 234 [github.com/gramarga/vcf2sm](https://github.com/gramarga/vcf2sm)).

235 **Linkage mapping:** A linkage map was constructed by (Mollinari  
 236 *et al.* 2019, in preparation) using the R package MAPPOLY (Mol-  
 237 linari and Garcia 2018, available at [https://github.com/mmollina/](https://github.com/mmollina/mappoly)  
 238 [mappoly](https://github.com/mmollina/mappoly)) (see File S3). In brief, we computed two-point recom-  
 239 bination fractions between all 38,701 non-redundant, high qual-  
 240 ity GBSpoly-based markers, and sorted the most likely linkage

241 phase between each marker pair. Markers were then grouped  
 242 into 15 linkage groups (LGs) by using the Unweighted Pair  
 243 Group Method with Arithmetic Mean (UPGMA) hierarchical  
 244 clustering method. For each LG, markers were first ordered us-  
 245 ing multidimensional scaling as implemented in the R package  
 246 MDSMAP (Preedy and Hackett 2016), and then local order was  
 247 refined based on the reference genomes (Wu *et al.* 2018). Finally,  
 248 map distances were re-estimate individual posterior probabil-  
 249 ities from SUPERMASSA dosage calls. The final integrated,  
 250 completely phased map was composed of 30,684 markers dis-  
 251 tributed along 15 LGs with a total length of 2,708.4 centiMorgans  
 252 (cM) and no major gaps between markers (11.35 markers every  
 253 cM, on average). Multi-point genotype conditional probabilities  
 254 of putative QTL were estimated for every individual given the  
 255 final map using an HMM algorithm (Lander and Green 1987;  
 256 Jiang and Zeng 1997) adapted for polyploids (Mollinari *et al.*  
 257 2019, in preparation) as implemented in MAPPOLY. Since 17  
 258 full-sibs were filtered out along the map construction (Mollinari  
 259 *et al.* 2019, in preparation), only the remaining 298 individuals  
 260 were ultimately used for QTL mapping.

### 261 QTL Mapping Analyses

262 Under some fairly conventional assumptions (exclusive bivalent  
 263 formation, no preferential pairing, and no double reduction),  
 264 an autopolyploid individual of a species with an even ploidy  
 265 level  $m$  can produce up to  $\binom{m}{m/2}$  or “ $m$  choose  $m/2$ ” different  
 266 gametes with the same probability. As an example, consider two  
 267 contrasting parents, A and B, of a hexaploid species (such as  
 268 sweetpotato) and their respective genotypes for a QTL as  $abcdef$   
 269 and  $ghijkl$ , each one with potentially six different alleles. Under  
 270 the previous assumptions, each parent can produce up to 20  
 271 different gametes. Therefore, the cross A  $\times$  B would generate up  
 272 to 400 possible different genotypes. We will use the number of  
 273 gametes and genotypes of a hexaploid species to define vector  
 274 and matrix dimensions in the QTL mapping model from now  
 275 on. Obviously, the model can be easily adapted to any polyploid  
 276 species with an even ploidy level by simply changing these  
 277 dimensions accordingly.

**REMIM model and hypothesis testing:** Taking a full-sib popu-  
 lation with  $n$  individuals derived from a cross between two  
 hexaploid parents, A and B, the multiple-QTL mapping model  
 is expressed by

$$\mathbf{y} = \mathbf{1}\mu + \sum_{q=1}^Q \mathbf{Z}_q \boldsymbol{\gamma}_q + \boldsymbol{\varepsilon} \quad (2)$$

278 where  $\mathbf{y}$  is the  $n \times 1$  vector of phenotypic values,  $\mu$  is the fixed ef-  
 279 fect of population mean,  $\boldsymbol{\gamma}_q$  is the  $400 \times 1$  random vector of geno-  
 280 type effects of QTL  $q$  ( $q = 1, \dots, Q$ ) with  $\boldsymbol{\gamma}_q \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Pi}\sigma_q^2)$ , and  $\boldsymbol{\varepsilon}$   
 281 is the  $n \times 1$  random vector of residual error with  $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}\sigma^2)$ .  
 282  $\mathbf{1}$  and  $\mathbf{I}$  are an  $n \times 1$  vector of 1’s and an  $n \times n$  identity matrix,  
 283 respectively,  $\mathbf{Z}_q$  is the  $n \times 400$  incidence matrix of genotype con-  
 284 ditional probabilities of QTL  $q$ , and  $\boldsymbol{\Pi}$  is a  $400 \times 400$  matrix of  
 285 proportion of shared alleles IBD between the 400 possible geno-  
 286 types. These IBD allele sharing proportions range from zero  
 287 (no shared alleles) to one (six shared alleles) and relate to the  
 288 additive effects of within-parent alleles. For a full-sib progeny,  
 289 the IBD expected value is 0.5.

Assuming that the random-effect QTL are uncorrelated, each  
 with expectation zero, the expectation of the vector of pheno-  
 typic values  $\mathbf{y}$  is

$$E(\mathbf{y}) = \mathbf{1}\mu$$

and its variance-covariance matrix is

$$\text{Var}(\mathbf{y}) = \sum_{q=1}^Q \mathbf{G}_q \sigma_q^2 + \mathbf{I}\sigma^2$$

where  $\mathbf{G}_q = \mathbf{Z}_q \boldsymbol{\Pi} \mathbf{Z}_q'$  is the  $n \times n$  additive relationship matrix  
 between all  $n$  full-sibs on the putative QTL  $q$ . Here, our interest  
 is in testing

$$H_0 : \sigma_q^2 = 0 \text{ vs. } H_a : \sigma_q^2 > 0$$

i.e., whether QTL  $q$  contributes to the variation in  $\mathbf{y}$  or not, so  
 that several tests have to be performed along the genome. As  
 part of the algorithm described next, we test for the presence of  
 multiple QTL in consecutive rounds. In practice, we compute  
 and store a  $\mathbf{G}_q$  matrix for every putative QTL  $q$ , representing  
 genomic positions at a certain step size (e.g., every 1 cM). In this  
 case, the Model 2 can be rewritten as

$$\mathbf{y} = \mathbf{1}\mu + \sum_{q=1}^Q \mathbf{g}_q + \boldsymbol{\varepsilon} \quad (3)$$

290 where  $\mathbf{g}_q$  is an  $n \times 1$  random vector of the individual effects for  
 291 the QTL  $q$  with  $\mathbf{g}_q \sim \mathcal{N}(\mathbf{0}, \mathbf{G}_q \sigma_q^2)$ .

292 We compute linear score statistics according to Qu *et al.* (2013)  
 293 at every position and compare its  $p$ -value to a prescribed critical  
 294 value. The  $p$ -values are continuous over the unit interval as a  
 295 result of weighted sums of the scores from the profiled likelihood.  
 296 The test is exact (nonasymptotic) when there is only one QTL,  
 297 while a moment-based approximation to the null distribution  
 298 is used when two or more QTL are present in the model (Qu  
 299 *et al.* 2013). Herein, we conveniently take the “logarithm of  $p$ ”  
 300 as  $LOP = -\log_{10}(p)$  for graphic representation and supporting  
 301 interval calculation purposes. Support intervals are defined as  
 302 the QTL peak neighboring region with  $LOP$  greater than or equal  
 303  $LOP - d$ , where  $d$  is a constant which subtracts the highest  $LOP$   
 304 (thus from the QTL peak) in that region, as similarly proposed  
 305 for the statistic LOD scores (Lander and Green 1987).

306 **QTL detection and characterization:** In order to select QTL, we  
 307 adapted the MIM methodology described by Kao *et al.* (1999) to  
 308 a random-effect model framework as follows:

- 309 1. *Forward search* adds one QTL at a time to the model at the  
 310 position with the highest score statistic if the  $p$ -value is  
 311 smaller than a pointwise significance threshold level (e.g.,  
 312  $p < 0.01$ ), and fits it into the model. Consecutive rounds  
 313 are carried out conditioning the search of a new QTL to the  
 314 one(s) in the model until no more positions can reach the  
 315 threshold. A window size (e.g., of 15 cM) is avoided on  
 316 either side of QTL already in the model when searching for  
 317 a new QTL;
- 318 2. *Model optimization* follows rounds of position refinement  
 319 and backward elimination when no more QTL can be added  
 320 in the forward search step. In turns, a QTL position is up-  
 321 dated conditional to all the other QTL in the model, and its  
 322 score statistic is reevaluated at a more stringent significance  
 323 threshold level (e.g.,  $p < 10^{-3}$  or  $10^{-4}$ ), when the QTL may  
 324 be dropped. The final set of QTL is defined when all se-  
 325 lected positions are significant and, thus, no more positions  
 326 change or QTL are dropped;
- 327 3. *Forward search* (now with a threshold value as stringent  
 328 as the one used for backward elimination, e.g.,  $p < 10^{-3}$ )

or  $10^{-4}$ ) as well as model optimization procedures are repeated until no more QTL are added (via forward search) or dropped (via backward elimination). Finally, *QTL profiling* is performed with the remaining significant QTL after the last round of model optimization has been carried out. The score statistics and their associated  $p$ -values are computed for all genomic positions conditional to the final set of QTL.

Notice that, as part of the strategy for selecting QTL, we were less stringent during the first step of *forward search*, so that we were able to allow more positions to be tested again during *model optimization*. In fact, power for detecting significant positions is expected to increase when conditioning the forward search as well as the backward elimination to other QTL already in the model (Da Costa E Silva *et al.* 2012a). For the forward search performed after the first backward elimination, we used the last threshold set from the backward elimination in order to avoid false positives.

Once the QTL were selected, we were able to estimate their variance components and compute QTL heritabilities,  $h_q^2$ , as the ratio between the QTL variance component and total variance. Given the parameter estimates, QTL-based breeding values are directly obtained as the best linear unbiased predictions (BLUPs) of the QTL genotypes (i.e.,  $\hat{\gamma}_q$ ) from Model 2. BLUPs of the 400 possible genotypes were further decomposed in order to estimate the additive allele effects (i.e., 6 for each parent as  $\{a, \dots, f\}$  and  $\{g, \dots, l\}$ ) as well as the additive allele combination effects among two (i.e., 15 combinations for each parent as  $\{ab, \dots, cd\}$  and  $\{ef, \dots, kl\}$ ), and three (i.e., 20 combinations for each parent as  $\{abc, \dots, def\}$  and  $\{ghi, \dots, jkl\}$ ) alleles (Kempthorne 1955). Notice that in an  $F_1$  population, we can only study QTL that are different in alleles within the parents, not between. Also, due to the model assumptions of zero mean for random effects, allele and allele combination effects sum up to zero. These effects should be interpreted as the heritable contributions from parent to offspring, hence providing straightforward estimation of QTL-based breeding values to be used for selection.

### Simulations

We examined the performance of REMIM with 1,000 simulated quantitative traits with three QTL each. The QTL heritabilities were simulated as  $h_1^2 = 0.27$ ,  $h_2^2 = 0.18$  and  $h_3^2 = 0.09$  following their respective QTL genotype effect distributions as  $g_1 \sim \mathcal{N}(0, G_1 0.6)$ ,  $g_2 \sim \mathcal{N}(0, G_2 0.4)$  and  $g_3 \sim \mathcal{N}(0, G_3 0.2)$ . The environmental error was simulated from a standard normal distribution, i.e.  $\varepsilon \sim \mathcal{N}(0, I)$ , while the population mean was simulated as zero, i.e.  $\mu = 0$ . The QTL were randomly assigned to the BT population linkage map ( $n = 298$ ), but no closer than 15 cM (our window size) from each other. One round of *forward search* followed by *model optimization* (steps 1 and 2 from the algorithm described above) was carried out for each simulated trait with combinations of different forward (0.01, 0.02 and 0.05) and backward ( $10^{-2}$ ,  $10^{-3}$ ,  $10^{-4}$  and  $10^{-5}$ ) pointwise significance  $p$ -value thresholds. For comparison, we ran the fixed-effect interval mapping (FEIM, Model 1) with the same simulated traits using different genome-wide significance levels of (0.20, 0.15, 0.10 and 0.05) based on 1,000 permutations as LOD thresholds to declare significant QTLs (Churchill and Doerge 1994). The same step size of 2 cM was used in both approaches.  $LOP - d$  (from REMIM) and  $LOD - d$  (from FEIM) support intervals were calculated for three different  $d$  values (1.0, 1.5 and 2.0).

Following the definitions and summary statistics from Da Costa E Silva *et al.* (2012b), all QTL kept after the model

optimization were considered “mapped”. A mapped QTL was considered “paired” (true QTL) if less than 15 cM apart from the simulated position, and a paired QTL was considered “matched” if included within a support interval of a mapped QTL. Finally, a mapped QTL was considered “mismatched” (false QTL) if it was not matched. We summarized detection power as the ratio between the number of paired QTL over the total number of simulated QTL, and the absolute distance differences between simulated and mapped positions were averaged out. Genome-wide type I error or false discovery rate (FDR) was estimated for each support interval as the ratio between the number of mismatched QTL over the total number of mapped QTL. Finally, the proportion of matched QTL (coverage) as an approximation of support intervals was provided for each  $d$  value.

### Software implementation

We implemented the algorithm for detection and characterization of multiple QTL based on REMIM model in an R package called QTLPOLY (available at <https://github.com/guilherme-pereira/qtlpoly>). We integrated functions from the R package VARCOMP (v0.2-0; Qu *et al.* 2013) to compute the score statistics. The rounds of QTL search and model optimization use the variance components estimated in the previous round, so that the new estimates iterate faster. In addition, calculations for different genomic positions were paralleled in order to speed up the process by using the R base package PARALLEL (v3.5.2; R Core Team 2018). Final models were fitted using the R package SOMMER (v3.6; Covarrubias-Pazarán 2016), from which BLUPs were extracted and used for estimation of allele effects and QTL-based breeding values. Both VARCOMP and SOMMER packages use REML estimation to compute the variance components from the random-effect QTL model. Functions for plotting QTL profiles, effects and support intervals were based on GG-PLOT2 (v3.1.0; Wickham 2016). Additional functions for running FEIM model and multi-threaded permutations were included in QTLPOLY and were based on the `lm()` function from R base package STATS (v3.5.2; R Core Team 2018).

### Gene expression profiling

A developmental time-course of expression profiling data of ‘Beauregard’ was reported previously (Wu *et al.* 2018) and used with a parallel time-series of development with ‘Tanzania’ roots (Gemenet *et al.* 2019, submitted). In brief, ‘Beauregard’ and ‘Tanzania’ roots were harvested from four biological replicates at 10, 20, 30, 40 and 50 days after transplanting (DAT), and classified at 30, 40, and 50 DAT into fibrous and storage roots based on diameter as described by Wu *et al.* (2018). RNA and RNA-sequencing (RNA-seq) from ‘Tanzania’ were generated in parallel with the ‘Beauregard’ samples as described in Wu *et al.* (2018). All reads were cleaned, aligned to the *I. trifida* genome (Wu *et al.* 2018), and fragments per kilobase exon model per million mapped reads (FPKM) determined as described previously in Lau *et al.* (2018) with the one exception that the ‘Tanzania’ 30 DAT storage root sample was sub-sampled for 30 million reads. To provide a comparison of expression abundances in the roots to leaves, ‘Beauregard’ and ‘Tanzania’ plants were grown as described in Lau *et al.* (2018) for control conditions and RNA-seq libraries from leaves processed as described above. For the final FPKM matrix, genes encoded by the chloroplast were removed.

### Data Availability

Raw sequence reads are available at NCBI, under BioProject numbers PRJNAXXXXXX (DNA), and PRJNA491292 and PRJ-

450 NAXXXXXX (RNA) [to be released upon publication]. The expres-  
 451 sion abundance matrix is available at Dryad Digital Repository  
 452 [to be released upon publication]. Remaining supplemental files  
 453 are available at FigShare [to be released upon publication]. File S1  
 454 contains phenotypic data. File S2 contains VCF files. File S3  
 455 contains genetic map information. QTLPOLY software used for  
 456 QTL mapping analyses and simulations is available at GitHub  
 457 (<https://github.com/guilherme-pereira/qtlpoly>).

## 458 Results

### 459 Trait heritabilities and correlations

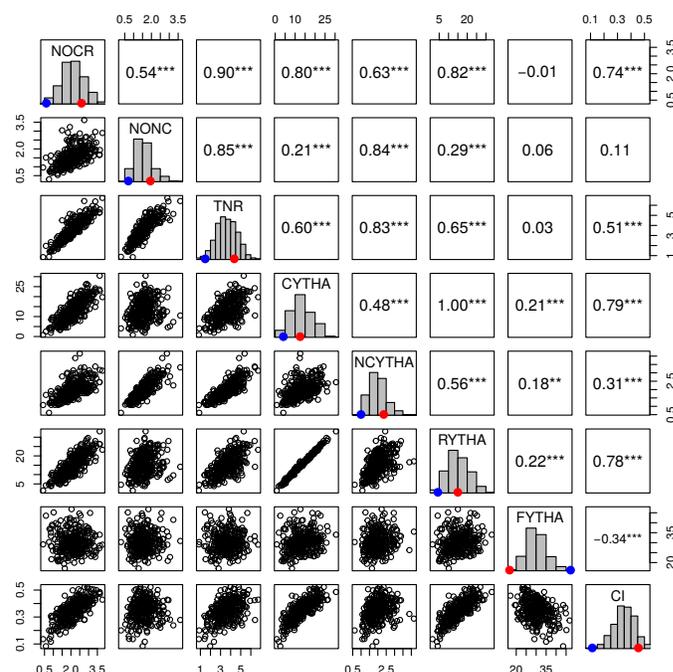
460 Each one of the eight yield-related traits from six environments  
 461 were analyzed using a multi-environment mixed-effect model,  
 462 from which we were able to obtain jointly predicted means for  
 463 each full-sib and variance component estimates (Table 1). Par-  
 464 ents showed contrasting means for all traits, with 'Beauregard'  
 465 presenting higher means for number of roots and root yield  
 466 (both commercial and non-commercial) and commercial index  
 467 when compared to 'Tanzania', which surpassed 'Beauregard'  
 468 only for foliage yield. Interestingly, transgressive segregation  
 469 was observed among the full-sibs for all traits, with emphasis  
 470 on several individuals with CYTHA higher than the most produc-  
 471 tive parent. Broad-sense heritabilities ranged from intermediate  
 472 (55.01% for FYTHA) to high values (80.50% for CI). Correla-  
 473 tions between the predicted means were estimated (Figure 1).  
 474 Low correlations (from 0.18\*\* to 0.22\*\*\*) were observed between  
 475 FYTHA and root yield traits. The highest correlation (0.99\*\*\*)  
 476 was between CYTHA and RYTHA, which was expected, since  
 477 most RYTHA is derived from CYTHA. Among the traits used  
 478 for CI calculation, CYTHA also had the highest correlation with  
 479 CI (0.79\*\*\*), likely because it is its main component. TNR com-  
 480 ponents were also highly correlated with TNR, namely NOCR  
 481 (0.90\*\*\*) and NONC (0.86\*\*\*). Finally, NOCR and NONC turned  
 482 out to be highly correlated with CYTHA (0.80\*\*\*) and NCYTHA  
 483 (0.84\*\*\*), respectively.

### 484 Mapping QTL in BT population

485 **Simulations:** The BT linkage map based on 298 F<sub>1</sub> progenies  
 486 was used to simulate 1,000 quantitative traits with three QTL  
 487 each. We ran FEIM (Model 1) and REMIM (Model 3) for each  
 488 simulated trait in order to assess their detection power and FDR  
 489 in such scenario (see Tables S1 and S2). For REMIM, different  
 490 forward *p*-value thresholds did not impact power or FDR (re-  
 491 sults not shown), but backward thresholds were critical. Figure  
 492 2 compares different threshold criteria for declaring a QTL dur-  
 493 ing FEIM and REMIM (for *p* < 0.01 forward threshold). From  
 494 both approaches, FDR greater than 20% were found when us-  
 495 ing less conservative criteria of genome-wide significance LOD  
 496 threshold of 0.20 for FEIM and *p* < 10<sup>-3</sup> backward threshold  
 497 for REMIM. However, detection power was higher with REMIM  
 498 (75.2%, on average) than with FEIM (65.6%, on average). Tak-  
 499 ing more conservative criteria such as genome-wide significance  
 500 LOD threshold of 0.05 for FEIM and *p* < 10<sup>-4</sup> backward thresh-  
 501 old for REMIM, FDR decreased to 14.3% and 13.0%, respectively,  
 502 whereas power was still higher with REMIM (67.3%, on average)  
 503 than with FEIM (59.9%, on average). In fact, although the QTL  
 504 with the highest heritability (*h*<sub>1</sub><sup>2</sup> = 0.27) was similarly detected  
 505 regardless of the method and criterion, a higher proportion of  
 506 QTL with intermediate and low heritabilities were mapped un-  
 507 der the multiple-QTL mapping approach. Interestingly, even  
 508 with the most stringent criteria of *p* < 10<sup>-5</sup> backward threshold

509 for REMIM, we were able to map as many QTL as using genome-  
 510 wide significance LOD threshold of 0.05 for FEIM, but with a  
 511 better FDR control (~10%). In general, the average absolute dif-  
 512 ference between the simulated and mapped QTL peak location  
 513 did not differ when comparing models or thresholds whatsoever  
 514 (see Table S1). From testing different *d* values for LOD - *d* and  
 515 LOP - *d*, we learned that *d* = 1.5 was a good approximation of  
 516 95% support interval for both FEIM and REMIM (see Table S2).

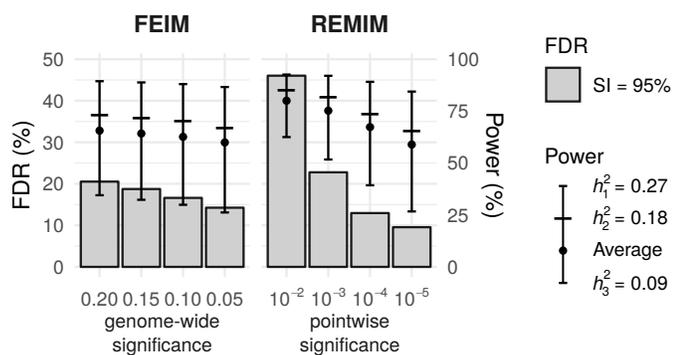
517 **Yield-related traits:** We adopted 0.01 and 10<sup>-3</sup> as the respec-  
 518 tive forward and backward *p*-value thresholds for detecting  
 519 QTL in eight yield-related traits in the BT population using  
 520 REMIM (Model 3; Figure 3, Table 2). In total, 41 QTL were iden-  
 521 tified, with *p*-values ranging from 1.64 × 10<sup>-9</sup> (QTL 2 for TNR)  
 522 to 7.20 × 10<sup>-4</sup> (QTL 1 for RYTHA). The number of QTL per trait  
 523 ranged from one (CYTHA and CI each) to ten (TNR). NOCR,  
 524 NONC, NCYTHA, RYTHA and FYTHA had seven, nine, five,  
 525 three and five QTL, respectively. All LGs except LG 5 harboured  
 526 QTL regions. LGs 4 and 15 harboured six QTL each, and LGs  
 527 10 and 13 harboured four QTL each. Approximate 95% support  
 528 intervals computed as LOP - 1.5 (see Figure S1) showed that  
 529 QTL hotspots may be found on LGs 1, 3, 4 and 15 as several  
 530 QTL regions seemed to be co-localized. On LG 1, QTL peaks for  
 531 NOCR, TNR and NCYTHA can be found between 131.50 and  
 532 140.03 cM. On LG 3, QTL peaks for NONC, NOCR and TNR  
 533 were localized either at 12.36 or 20.18 cM. LG 4 seemed to have  
 534 a second QTL hotspot: a first region with QTL peaks between  
 535 32.06 and 76.48 cM for NONC, NOCR, TNR and FYTHA, in



**Figure 1** Pearson's correlations (\*\**p* < 0.01, \*\*\**p* < 0.001) among predicted means of eight yield-related traits from 'Beauregard' × 'Tanzania' (BT) full-sib family. Parental means are represented by red (B) and blue (T) dots. Trait abbreviations: number of commercial (NOCR) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in t · ha<sup>-1</sup>, foliage yield (FYTHA) in t · ha<sup>-1</sup>, and commercial index (CI).

**Table 1** Phenotypic analysis summary of eight yield-related traits from ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family. Parental ( $\bar{B}$  and  $\bar{T}$ ) and progeny ( $\bar{F}_1$ ) means, minimum and maximum  $F_1$  means, and genetic ( $\sigma_g^2$ ), genotype-by-environment interaction ( $\sigma_{ge}^2$ ) and residual ( $\sigma^2$ ) variance components and heritability ( $H^2$ ) estimates are shown for eight traits: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).

	NOCR	NONC	TNR	CYTHA	NCYTHA	RYTHA	FYTHA	CI
$\bar{B}$	2.751	1.876	4.622	12.230	2.384	15.070	11.270	0.478
$\bar{T}$	0.697	0.796	1.448	3.720	0.971	4.680	50.240	0.108
$\bar{F}_1$	2.121	1.577	3.624	13.201	2.006	15.275	30.204	0.350
min( $F_1$ )	0.421	0.292	0.850	0.780	0.613	1.430	17.070	0.083
max( $F_1$ )	3.786	3.619	6.730	30.480	4.155	33.130	46.810	0.523
$\sigma_g^2$	0.387	0.277	1.117	27.620	0.313	31.580	23.940	$5.880 \times 10^{-3}$
$\sigma_{ge}^2$	0.277	0.215	0.581	17.330	0.317	18.870	36.040	$2.352 \times 10^{-3}$
$\sigma^2$	0.678	0.552	1.446	31.640	1.064	34.420	48.090	$5.120 \times 10^{-3}$
$H^2$ (%)	69.05	67.00	75.35	73.33	58.76	74.28	55.01	80.50



**Figure 2** Detection power and false discovery rate (FDR) from QTL mapping analyses of 1,000 simulated traits in ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family. Each trait was simulated with three QTL with different heritabilities ( $h_q^2 = \{0.27, 0.18, 0.09\}$ ), randomly positioned on the BT linkage map ( $n = 298$ ). Fixed-effect interval mapping (FEIM) used different genome-wide significance LOD thresholds (0.20, 0.15, 0.10, 0.05) based on 1,000 permutation tests. Random-effect multiple interval mapping (REMIM) was performed using different score-based  $p$ -value thresholds during backward elimination ( $p < \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ ) after forward search using  $p < 0.01$ . Power (vertical lines, right axis) represents the proportion of true QTL over the total number of simulated QTL. FDR (bars, left axis) depicts the proportion of false QTL over the total number of mapped QTL for a  $\sim 95\%$  support interval (SI) coverage.

536 addition to a second region with another QTL for TNR and one  
 537 for NCYTHA at 200.09 and 180.36 cM, respectively. Similarly,  
 538 on LG 15, a first hotspot had QTL peaks either at 4.19 or 5.27  
 539 cM for CYTHA, RYTHA and CI, whereas a second hotspot had  
 540 QTL peaks at 109.10 and 105.02 cM for TNR and NCYTHA,  
 541 respectively.

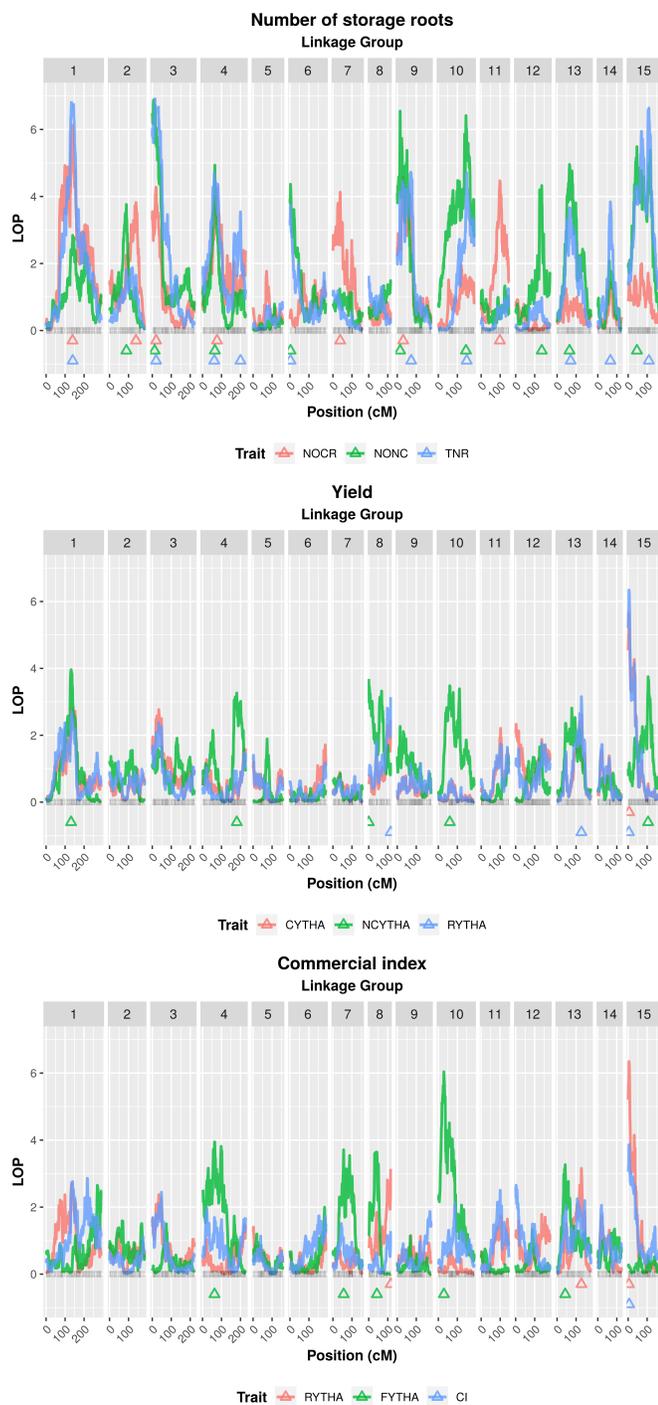
542 QTL variances ( $\sigma_q^2$ ) and heritabilities ( $h_q^2$ ) estimates from  
 543 Model 2 are shown in Table 2, where the subscript  $q$  denotes the

544 QTL number for a specific trait. QTL heritabilities ranged from  
 545 3.68 (QTL 4 for TNR) to 22.69% (QTL 2 for TNR), representing  
 546 the proportion of the total variance explained by that QTL, condi-  
 547 tional to all the other QTL in the model. Out of 41 QTL, 14 were  
 548 considered major QTL ( $h_q^2 > 10\%$ ), with one major QTL for TNR  
 549 ( $h_2^2 = 22.69\%$ ), CYTHA ( $h_1^2 = 15.53\%$ ), and CI ( $h_1^2 = 20.65\%$ ),  
 550 two major QTL for NONC ( $h_2^2 = 10.22\%$  and  $h_5^2 = 10.47\%$ ),  
 551 NOCR ( $h_1^2 = 10.84\%$  and  $h_3^2 = 16.53\%$ ), NCYTHA ( $h_1^2 = 10.28\%$   
 552 and  $h_5^2 = 12.97\%$ ), and RYTHA ( $h_2^2 = 10.80\%$  and  $h_3^2 = 20.13\%$ ),  
 553 and three major QTL for FYTHA ( $h_1^2 = 10.65\%$ ,  $h_4^2 = 13.48\%$   
 554 and  $h_5^2 = 10.40\%$ ). Altogether, mapped QTL explained as much  
 555 as 76.64%, 68.49% and 78.66% of the total variance for NONC,  
 556 NOCR and TNR, respectively. As less QTL were identified for  
 557 root yield traits, a relatively smaller portion of the total variance  
 558 was explained for NCYTHA (48.99%), RYTHA (38.77%) and  
 559 FYTHA (51.22%). Interestingly, most of the major QTL lies on  
 560 hotspots, such as those on the beginning of LGs 3 and 15 (see  
 561 Figure S2). In order to compare QTL detection results, we used a  
 562 rather relaxed genome-wide significance of 0.20 for FEIM, whose  
 563 permutation-based LOD score thresholds ranged from 6.81 to  
 564 6.89 depending on the trait ( $LOD \approx 6.85$ , on average) (see Figure  
 565 S4). A total of 17 QTL were mapped (see Table S3): one for each  
 566 CYTHA, RYTHA and CI, two for each NOCR, NCYTHA and  
 567 FYTHA, and four for each NONC and TNR. Six LGs harboured  
 568 QTL: LGs 15 and 3 had six and four QTL, respectively, LG 1  
 569 had three QTL, LG 10 had two QTL each, and LGs 4 and 9 had  
 570 only one QTL each. In fact, the most significant QTL ( $LOD > 9$ )  
 571 were found on LGs 1, 3 and 15, similar to REMIM results. In  
 572 addition to QTL not detected on several other LGs, FEIM did  
 573 not detect any QTL for number of roots on LG 4, identified as a  
 574 QTL hotspot by REMIM, although for FYTHA a QTL reached  
 575 the threshold at 95.01 cM ( $LOD = 6.91$ ). On the other hand,  
 576 REMIM missed a QTL for NCYTHA on LG 3 at 32.56 cM, which  
 577 was detected using FEIM. LOD thresholds for a genome-wide  
 578 significance of 0.05 ranged from 7.68 to 7.98 ( $LOD \approx 7.77$ , on  
 579 average), and one would have mapped 11 QTL, instead. One  
 580 QTL of each NONC (LG 1), TNR (LG 9), NCYTHA (LG 3) and  
 581 RYTHA (LG 15), and both QTL for FYTHA (LGs 4 and 10) would

**Table 2 Random-effect multiple interval mapping (REMIM) of yield-related traits from 'Beauregard' × 'Tanzania' (BT) full-sib family. Linkage group (LG), map position (in centiMorgans) and its ~95% support interval (SI), score statistic and its corresponding  $p$ -value, variance ( $\sigma_q^2$ ) and heritability ( $h_q^2$  in percentage) of mapped QTL.**

Trait <sup>a</sup>	QTL	LG	Position (SI)	Score	$p$ -value	$\sigma_q^2$	$h_q^2$ (%)
NOCR	1	1	139.24 (99.06-152.48)	229.86	$7.40 \times 10^{-7}$	0.0696	10.84
	2	2	139.30 (98.83-152.50)	162.91	$1.52 \times 10^{-4}$	0.0620	9.65
	3	3	20.18 (0.00-37.44)	164.37	$5.27 \times 10^{-5}$	0.1061	16.53
	4	4	76.48 (57.01-100.14)	161.69	$5.67 \times 10^{-5}$	0.0437	6.80
	5	7	38.27 (2.18-99.12)	161.55	$7.40 \times 10^{-5}$	0.0578	9.00
	6	9	31.15 (22.24-73.45)	177.01	$2.65 \times 10^{-5}$	0.0497	7.74
	7	11	98.39 (83.07-123.22)	177.88	$3.40 \times 10^{-5}$	0.0508	7.92
NONC	1	2	88.62 (72.44-95.72)	144.97	$1.73 \times 10^{-4}$	0.0353	7.56
	2	3	12.36 (0.00-21.01)	247.82	$5.06 \times 10^{-8}$	0.0477	10.22
	3	4	65.09 (54.35-70.05)	172.49	$1.16 \times 10^{-5}$	0.0379	8.11
	4	6	3.12 (0.00-25.13)	156.65	$4.30 \times 10^{-5}$	0.0333	7.14
	5	9	15.11 (13.48-51.17)	229.78	$2.85 \times 10^{-7}$	0.0489	10.47
	6	10	145.40 (137.16-164.81)	226.77	$3.86 \times 10^{-7}$	0.0420	9.01
	7	12	137.16 (125.03-147.31)	152.11	$4.71 \times 10^{-5}$	0.0349	7.47
	8	13	65.36 (50.68-98.41)	183.08	$1.10 \times 10^{-5}$	0.0408	8.74
	9	15	46.19 (32.15-119.04)	207.76	$3.26 \times 10^{-6}$	0.0368	7.90
TNR	1	1	140.03 (130.13-148.23)	258.64	$2.72 \times 10^{-8}$	0.1769	8.88
	2	3	20.18 (18.43-25.28)	295.97	$1.64 \times 10^{-9}$	0.4520	22.69
	3	4	62.03 (43.43-86.11)	165.39	$1.98 \times 10^{-5}$	0.1507	7.57
	4	4	200.09 (164.01-205.15)	128.48	$2.85 \times 10^{-4}$	0.0733	3.68
	5	6	3.12 (0.00-14.33)	134.07	$1.85 \times 10^{-4}$	0.0919	4.61
	6	9	73.45 (13.48-82.1)	170.46	$2.01 \times 10^{-5}$	0.1298	6.52
	7	10	149.31 (144.46-185.31)	164.53	$1.98 \times 10^{-5}$	0.1250	6.27
	8	13	71.22 (44.23-120.57)	133.47	$2.34 \times 10^{-4}$	0.1140	5.72
	9	14	66.19 (59.27-79.27)	137.79	$1.45 \times 10^{-4}$	0.1164	5.84
	10	15	109.10 (65.09-118)	237.56	$2.29 \times 10^{-7}$	0.1370	6.88
CYTHA	1	15	5.27 (0.00-32.15)	258.61	$2.26 \times 10^{-6}$	6.7290	20.65
NCYTHA	1	1	131.5 (113.05-152.48)	166.05	$1.09 \times 10^{-4}$	0.0399	10.28
	2	4	180.36 (43.43-215.18)	137.92	$5.43 \times 10^{-4}$	0.0310	8.00
	3	8	0.00 (0.00-76.43)	156.90	$2.07 \times 10^{-4}$	0.0304	7.83
	4	10	60.24 (35.15-114.14)	148.63	$3.30 \times 10^{-4}$	0.0385	9.92
	5	15	105.02 (75.28-125.30)	161.27	$1.77 \times 10^{-4}$	0.0503	12.97
RYTHA	1	8	115.54 (40.19-115.54)	140.37	$7.20 \times 10^{-4}$	3.3344	7.84
	2	13	128.02 (90.20-145.54)	141.40	$6.98 \times 10^{-4}$	4.5906	10.80
	3	15	4.19 (0.00-8.02)	275.67	$4.53 \times 10^{-7}$	8.5576	20.13
FYTHA	1	4	63.08 (0.00-112.02)	164.77	$1.14 \times 10^{-4}$	3.9400	10.65
	2	7	56.40 (34.12-109.48)	159.38	$1.94 \times 10^{-4}$	3.4774	9.40
	3	8	43.00 (27.20-54.33)	159.45	$2.27 \times 10^{-4}$	2.6997	7.29
	4	10	29.09 (18.40-37.32)	249.37	$9.15 \times 10^{-7}$	4.9884	13.48
	5	13	42.60 (31.67-68.18)	140.27	$5.41 \times 10^{-4}$	3.8502	10.40
CI	1	15	5.27 (0.00-36.02)	182.71	$1.36 \times 10^{-4}$	0.0010	15.53

<sup>a</sup> Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).



**Figure 3** Logarithm of  $p$ -value ( $LOP$ ) profiles from random-effect multiple interval mapping (REMIM) of eight yield-related traits from 'Beauregard'  $\times$  'Tanzania' (BT) full-sib family. Triangles show the QTL peak location. Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).

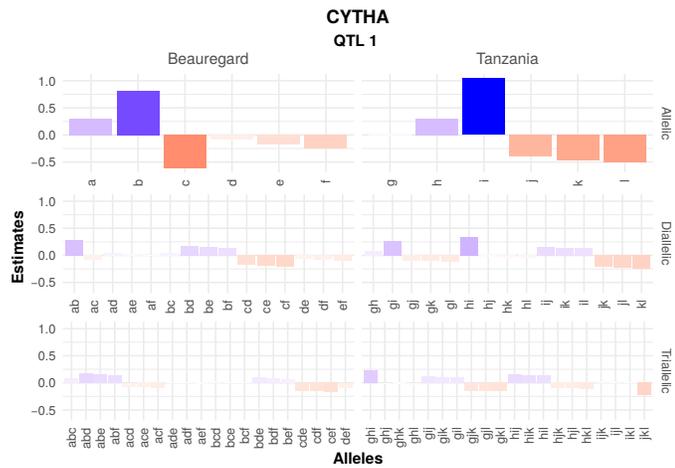
582 have been missed under this more conservative criterion.

583 From REMIM, allele and allele combination additive effects  
 584 of each QTL (see Table S4) were derived from the QTL geno-  
 585 type BLUPs for each trait (Model 2). These effects represent the  
 586 parental contribution to the population mean, i.e. how much  
 587 one adds to or subtracts from the mean given one of the 400 pos-  
 588 sible genotypes. For instance, Figure 4 shows the allele-specific  
 589 and allele combination additive effects of QTL 1 for CYTHA.  
 590 Inferences on which alleles contribute more to the mean as well  
 591 as which ones ought to be selected for breeding purposes are  
 592 straightforward. For example, individuals with the haplotypes  
 593  $b$  from 'Beauregard' and  $i$  from 'Tanzania', and without the hap-  
 594 lotypes  $c$  and  $j$  through  $l$  from the respective parents will have  
 595 the highest QTL-based breeding value estimates for CYTHA.  
 596 In order to allow comparison among the allele effects from 41  
 597 QTL, proportional contribution of a specific allele was calculat-  
 598 ed as the ratio between its absolute effect and the sum of all  
 599 12 absolute effects for each QTL, so that effects would range  
 600 within the unit interval. Out of 492 effects, 75 (15.24%) showed  
 601 a contribution of at least 15% (almost double of the average).  
 602 Among these most important effects, 'Beauregard' provided 34  
 603 effects, with 15 negative and 19 positive effects summing up  
 604 to  $-2.833$  and  $+3.908$ , while 'Tanzania' contributed with 41 ef-  
 605 fects, with 22 negative and 19 positive effects summing up to  
 606  $-4.308$  and  $+3.594$ . Although most QTL had approximately  
 607 50%:50% allele effect contribution from each respective parent,  
 608 we observed skewed contributions towards 'Beauregard' (from  
 609 61%:39% to 72%:28%) for seven QTL, all related with number  
 610 of roots, and towards 'Tanzania' (from 40%:60% to 30%:70%)  
 611 for eight QTL, with four related to number of roots and four  
 612 to yield traits. By computing QTL-based breeding values we  
 613 could hypothesize on the genetic basis of trait correlation. For  
 614 example, the single locus (QTL 1) detected for CI has shown ad-  
 615 ditive effects in the same direction as those of QTL 1 for CYTHA.  
 616 Therefore, correlation between QTL-based breeding values from  
 617 these two traits were indeed as high as  $0.97^{***}$  (see Figure S3).  
 618 In addition, other important breeding lessons may be learned  
 619 from these correlations. For instance, very low, non-significant  
 620 correlation between CYTHA and NCYTHA ( $0.11$ ) indicates that  
 621 improving the former will not affect the latter. An increase of  
 622 NOCR would bring a relative increase of NONC, though, as  
 623 their QTL-based breeding values showed higher, significant cor-  
 624 relation ( $0.46^{***}$ ). More interestingly, QTL-based breeding values  
 625 for FYTHA did not seem to correlate to any other root-related  
 626 trait. Finally, the absolute positive correlation of  $1.00^{***}$  between  
 627 predicted means from CYTHA and RYTHA (Figure 1) could be  
 628 only partially explained by a single co-localized QTL, since  
 629 the correlation between QTL-based breeding values was smaller  
 630 ( $0.73^{***}$ ), although still high.

### 631 **Candidate genes underlying QTL hotspots**

632 This study identified numerous QTL corresponding to yield-  
 633 related traits and identification of causal genes within all QTL  
 634 will require further research efforts due to the large number  
 635 of genes that are involved in storage root formation and yield  
 636 underlying these QTL. As we have access to an expression pro-  
 637 filing dataset for 'Beauregard' and 'Tanzania' roots, we elected  
 638 to examine putative candidate genes under two QTL that had  
 639 high heritability and effect sizes: the QTL for TNR on LG 3 (co-  
 640 localized for NOCR and NONC) and the QTL for CYTHA on  
 641 LG 15 (co-localized for RYTHA and CI).

642 The QTL peak for TNR on LG 3 was at 1,591,872 bp with a SI



**Figure 4** Allele and allele combination additive effects from the decomposed best linear unbiased predictions (BLUPs) for the QTL 1 (on linkage group 15 at 5.27 cM) of commercial root yield in  $t \cdot ha^{-1}$  (CYTHA) in a hexaploid sweetpotato full-sib family ('Beauregard'  $\times$  'Tanzania'). Marker-assisted selection for increasing CYTHA would have to focus on selection for alleles *b* and *i*, and against alleles *c* and *j* through *l* from the respective parents.

643 between 1,428,660 and 1,971,958 bp (see Table S5) that contains  
 644 75 genes. Examination of functional annotation of these 75 genes,  
 645 coupled with expression profiles in leaves as well as a time  
 646 course of developing roots in both 'Beauregard' and 'Tanzania'  
 647 (Gemenet *et al.* 2019, submitted) revealed three candidate genes  
 648 of interest (see Figure S5). The *I. trifida* *itf03g02930* gene encodes  
 649 a homolog of SKU5, a glycosyl phosphatidylinositol modified  
 650 protein in *Arabidopsis thaliana* with similarity to multiple-copper  
 651 oxidases that are localized to the plasma membrane and cell  
 652 wall. In sweetpotato, the *itf03g02930* homolog was expressed  
 653 in leaves and roots, although the expression in roots is substan-  
 654 tially higher than that in leaves. A second candidate gene is  
 655 *itf03g03280* which encodes a protein with sequence similarity  
 656 to annexin. The *itf03g03280* homolog shares 71% identity (85%  
 657 similarity) with 99% coverage with ANN2 (AT5G65020) and  
 658 66% identity (81% similarity) over 100% coverage with ANN1  
 659 (AT1G35720). *itf03g03280* was expressed in both leaves and roots  
 660 (total roots, fibrous roots, and storage roots) although expression  
 661 in roots is nearly twice that of leaves. The last candidate gene,  
 662 *itf03g03460*, encodes a protein with 51% identity (62% similar-  
 663 ity) to the WUSCHEL homeobox family protein (AtWOX13).  
 664 While *itf03g03460* was lowly expressed in leaves, it was highly  
 665 expressed in roots of both 'Beauregard' and 'Tanzania'.

666 On LG 15, a major QTL for CYTHA with the peak at 477,772  
 667 bp spanned positions from 21,822 to 1,939,509 bp (see Table S5)  
 668 and 310 genes. As this was too large of a distance to manually  
 669 curate candidate genes responsible for the trait, we restricted  
 670 our query to 25 genes distal and proximal to the most signifi-  
 671 cant marker. Within this region, two genes encoded functions  
 672 that may be associated with storage root development and had  
 673 expression profiles that supported a role in storage root devel-  
 674 opment (see Figure S5). The hormone ethylene has diverse roles  
 675 in cell proliferation and elongation, and the *I. trifida* *itf15g01020*  
 676 gene encodes a protein with similarity to the *A. thaliana* *CON-*  
 677 *STITUTIVE TRIPLE RESPONSE 1* gene (*CTR1*) with functions in  
 678 the ethylene signaling pathway. In sweetpotato, the *itf15g01020*

679 homolog was expressed in leaves but expressed at twice the  
 680 levels in developing roots. Storage roots are grown for their high  
 681 starch content, and *itf15g01120* encodes a protein with similarity  
 682 to starch branching enzyme 2.2, involved in starch biosynthesis.  
 683 *itf15g01120* was expressed in leaves and roots with the highest  
 684 expression levels detected in storage, not fibrous or developing  
 685 roots.

## 686 Discussion

687 Most of the linkage and QTL mapping work done for sweet-  
 688 potato so far has relied on strategies based on a double pseudo-  
 689 testcross approach for diploid species (Grattapaglia and Sederoff  
 690 1994). For example, separate parental maps have been built  
 691 based on this diploid-based simplification, using qualitative  
 692 marker systems such as randomly amplified polymorphic DNA  
 693 (RAPD; Ukoskit and Thompson 1997), amplified fragment  
 694 length polymorphism (AFLP; Kriegner *et al.* 2003; Cervantes-  
 695 Flores *et al.* 2008a; Nakayama *et al.* 2012), retrotransposon inser-  
 696 tion polymorphisms (Monden *et al.* 2015) and simple sequence  
 697 repeats (SSR; Kim *et al.* 2017). A recent map was developed from  
 698 a selfing population and used only single-dose SNPs, resulting  
 699 in higher marker saturation in comparison to the previous maps  
 700 (Shirasawa *et al.* 2017), though the map was still not integrated.  
 701 In some of these cases, QTL mapping analyses were performed  
 702 for several traits, mostly related to quality (Cervantes-Flores  
 703 *et al.* 2011; Zhao *et al.* 2013; Yu *et al.* 2014; Kim *et al.* 2017) and  
 704 resistance to biotic stresses (Cervantes-Flores *et al.* 2008b; Yada  
 705 *et al.* 2017a). For yield-related traits, only two studies have been  
 706 reported to date (Chang *et al.* 2009; Li *et al.* 2014). The use of  
 707 DNA markers with unknown DNA sequence limited our ability  
 708 to compare their results with *I. trifida* and *I. triloba* genomes (Wu  
 709 *et al.* 2018), and ultimately with our present QTL study (see Table  
 710 S5). Moreover, although these diploid-based strategies were the  
 711 state-of-the-art at that time for qualitative marker-based, low  
 712 density genetic maps, they imposed significant restrictions on  
 713 statistical power for QTL detection and its genetic interpretation.  
 714 Recently, more improved methods and computational tools that  
 715 take into account autopolyploid complexity for dosage SNP call-  
 716 ing (Voorrips *et al.* 2011; Serang *et al.* 2012; Schmitz Carley *et al.*  
 717 2017; Gerard *et al.* 2018) and integrated linkage map construc-  
 718 tion (Hackett *et al.* 2016; Bourke *et al.* 2018; Mollinari and Garcia  
 719 2018) have become available, mostly dedicated to tetraploids.  
 720 Taking advantage of the newly developed MAPPOLY package,  
 721 (Mollinari *et al.* 2019, in preparation) built the first integrated  
 722 genetic map for sweetpotato, from the BT population used here.  
 723 For a hexaploid species, this has opened up new opportunities  
 724 for more interpretable QTL genetic models due to MAPPOLY im-  
 725 plementation of a HMM that delivers QTL genotype conditional  
 726 probabilities along a fully integrated genetic map (Mollinari and  
 727 Garcia 2018).

728 More specifically, QTL mapping in autopolyploid species has  
 729 been limited to a fixed-effect interval mapping (FEIM) model  
 730 proposed for tetraploids (Hackett *et al.* 2001) and eventually  
 731 expanded for hexaploids (van Geest *et al.* 2017). Consisting of a  
 732 single-QTL model,  $2m - 2$  main effects are fitted ( $m$  is the ploidy  
 733 level), and this model is compared to a null model (with no QTL)  
 734 using LRT, ultimately expressed as LOD scores. Permutation-  
 735 based genome-wide significance LOD thresholds are then used  
 736 to declare a QTL. Trying to add more QTL into FEIM could  
 737 rapidly lead the model to over-parameterization, since each QTL  
 738 requires as much as six (for tetraploids), ten (for hexaploids) or  
 739 14 (for octoploids) parameters to be estimated. Furthermore,

740 new rounds of permutation, based on a model with QTL, would  
741 need to be carried out in order to provide an updated LOD  
742 score threshold (Klaassen *et al.* 2019). In contrast, the random-  
743 effect multiple interval mapping (REMIM) model presented  
744 here is designed to fit multiple random-effect QTL by estimating  
745 only one single parameter ( $\sigma_q^2$ ) per QTL. Score statistic tests are  
746 performed in order to assess whether a QTL variance component  
747 is zero or not, conditional to other QTL in the model. These tests  
748 provide an approach for comparing two nested models with  
749 the reduced model having a random effect excluded, just like  
750 residual LRT (RLRT) would do. However, (R)LRT is more prone  
751 to numerical errors because the null hypothesis ( $H_0 : \sigma_q^2 = 0$ )  
752 falls on the boundary of the parameter space, whereas score-  
753 based methods can be robust to eventual misspecification of the  
754 distribution of random effects (Verbeke and Molenberghs 2003).

755 We used the BT population genetic map to simulate quantita-  
756 tive traits in order to compare FEIM and REMIM performances,  
757 and also to assess the impact of using different thresholds for  
758 QTL detection (Figure 2). According to our simulations, both ap-  
759 proaches would detect similar number of simulated QTL at more  
760 stringent criteria, with REMIM delivering less false positives.  
761 The results also suggested that one could use a more relaxed  
762 criteria in order to increase the power of detection while still  
763 maintaining an acceptable level of FDR. Although conclusions  
764 may be limited to the simulated scenario, multiple-QTL model  
765 approaches have been proven to provide greater power and  
766 better FDR control than single-QTL models for both univariate  
767 (Zeng *et al.* 1999; Laurie *et al.* 2014) and multivariate models  
768 (Da Costa E Silva *et al.* 2012b), mostly due to the differences  
769 in detecting QTL with smaller effects. In fact, this is rather expected  
770 as a multiple-QTL model has a smaller residual variance which  
771 helps to detect additional QTL. Multiple-QTL models are also  
772 supposed to improve detection of more than one QTL on the  
773 same LG (Mayer 2005), as they are usually hard to separate from  
774 each other due to the high extension of linkage disequilibrium in  
775 mapping populations. For polyploids, a non-optimal approach  
776 of using residuals from a fitted single-QTL model as phenotypic  
777 data to find a second linked QTL has been proposed (Mengist  
778 *et al.* 2018). In QTL mapping analysis, it is important to have a  
779 reasonable balance between detection power and FDR, as we are  
780 interested in mapping as many true QTL as possible. It should  
781 also consider the goals of the study, i.e. whether it is intended  
782 to use a few very reliable QTL for marker-assisted breeding,  
783 or to discover as many QTL-related putative genes as possible  
784 for further validation. Additional suggestive QTL also increase  
785 the number of hypothesized regions that affect trait variation  
786 and may be targeted for selection. For yield-related traits in  
787 our BT population, FEIM results were limited to 15 QTL (see  
788 Figure S4), most of which also happened to be mapped using  
789 REMIM (Figure 3). In fact, by performing REMIM, we found  
790 27 minor QTL ( $3.68\% \leq h_q^2 \leq 9.92\%$ ) in addition to 14 major  
791 ones ( $10.22\% \leq h_q^2 \leq 22.69\%$ ) (Table 2, see Figure S2). Based on  
792 double pseudo-testcross approach, previous estimates of propor-  
793 tion of variance explained (PVE) by nine QTL for storage root  
794 yield ranged from 17.7 to 59.3% for 202 individuals from a cross  
795 between two Chinese sweetpotato varieties (Li *et al.* 2014). In  
796 another study, analyses of two reciprocal full-sib populations  
797 with less than 120 individuals each detected seven QTL for root  
798 and top (foliage) weight ( $16.0\% \leq \text{PVE} \leq 29.5\%$ ), and only one  
799 QTL detected for root number ( $\text{PVE} = 14.8\%$ ) (Chang *et al.* 2009).  
800 Because of likely estimation bias due to reduced sample size and  
801 the use of not very informative markers and linkage maps, these

802 previous PVE findings are hard to compare with our results. Ad-  
803 justed  $R^2$  from FEIM accounted for  $7.8\% \leq \text{PVE} \leq 12.6\%$  (see  
804 Table 3), which may not be comparable with  $h_q^2$  from REMIM  
805 due to different approaches (single vs. multiple QTL models).

806 In general, the number of mapped QTL was compatible  
807 with the proportion of total variance explained by the QTL  
808 ( $h_{\text{total}}^2$ ) (Table 2). From seven to ten QTL ( $68.48\% \leq h_{\text{total}}^2 \leq$   
809  $78.67\%$ ) were mapped for traits related with number of roots  
810 ( $67.00\% \leq H^2 \leq 75.35\%$ ), whereas from one to five QTL  
811 ( $20.65\% \leq h_{\text{total}}^2 \leq 48.99\%$ ) were found underlying root yield  
812 variation ( $58.76\% \leq H^2 \leq 74.28\%$ ). Finally, there were five  
813 QTL ( $h_{\text{total}}^2 = 51.22\%$ ) for FYTHA ( $H^2 = 55.01$ ), but only one  
814 ( $h_{\text{total}}^2 = 15.53\%$ ) for CI ( $H^2 = 80.50$ ). Although number of roots  
815 seemed to be as heritable as root yield (Table 1), the latter traits  
816 are likely more complex in terms of their genetic architecture  
817 than the former ones. That is, not only number of roots con-  
818 tributes to yield, but also size and composition, so we can expect  
819 that more regions are involved in root yield, in addition to those  
820 involved in number of roots. Nevertheless, Yada *et al.* (2017b)  
821 found a rather low trait heritability (likely individual-basis) for  
822 commercial root yield ( $H^2 = 24\%$ ) among 278 full-sibs of a cross  
823 between 'New Kawogo', a Ugandan landrace, and 'Beauregard',  
824 possibly due to stronger G×E interaction, which adds to the  
825 trait complexity. Here, G×E interaction seemed important for all  
826 traits and its consequences to QTL mapping and breeding will  
827 be explored in future studies. As QTL mapping targets major  
828 QTL, usually stable across environments, most of the minor ones  
829 must have gone undetected. Moreover, additional genetic varia-  
830 tion could be due to higher order allele interactions and genetic  
831 epistasis, which the current models do not account for. In fact,  
832 only a few minor QTL co-localized among number of roots and  
833 yield traits, which explains lower correlations among QTL-based  
834 breeding values (from  $-0.02$  to  $0.63^{***}$ , see Figure S3) relative  
835 to correlations among predicted means (from  $0.21^{***}$  to  $0.84^{***}$ ,  
836 Figure 1) between these sets of traits. Based on the correlation  
837 between QTL-based breeding values, FYTHA does not seem to  
838 be useful in indirect selection for CYTHA ( $-0.02$ , see Figure S3),  
839 as suggested previously (Chang *et al.* 2009), even though some  
840 correlation ( $0.21^{**}$ ) among their predicted means was observed  
841 (Figure 1). Although 'Beauregard' and 'Tanzania' contributed  
842 more importantly with positive and negative major effects, re-  
843 spectively, the presence of both favorable and unfavorable QTL  
844 alleles in either parents possibly explains the presence of trans-  
845 gressive segregants for all traits. Transgression in polyploids  
846 seems to be due to not only cumulative complementary alleles  
847 at different loci (Tanksley 1993), but also from the same QTL.  
848 In fact, increased heterozygosity has been suggested as one of  
849 the major forces of polyploid evolutionary success, as a broader  
850 allele repertoire may result in the variation of gene expression  
851 and regulation needed to thrive in more diverse environmental  
852 conditions (Van De Peer *et al.* 2009). As an example, 'Tanzania'  
853 exhibited allele contributing to increase CYTHA from a major  
854 QTL (Figure 4), although this landrace was not very productive  
855 in our environments overall. The additive effects are the most  
856 important when performing selection as for a breeding point-of-  
857 view. However, one could easily estimate eventual dominance  
858 effects from the detected QTL using simpler biallelic-based mod-  
859 els as proposed previously (Hackett *et al.* 2014; Chen *et al.* 2018).  
860 The effective use of higher allele interactions in QTL detection  
861 remains limited, though.

862 Several studies have looked at genes involved in storage  
863 root initiation and development in sweetpotato as reviewed by

864 Khan *et al.* (2016). The storage roots differentiate from lateral  
865 roots by development of cambia around the protoxylem and  
866 secondary xylem, while lignification of the steles of some lateral  
867 roots inhibits this transformation (Villordon *et al.* 2012). The  
868 transformation is genetically and environmentally controlled.  
869 Using the expression profile of the parents of the current map-  
870 ping population, we found genes in leaves and roots (see Figure  
871 S5) related to root development and sugar transport within the  
872 QTL hotspot on LG 3 associated with number of storage roots,  
873 implying that both root restructuring and carbon supply is likely  
874 involved in the number of lateral root that transform to storage  
875 root. In *A. thaliana*, SKU5 (*itf03g02930* homolog) was shown to  
876 have a role in regulating directional root growth as mutants in  
877 SKU5 were shorter than wild-type and altered in the angle of tip  
878 growth when grown on agar (Sedbrook *et al.* 2002). ANN1 and  
879 ANN2 (*itf03g03280* homolog) have been shown to be involved in  
880 post-phloem sugar transport to the root tip (Wang *et al.* 2018), a  
881 phenotype essential to development of storage roots. AtWOX13  
882 (*itf03g03460* homolog) was shown to take part in lateral root de-  
883 velopment (Kreis *et al.* 2008). Other genes such as *SRF1* through  
884 *SRF10* (Tanaka *et al.* 2005), *knotted1*-like homeobox (*KNOXI*;  
885 Tanaka *et al.* 2008), MADS-box genes (Kim 2002), expansin (EXP)  
886 genes and BEL1-like homeodomain (Ponniah *et al.* 2017) have  
887 been strongly implicated in storage root formation and devel-  
888 opment in sweetpotato. Though we did not find evidence of  
889 differential expression of these genes in the available transcrip-  
890 tomic data, two MADS-box transcription factors (*itf03g02230*  
891 and *itf03g02240*), a BEL1-like homeodomain (*itf03g02670*) and an  
892 EXP (*itf03g05010*) were all found within the QTL region on LG 3.  
893 The association between these genes and the genes described in  
894 this study is yet to be defined and suggests the complex nature  
895 of storage root formation and development. On the QTL hotspot  
896 related to storage root weight on LG 15, we found the *CTR1*  
897 gene (*itf15g01020* homolog), which encodes a serine-threonine  
898 kinase and functions in the ethylene signaling pathway leading  
899 to inhibition of cell proliferation (Ramzan *et al.* 2015) had vari-  
900 able expression in the sampled roots. Rose *et al.* (1997) showed  
901 that inhibition of ethylene biosynthesis led to inhibition of *EXP1*  
902 gene in tomato. In sweetpotato, down-regulation of an *EXP1*  
903 homologue (*IbEXP1*) enhanced storage root development (Noh  
904 *et al.* 2013). While little is known about the role of ethylene in  
905 storage root development, the complex interactions of multiple  
906 hormones in storage root formation would suggest ethylene may  
907 be involved in storage root development. The main component  
908 of the sweetpotato storage root is starch. We found differen-  
909 tial expression of a gene encoding starch branching enzyme  
910 (*itf15g01120* homolog). Starch biosynthesis involves four major  
911 classes of enzymes: ADP-glucose pyrophosphorylases, starch  
912 synthases, starch branching enzymes and starch debranching  
913 enzymes (Li *et al.* 2014). Starch branching enzymes influence  
914 the structure of starch through formation of  $\alpha$ -1,6-branch points  
915 with different frequencies and chain length (Tetlow and Emes  
916 2014). Given the number and magnitude of QTL identified in  
917 the current study as associated with yield and yield component  
918 traits, the results indicate that the candidate genes identified  
919 in the current study may interact with those from other loci to  
920 determine the final yield in terms of number, composition and  
921 weight of storage roots.

922 Here, we present a stepwise-based algorithm for multiple-  
923 QTL model selection in full-sib populations of autopolyploid  
924 species with a fully integrated map, from which QTL genotype  
925 conditional probabilities can be calculated. The use of score

926 statistics is a key component of this new method, which de-  
927 pends on a dynamic and fast-computing test for model selection  
928 during the QTL search process. Simulations were performed in  
929 order to assess the impact of using different threshold criteria  
930 for QTL detection and to provide some empirical sense on how  
931 to use the method in practice. REMIM has been carried out in  
932 a hexaploid sweetpotato population to detect both minor and  
933 major loci contributing to the variation of yield-related traits  
934 that may be targeted in molecular-assisted breeding. The use of  
935 random-effect models has created the context for fitting multiple  
936 QTL, providing straightforward information on variance com-  
937 ponents, important for computing QTL heritabilities. Finally,  
938 QTL genotype predictions allowed us to estimate allele-specific  
939 additive effects, for characterizing major additive allele contribu-  
940 tions, and compute QTL-based breeding values, that can be used  
941 for performing selection. This novel approach may enable more  
942 complex models, such as those accounting for interaction among  
943 QTL as well as multiple traits or multiple environments in order  
944 to study shared genetic control in different traits/environments  
945 and G×E interaction at QTL level. Understanding the genetic  
946 architecture of root yield and other traits related to quality and  
947 resistance to biotic and abiotic stresses represents great opportu-  
948 nity for improving interesting characteristics in sweetpotato and  
949 other polyploids. Most of these important traits are polygenic  
950 in nature and only assessed later in a breeding program, where  
951 marker-assisted selection could help to speed up the process.

## 952 Author contributions

953 DCG, WJG, AK, GCY and ZBZ conceived and designed the  
954 experiments. BAO performed DNA sequencing. DCG, FD, VM,  
955 WJG and AK carried out field experiments. GSP, MM and ZBZ  
956 developed tools and analyzed the data. CRB, JCW and DCG  
957 carried out candidate gene expression profiling. GSP wrote the  
958 manuscript. All authors read and approved the manuscript.

## 959 Acknowledgments

960 This work was supported by the Bill & Melinda Gates Founda-  
961 tion [OPP1052983] as part of the Genomic Tools for Sweet-  
962 potato (GT4SP) Project. We acknowledge CIP sweetpotato breed-  
963 ing technical team in Peru for running experiments and col-  
964 lecting phenotypic data. Research at CIP was undertaken as  
965 part of the CGIAR Research Program on Roots, Tubers and  
966 Bananas (RTB), which is supported by CGIAR Fund Donors  
967 (<http://www.cgiar.org/about-us/our-funders/>).

## 968 Literature Cited

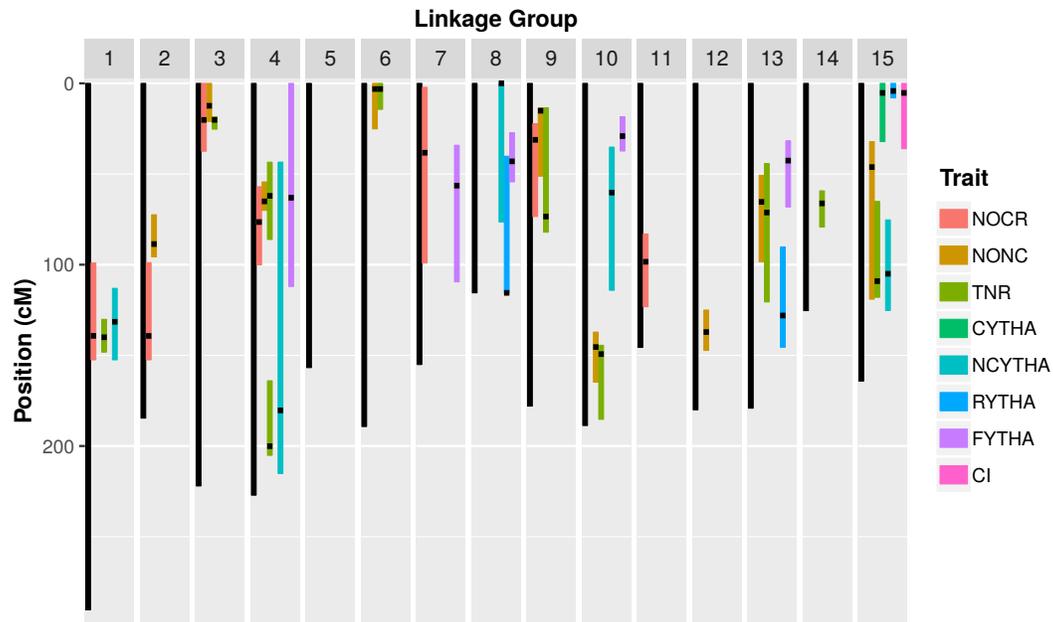
- 969 Almasy, L. and J. Blangero, 2010 Variance component methods  
970 for analysis of complex phenotypes. *Cold Spring Harbor Pro-*  
971 *tocols* 5: 1–12.
- 972 Balsalobre, T. W. A., G. S. Pereira, G. R. A. Margarido, R. Gazaffi,  
973 F. Z. Barreto, *et al.*, 2017 GBS-based single dosage markers for  
974 linkage and QTL mapping allow gene mining for yield-related  
975 traits in sugarcane. *BMC Genomics* 18: 1–19.
- 976 Bourke, P. M., G. Van Geest, R. E. Voorrips, J. Jansen, T. Kranen-  
977 burg, *et al.*, 2018 polypmapR - linkage analysis and genetic map  
978 construction from F1 populations of outcrossing polyploids.  
979 *Bioinformatics* 34: 3496–3502.
- 980 Broman, K. W., H. Wu, S. Sen, and G. A. Churchill, 2003 R/qtl:  
981 QTL mapping in experimental crosses. *Bioinformatics* 19: 889–  
982 890.

- 983 Cervantes-Flores, J. C., B. Sosinski, K. V. Pecota, R. O. Mwangi, 1045  
984 G. L. Catignani, *et al.*, 2011 Identification of quantitative trait 1046  
985 loci for dry-matter, starch, and  $\beta$ -carotene content in sweet- 1047  
986 potato. *Molecular Breeding* **28**: 201–216. 1048
- 987 Cervantes-Flores, J. C., G. C. Yencho, A. Kriegner, K. V. Pecota, 1049  
988 M. A. Faulk, *et al.*, 2008a Development of a genetic link- 1050  
989 age map and identification of homologous linkage groups 1051  
990 in sweetpotato using multiple-dose AFLP markers. *Molecular* 1052  
991 *Breeding* **21**: 511–532. 1053
- 992 Cervantes-Flores, J. C., G. C. Yencho, K. V. Pecota, B. Sosinski, 1054  
993 and R. O. M. Mwangi, 2008b Detection of Quantitative Trait 1055  
994 Loci and Inheritance of Root-knot Nematode Resistance in 1056  
995 Sweetpotato. *Journal of American Horticultural Science* **133**: 1057  
996 844–851. 1058
- 997 Chang, K. Y., H. F. Lo, Y. C. Lai, P. J. Yao, K. H. Lin, *et al.*, 2009 1059  
998 Identification of quantitative trait loci associated with yield- 1060  
999 related traits in sweet potato (*Ipomoea batatas*). *Botanical* 1061  
1000 *Studies* **50**: 43–55. 1062
- 1001 Chen, J., F. Zhang, L. Wang, L. Leach, and Z. Luo, 2018 Orthogo- 1063  
1002 nal contrast based models for quantitative genetic analysis in 1064  
1003 autotetraploid species. *New Phytologist* **220**: 332–346. 1065
- 1004 Churchill, G. A. and R. W. Doerge, 1994 Empirical threshold 1066  
1005 values for quantitative trait mapping. *Genetics* **138**: 963–71. 1067
- 1006 Comai, L., 2005 The advantages and disadvantages of being 1068  
1007 polyploid. *Nature Reviews Genetics* **6**: 836–846. 1069
- 1008 Covarrubias-Pazarán, G., 2016 Genome-assisted prediction of 1070  
1009 quantitative traits using the R package sommer. *PLoS ONE* 1071  
1010 **11**: 1–15. 1072
- 1011 Crepieux, S., C. Lebreton, P. Flament, and G. Charmet, 2005 1073  
1012 Application of a new IBD-based QTL mapping method to 1074  
1013 common wheat breeding population: Analysis of kernel hard- 1075  
1014 ness and dough strength. *Theoretical and Applied Genetics* 1076  
1015 **111**: 1409–1419. 1077
- 1016 Da Costa E Silva, L., S. Wang, and Z.-B. Zeng, 2012a Composite 1078  
1017 Interval Mapping and Multiple Interval Mapping: Procedures 1079  
1018 and Guidelines for Using Windows QTL Cartographer. In 1080  
1019 *Quantitative Trait Loci (QTL): Methods and Protocols*, edited by 1081  
1020 S. A. Rifkin, *Methods in Molecular Biology*, chapter 6, pp. 1082  
1021 75–119, Humana Press, Totowa, NJ. 1083
- 1022 Da Costa E Silva, L., S. Wang, and Z.-B. Zeng, 2012b Multiple 1084  
1023 trait multiple interval mapping of quantitative trait loci from 1085  
1024 inbred line crosses. *BMC Genetics* **13**: 1–24. 1086
- 1025 Druet, T., S. Fritz, M. Boussaha, S. Ben-Jemaa, F. Guillaume, 1087  
1026 *et al.*, 2008 Fine mapping of quantitative trait loci affecting 1088  
1027 female fertility in dairy cattle on BTA03 using a dense single- 1089  
1028 nucleotide polymorphism map. *Genetics* **178**: 2227–2235. 1090
- 1029 FAO, 2019 FAOSTAT Crops. Available at 1091  
1030 <http://www.fao.org/faostat/>. Accessed on Feb-18-2019. 1092
- 1031 Gemenet, D. C., G. S. Pereira, B. D. Boeck, J. C. Wood, M. Mol- 1093  
1032 linari, *et al.*, 2019 Quantitative trait loci mapping and candi- 1094  
1033 date gene profiling reveal the genetic architecture controlling 1095  
1034  $\beta$ -carotene accumulation and the associated altered carbon 1096  
1035 partitioning into starch in hexaploid sweetpotato [*Ipomoea* 1097  
1036 *batatas* (L.) Lam.]. Submitted pp. 1–23. 1098
- 1037 Gerard, D., L. F. V. Ferrão, A. A. F. Garcia, and M. Stephens, 1099  
1038 2018 Genotyping Polyploids from Messy Sequencing Data. 1100  
1039 *Genetics* **210**: 789–807. 1101
- 1040 Grattapaglia, D. and R. Sederoff, 1994 Genetic linkage maps of 1102  
1041 *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo- 1103  
1042 testcross: mapping strategy and RAPD markers. *Genetics* 1104  
1043 **1137**: 1121–1137. 1105
- 1044 Hackett, C. A., B. Boskamp, A. Vogogias, K. F. Preedy, I. Milne, 1106  
*et al.*, 2016 TetraploidSNPMap: software for linkage analysis 1107  
and QTL mapping in autotetraploid populations using SNP 1108  
dosage data. *Journal of Heredity* **108**: 438–442. 1109
- Hackett, C. A., J. E. Bradshaw, and G. J. Bryan, 2014 QTL map- 1110  
ping in autotetraploids using SNP dosage information. *Theo- 1111  
retical and Applied Genetics* **127**: 1885–1904. 1112
- Hackett, C. A., J. E. Bradshaw, and J. W. McNicol, 2001 Interval 1113  
mapping of quantitative trait loci in autotetraploid species. 1114  
*Genetics* **159**: 1819–1832. 1115
- Hulse-Kemp, A. M., J. Lemm, J. Plieske, H. Ashrafi, R. Buy- 1116  
yarapu, *et al.*, 2015 Development of a 63K SNP Ar- 1117  
ray for Cotton and High-Density Mapping of Intraspe- 1118  
cific and Interspecific Populations of *Gossypium* spp. *G3 1119  
Genes | Genomes | Genetics* **5**: 1187–1209. 1120
- Jiang, C. and Z.-B. Zeng, 1997 Mapping quantitative trait loci 1121  
with dominant and missing markers in various crosses from 1122  
two inbred lines. *Genetica* **101**: 47–58. 1123
- Kao, C.-H., Z.-B. Zeng, and R. D. Teasdale, 1999 Multiple interval 1124  
mapping for quantitative trait loci. *Genetics* **152**: 1203–1216. 1125
- Kempthorne, O., 1955 The correlation between relatives in a 1126  
simple autotetraploid population. *Genetics* **40**: 168–174. 1127
- Khan, M. A., D. C. Gemenet, and A. Villordon, 2016 Root System 1128  
Architecture and Abiotic Stress Tolerance: Current Knowledge 1129  
in Root and Tuber Crops. *Frontiers in Plant Science* **7**: 1–13. 1130
- Kim, J.-H., I. K. Chung, and K.-M. Kim, 2017 Construction of a 1131  
genetic map using EST-SSR markers and QTL analysis of ma- 1132  
jor agronomic characters in hexaploid sweet potato (*Ipomoea 1133  
batatas* (L.) Lam.). *PLoS ONE* **12**: e0185073. 1134
- Kim, S.-H., 2002 Isolation of MADS-box Genes from Sweet 1135  
Potato (*Ipomoea batatas* (L.) Lam.) Expressed Specifically in 1136  
Vegetative Tissues. *Plant and Cell Physiology* **43**: 314–322. 1137
- Klaassen, M. T., P. M. Bourke, C. Maliepaard, and L. M. Trindade, 1138  
2019 Multi-allelic QTL analysis of protein content in a bi- 1139  
parental population of cultivated tetraploid potato. *Euphytica 1140  
215: 18. 1141*
- Kreis, M., V. Thareau, A. Lecharny, Y. Deveaux, H. Moreau, *et al.*, 1142  
2008 Genes of the most conserved WOX clade in plants affect 1143  
root and flower development in Arabidopsis. *BMC Evolution- 1144  
ary Biology* **8**: 291. 1145
- Kriegner, A., J. C. Cervantes, K. Burg, R. O. M. Mwangi, and 1146  
D. Zhang, 2003 A genetic linkage map of sweetpotato [*Ip- 1147  
omoea batatas* (L.) Lam.] based on AFLP markers. *Molecular 1148  
Breeding* **11**: 169–185. 1149
- Lander, E. S. and P. Green, 1987 Construction of multilocus 1150  
genetic linkage maps in humans. *Proceedings of the National 1151  
Academy of Sciences of the United States of America* **84**: 2363– 1152  
2367. 1153
- Langmead, B. and S. L. Salzberg, 2013 Fast gapped-read align- 1154  
ment with Bowtie 2. *Nature Methods* **9**: 357–359. 1155
- Lau, K. H., M. del Rosario Herrera, E. Crisovan, S. Wu, Z. Fei, 1156  
*et al.*, 2018 Transcriptomic analysis of sweet potato under dehy- 1157  
dration stress identifies candidate genes for drought tolerance. 1158  
*Plant Direct* **2**. 1159
- Laurie, C., S. Wang, L. A. Carlini-Garcia, and Z.-B. B. Zeng, 2014 1160  
Mapping epistatic quantitative trait loci. *BMC Genetics* **15**: 1161  
112. 1162
- Li, H., N. Zhao, X. Yu, Y. Liu, H. Zhai, *et al.*, 2014 Identifica- 1163  
tion of QTLs for storage root yield in sweetpotato. *Scientia 1164  
Horticulturae* **170**: 182–188. 1165
- Lippert, C., J. Xiang, D. Horta, C. Widmer, C. Kadie, *et al.*, 2014 1166  
Greater power and computational efficiency for kernel-based 1167  
association testing of sets of genetic variants. *Bioinformatics 1168*

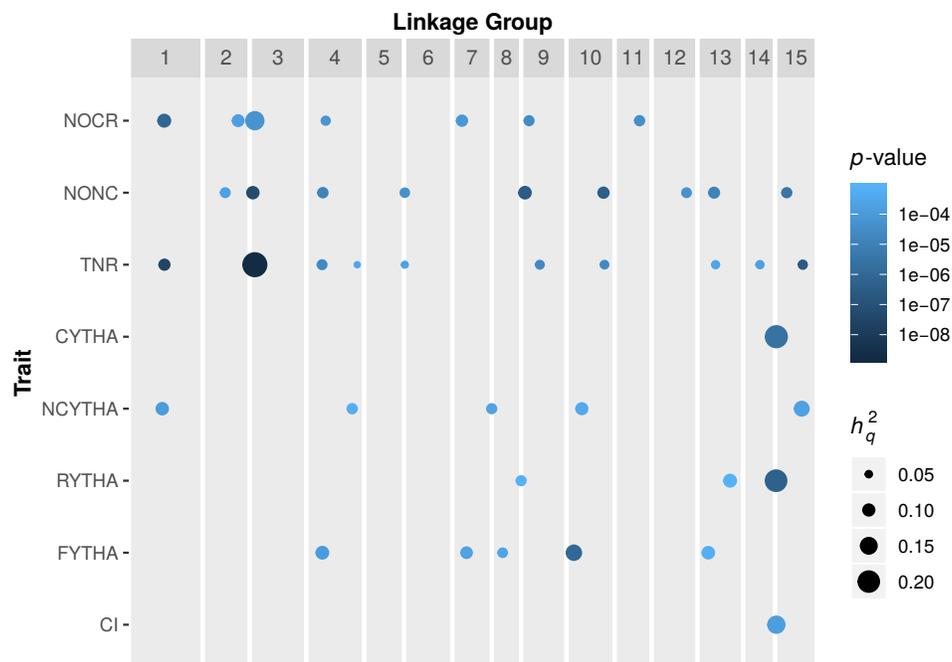
- 1107 **30**: 3206–3214. 1169
- 1108 Low, J. W., R. O. Mwangi, M. Andrade, E. Carey, and A. M. Ball, 1170
- 1109 2017 Tackling vitamin A deficiency with biofortified sweet- 1171
- 1110 potato in sub-Saharan Africa. *Global Food Security* **14**: 23–30. 1172
- 1111 Margarido, G. R. A., A. P. Souza, and A. A. F. Garcia, 2007 1173
- 1112 OneMap: software for genetic mapping in outcrossing species. 1174
- 1113 *Hereditas* **144**: 78–79. 1175
- 1114 Massa, A. N., N. C. Manrique-Carpintero, J. Coombs, K. G. 1176
- 1115 Haynes, P. C. Bethke, *et al.*, 2018 Linkage analysis and QTL 1177
- 1116 mapping in a tetraploid russet mapping population of potato. 1178
- 1117 *BMC Genetics* **19**: 1–13. 1179
- 1118 Mayer, M., 2005 A comparison of regression interval mapping 1180
- 1119 and multiple interval mapping for linked QTL. *Heredity* **94**: 1181
- 1120 599–605. 1182
- 1121 Mengist, M. F., S. Alves, D. Griffin, J. Creedon, M. J. McLaughlin, 1183
- 1122 *et al.*, 2018 Genetic mapping of quantitative trait loci for tuber- 1184
- 1123 cadmium and zinc concentration in potato reveals associations 1185
- 1124 with maturity and both overlapping and independent com- 1186
- 1125 ponents of genetic control. *Theoretical and Applied Genetics* 1187
- 1126 **131**: 929–945. 1188
- 1127 Mollinari, M. and A. A. F. Garcia, 2018 Linkage analysis and 1189
- 1128 haplotype phasing in experimental autopolyploid populations 1190
- 1129 with high ploidy level using hidden Markov models. *bioRxiv* 1191
- 1130 pp. 1–30. 1192
- 1131 Mollinari, M., B. Olokulu, G. S. Pereira, D. Gemenet, C. Yencho, 1193
- 1132 *et al.*, 2019 Unraveling sweetpotato complex inheritance using 1194
- 1133 ultra-dense multilocus genetic mapping. In preparation pp. 1195
- 1134 1–39. 1196
- 1135 Monden, Y., T. Hara, Y. Okada, O. Jahana, A. Kobayashi, *et al.*, 1197
- 1136 2015 Construction of a linkage map based on retrotransposon 1198
- 1137 insertion polymorphisms in sweetpotato via high-throughput 1199
- 1138 sequencing. *Breeding Science* **65**: 145–153. 1200
- 1139 Mwangi, R. O., M. I. Andrade, E. E. Carey, J. W. Low, G. C. Yen- 1201
- 1140 cho, *et al.*, 2017 Sweetpotato (*Ipomoea batatas* L.). In *Genetic* 1202
- 1141 *Improvement of Tropical Crops*, edited by H. Campos and P. D. 1203
- 1142 Caligari, chapter 6, pp. 181–218, Springer Nature. 1204
- 1143 Nakayama, H., M. Tanaka, Y. Takahata, K. Matsui, H. Iwahori, 1205
- 1144 *et al.*, 2012 Development of AFLP-derived SCAR markers as- 1206
- 1145 sociated with resistance to two races of southern root-knot 1207
- 1146 nematode in sweetpotato. *Euphytica* **188**: 175–185. 1208
- 1147 Noh, S. A., H.-S. Lee, Y.-S. Kim, K.-H. Paek, J. S. Shin, *et al.*, 2013 1209
- 1148 Down-regulation of the IbEXP1 gene enhanced storage root 1210
- 1149 development in sweetpotato. *Journal of Experimental Botany* 1211
- 1150 **64**: 129–142. 1212
- 1151 Pereira, G. S., A. A. F. Garcia, and G. R. A. Margarido, 2018 1213
- 1152 A fully automated pipeline for quantitative genotype calling 1214
- 1153 from next generation sequencing data in autopolyploids. *BMC* 1215
- 1154 *Bioinformatics* **19**: 398. 1216
- 1155 Ponniah, S. K., J. Thimmapuram, K. Bhide, V. K. Kalavacharla, 1217
- 1156 and M. Manoharan, 2017 Comparative analysis of the root 1218
- 1157 transcriptomes of cultivated sweetpotato (*Ipomoea batatas* 1219
- 1158 [L.] Lam) and its wild ancestor (*Ipomoea trifida* [Kunth] G. 1220
- 1159 Don). *BMC Plant Biology* **17**: 1–14. 1221
- 1160 Preedy, K. F. and C. A. Hackett, 2016 A rapid marker ordering 1222
- 1161 approach for high-density genetic linkage maps in experimen- 1223
- 1162 tal autotetraploid populations using multidimensional scaling. 1224
- 1163 *Theoretical and Applied Genetics* **129**: 2117–2132. 1225
- 1164 Qu, L., T. Guennel, and S. L. Marshall, 2013 Linear score tests for 1226
- 1165 variance components in linear mixed models and applications 1227
- 1166 to genetic association studies. *Biometrics* **69**: 883–892. 1228
- 1167 R Core Team, 2018 R: A Language and Environment for Statistical 1229
- 1168 Computing. 1230
- Ramzan, A., X. Wang, Y. Zubo, S. N. Shakeel, S. Aman, *et al.*, 1169
- 2015 Ethylene Inhibits Cell Proliferation of the Arabidopsis 1170
- Root Meristem. *Plant Physiology* **169**: 338–350. 1171
- Revelle, W., 2018 psych: Procedures for Psychological, Psycho- 1172
- metric, and Personality Research. 1173
- Rose, J. K. C., H. H. Lee, and A. B. Bennett, 1997 Expression 1174
- of a divergent expansin gene is fruit-specific and ripening- 1175
- regulated. *Proceedings of the National Academy of Sciences* 1176
- 94**: 5955–5960. 1177
- Schmitz Carley, C. A., J. J. Coombs, D. S. Douches, P. C. Bethke, 1178
- J. P. Palta, *et al.*, 2017 Automated tetraploid genotype calling 1179
- by hierarchical clustering. *Theoretical and Applied Genetics* 1180
- 130**: 717–726. 1181
- Schumann, M. J., Z. B. Zeng, M. E. Clough, and G. C. Yencho, 1182
- 2017 Linkage map construction and QTL analysis for inter- 1183
- nal heat necrosis in autotetraploid potato. *Theoretical and* 1184
- Applied Genetics* **130**: 2045–2056. 1185
- Sedbrook, J. C., K. L. Carroll, K. F. Hung, P. H. Masson, and 1186
- C. R. Somerville, 2002 The Arabidopsis SKU5 Gene Encodes an 1187
- Extracellular Glycosyl Phosphatidylinositol-Anchored Gly- 1188
- coprotein Involved in Directional Root Growth. *The Plant Cell* 1189
- 14**: 1635–1648. 1190
- Serang, O., M. Mollinari, and A. A. F. Garcia, 2012 Efficient 1191
- exact maximum a posteriori computation for Bayesian SNP 1192
- genotyping in polyploids. *PLoS ONE* **7**: 1–13. 1193
- Shirasawa, K., M. Tanaka, Y. Takahata, D. Ma, Q. Cao, *et al.*, 1194
- 2017 A high-density SNP genetic map consisting of a com- 1195
- plete set of homologous groups in autohexaploid sweetpotato 1196
- (*Ipomoea batatas*). *Scientific Reports* **7**: 1–8. 1197
- Stam, P., 1993 Construction of integrated genetic linkage maps 1198
- by means of a new computer package: Join Map. *The Plant* 1199
- Journal* **3**: 739–744. 1200
- Tanaka, M., N. Kato, H. Nakayama, M. Nakatani, and Y. Taka- 1201
- hata, 2008 Expression of class I knotted1-like homeobox genes 1202
- in the storage roots of sweetpotato (*Ipomoea batatas*). *Journal* 1203
- of Plant Physiology* **165**: 1726–1735. 1204
- Tanaka, M., Y. Takahata, and M. Nakatani, 2005 Analysis of genes 1205
- developmentally regulated during storage root formation of 1206
- sweet potato. *Journal of Plant Physiology* **162**: 91–102. 1207
- Tanksley, S. D., 1993 Mapping polygenes. *Annu. Rev. Genet.* **27**: 1208
- 205–233. 1209
- Tetlow, I. J. and M. J. Emes, 2014 A review of starch-branching 1210
- enzymes and their role in amylopectin biosynthesis. *IUBMB* 1211
- Life* **66**: 546–558. 1212
- Ukoskit, K. and P. G. Thompson, 1997 Autopolyploidy versus 1213
- Allopolyploidy and Low-density Randomly Amplified Poly- 1214
- morphic DNA Linkage Maps of Sweetpotato. *J. Amer. Soc.* 1215
- Hort. Sci.* **122**: 822–828. 1216
- Van De Peer, Y., S. Maere, and A. Meyer, 2009 The evolutionary 1217
- significance of ancient genome duplications. *Nature Reviews* 1218
- Genetics* **10**: 725–732. 1219
- van Geest, G., P. M. Bourke, R. E. Voorrips, A. Marasek- 1220
- Ciolakowska, Y. Liao, *et al.*, 2017 An ultra-dense integrated 1221
- linkage map for hexaploid chrysanthemum enables multi- 1222
- allelic QTL analysis. *Theoretical and Applied Genetics* **130**: 1223
- 2527–2541. 1224
- Verbeke, G. and G. Molenberghs, 2003 The use of score tests for 1225
- inference on variance components. *Biometrics* **59**: 254–262. 1226
- Villordon, A., D. LaBonte, J. Solis, and N. Firon, 2012 Charac- 1227
- terization of lateral root development at the onset of storage 1228
- root initiation in ‘Beauregard’ sweetpotato adventitious roots. 1229
- HortScience* **47**: 961–968. 1230

- 1231 Voorrips, R. E., G. Gort, and B. Vosman, 2011 Genotype calling  
1232 in tetraploid species from bi-allelic marker data using mixture  
1233 models. *BMC Bioinformatics* **12**: 172.
- 1234 VSN International, 2014 GenStat for Windows 16th Edition.
- 1235 Wadl, P. A., B. A. Olukolu, S. E. Branham, R. L. Jarret, G. C.  
1236 Yencho, *et al.*, 2018 Genetic Diversity and Population Structure  
1237 of the USDA Sweetpotato (*Ipomoea batatas*) Germplasm  
1238 Collections Using GBSpoly. *Frontiers in Plant Science* **9**: 1–13.
- 1239 Wang, J., J. Song, G. Clark, and S. J. Roux, 2018 ANN1 and  
1240 ANN2 Function in Post-Phloem Sugar Transport in Root Tips  
1241 to Affect Primary Root Growth. *Plant Physiology* **178**: 390–  
1242 401.
- 1243 Wickham, H., 2016 *ggplot2: elegant graphics for data analysis*.  
1244 Springer.
- 1245 Wu, Q. H., Y. X. Chen, S. H. Zhou, L. Fu, J. J. Chen, *et al.*,  
1246 2015 High-density genetic linkage map construction and QTL  
1247 mapping of grain shape and size in the wheat population  
1248 Yanda1817 × Beinong6. *PLoS ONE* **10**: 1–17.
- 1249 Wu, S., K. H. Lau, Q. Cao, J. P. Hamilton, H. Sun, *et al.*, 2018  
1250 Genome sequences of two diploid wild relatives of cultivated  
1251 sweetpotato reveal targets for genetic improvement. *Nature*  
1252 *Communications* **9**: 4580.
- 1253 Yada, B., A. Alajo, G. N. Ssemakula, R. O. Mwanga, G. Brown-  
1254 Guedira, *et al.*, 2017a Selection of simple sequence repeat mark-  
1255 ers associated with inheritance of sweetpotato virus disease  
1256 resistance in sweetpotato. *Crop Science* **57**: 1421–1430.
- 1257 Yada, B., G. Brown-Guedira, A. Alajo, G. N. Ssemakula,  
1258 E. Owusu-Mensah, *et al.*, 2017b Genetic analysis and asso-  
1259 ciation of simple sequence repeat markers with storage root  
1260 yield, dry matter, starch and  $\beta$ -carotene content in sweetpotato.  
1261 *Breeding Science* **67**: 140–150.
- 1262 Yu, X. X., N. Zhao, H. Li, Q. Jie, H. Zhai, *et al.*, 2014 Identification  
1263 of QTLs for Starch Content in Sweetpotato (*Ipomoea batatas*  
1264 (L.) Lam.). *Journal of Integrative Agriculture* **13**: 310–315.
- 1265 Zeng, Z.-B. B., C.-H. H. Kao, and C. J. Basten, 1999 Estimating the  
1266 genetic architecture of quantitative traits. *Genetical Research*  
1267 **74**: 279–289.
- 1268 Zhao, N., X. Yu, Q. Jie, H. Li, H. Li, *et al.*, 2013 A genetic linkage  
1269 map based on AFLP and SSR markers and mapping of QTL  
1270 for dry-matter content in sweetpotato. *Molecular Breeding* **32**:  
1271 807–820.

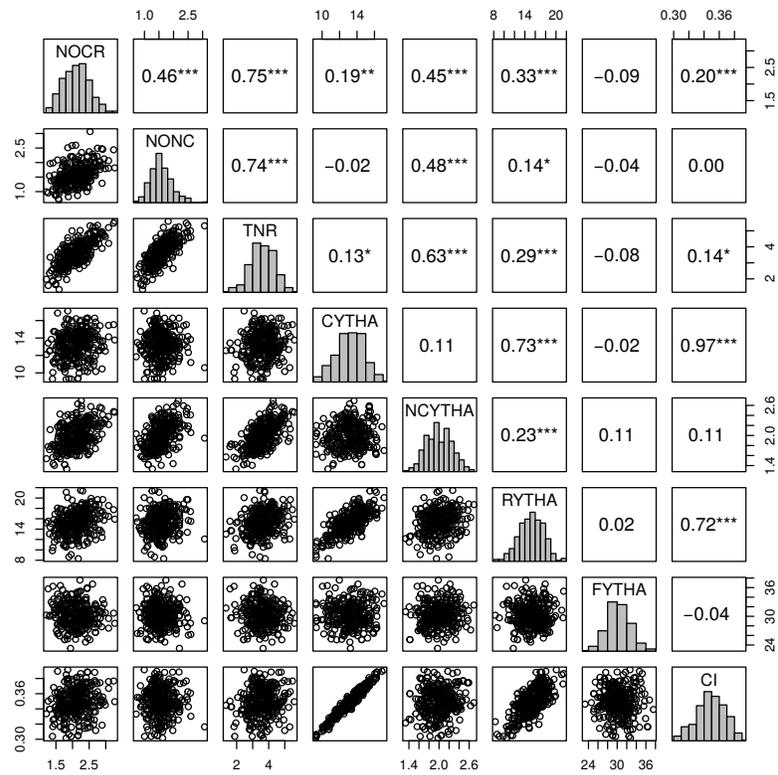
## Supplemental Figures



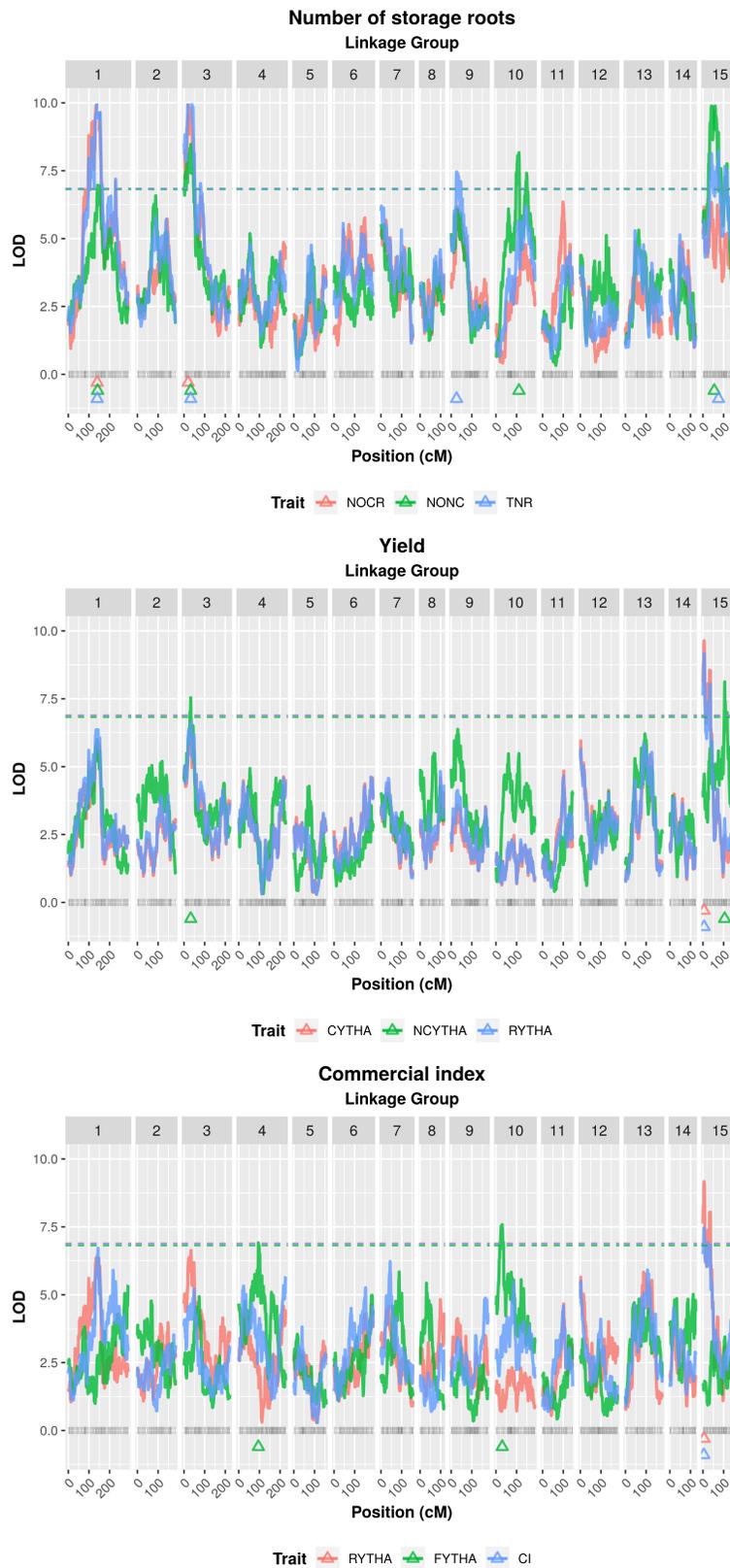
**Figure S1** QTL support intervals from random-effect multiple interval mapping (REMIM) of yield-related traits from ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family. Black dots represent the QTL peaks, and colored bars represent the ~95% support interval computed as  $LOP - 1.5$ . Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).



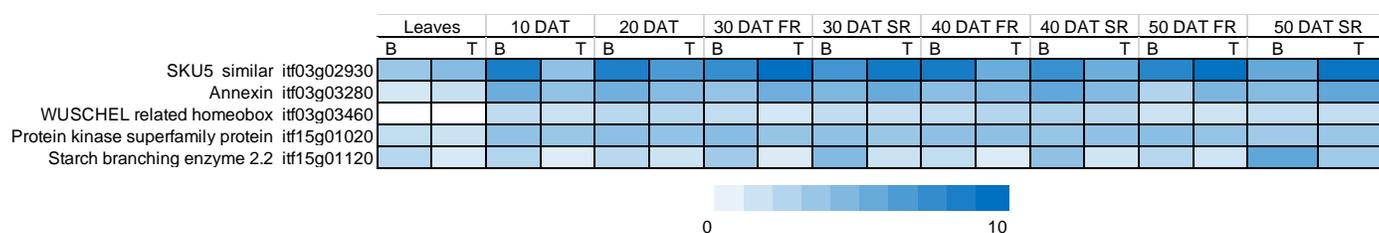
**Figure S2** Score-based  $p$ -values and QTL heritabilities ( $h_q^2$ ) from random-effect multiple interval mapping (REMIM) of eight yield-related traits from ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family. Dots are positioned relative to the QTL peaks: color gradient represents the  $p$ -values, while sizes are proportional to heritabilities of mapped QTL. Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).



**Figure S3** Pearson's correlations ( $*p < 0.05$ ,  $**p < 0.01$ ,  $***p < 0.001$ ) among QTL-based breeding values for eight yield-related traits from 'Beauregard'  $\times$  'Tanzania' (BT) full-sib family. Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI).



**Figure S4** Logarithm of the odds (LOD score) profiles from fixed-effect interval mapping (FEIM) of eight yield-related traits from 'Beaugard' × 'Tanzania' (BT) full-sib family. Triangles represent the QTL peaks. Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in  $t \cdot ha^{-1}$ , foliage yield (FYTHA) in  $t \cdot ha^{-1}$ , and commercial index (CI). Dashed horizontal lines represent the permutation-based genome-wide significance LOD threshold of 0.20.



**Figure S5** Heatmap with expression abundances in fragments per kilobase exon model per million mapped reads (FPKM,  $\log_2$ -transformed) for five genes in leaves and roots of 'Beauregard' (B) and 'Tanzania' (T). DAT: days after transplanting. SR: storage roots. FR: fibrous roots.

## Supplemental Tables

**Table S1** Detection power (in percentage) and absolute difference between simulated and mapped QTL peak position (on average, in centiMorgans) from 1,000 simulated quantitative traits with three QTLs with different heritabilities ( $h_q^2 = \{0.27, 0.18, 0.09\}$ ). Fixed-effect interval mapping (FEIM) and random-effect multiple interval mapping (REMIM) were carried out under different genome-wide significance LOD and pointwise backward  $p$ -value thresholds, respectively, in ‘Beauregard’ × ‘Tanzania’ (BT) full-sib population

$h_q^2$	FEIM			REMIM		
	Genome-wide significance	Power (%)	Difference (cM) <sup>a</sup>	Pointwise significance	Power (%)	Difference (cM) <sup>a</sup>
0.27	0.20	89.40	2.41 (3.31)	$10^{-2}$	92.60	2.75 (3.37)
0.18		73.00	3.16 (3.67)		84.90	3.51 (3.93)
0.09		34.50	4.06 (4.19)		62.50	4.12 (4.13)
0.27	0.15	88.80	2.40 (3.31)	$10^{-3}$	92.00	2.65 (3.36)
0.18		71.60	3.16 (3.67)		81.80	3.29 (3.81)
0.09		32.30	3.99 (4.21)		51.70	3.77 (3.97)
0.27	0.10	88.00	2.40 (3.32)	$10^{-4}$	89.10	2.48 (3.24)
0.18		70.00	3.16 (3.69)		73.40	3.14 (3.73)
0.09		29.90	3.85 (4.08)		39.30	3.67 (3.89)
0.27	0.05	86.60	2.38 (3.33)	$10^{-5}$	84.40	2.38 (3.17)
0.18		66.80	3.13 (3.68)		65.60	2.89 (3.59)
0.09		27.20	3.82 (4.07)		26.70	3.43 (3.86)

<sup>a</sup> Standard errors of means are between parentheses.

**Table S2** False discovery rate (FDR, in percentage) and proportion of matched QTL (coverage, in percentage) relative to support intervals calculated for each  $d = \{1.0, 1.5, 2.0\}$  from 1,000 simulated quantitative traits. Fixed-effect interval mapping (FEIM) and random-effect multiple interval mapping (REMIM) was carried out under different genome-wide significance LOD and pointwise backward  $p$ -value thresholds, respectively, in ‘Beauregard’ × ‘Tanzania’ (BT) full-sib population

$d$	FEIM			REMIM		
	Genome-wide significance	FDR (%)	Coverage (%)	Pointwise significance	FDR (%)	Coverage (%)
1.0	0.20	27.49	85.98	$10^{-2}$	48.95	90.21
1.5		20.53	94.51		46.03	95.67
2.0		17.53	98.32		45.25	97.08
1.0	0.15	25.98	85.99	$10^{-3}$	26.23	91.26
1.5		18.75	94.50		22.77	95.79
2.0		15.66	98.34		21.71	97.16
1.0	0.10	24.01	86.00	$10^{-4}$	16.73	91.82
1.5		16.61	94.52		12.96	95.94
2.0		13.50	98.30		11.92	97.13
1.0	0.05	21.71	86.08	$10^{-5}$	12.90	92.36
1.5		14.25	94.49		9.56	95.87
2.0		10.97	98.27		8.62	97.00

**Table S3** Summary of fixed-effect interval mapping (FEIM) for eight yield-related traits in ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family. Linkage group (LG), map position (in cM) and its ~95% support interval (SI), likelihood-ratio test (LRT), its corresponding logarithm of the odds (LOD) and the adjusted  $R^2$  (in percentage) are shown for each mapped QTL using permutation-based genome-wide significance LOD threshold of 0.20.

Trait <sup>a</sup>	QTL	LG	Position (SI)	LRT	LOD	$R^2_{adj}$ (%)
NOCR	1	1	140.03 (113.05-154.02)	47.71	10.36	11.83
	2	3	20.18 (16.60-39.52)	47.41	10.3	11.74
NONC	1	1	142.10 (127.44-162.44)	32.08	6.97	7.08
	2	3	32.56 (4.68-50.15)	39.02	8.47	9.22
	3	10	111.40 (97.02-154.13)	37.61	8.17	8.79
	4	15	54.49 (34.37-75.28)	47.48	10.31	11.76
TNR	1	1	140.03 (131.50-155.34)	49.28	10.7	12.29
	2	3	32.56 (18.43-44.03)	50.48	10.96	12.64
	3	9	26.14 (15.11-54.08)	34.35	7.46	7.78
	4	15	75.28 (33.13-118.00)	37.88	8.23	8.87
CYTHA	1	15	5.27 (1.07-36.02)	44.4	9.64	10.84
NCYTHA	1	3	32.56 (24.47-44.03)	34.71	7.54	7.89
	2	15	105.02 (103.05-117.01)	37.43	8.13	8.73
RYTHA	1	15	5.27 (1.07-36.02)	42.23	9.17	10.19
FYTHA	1	4	95.01 (68.09-112.02)	31.84	6.91	7.00
	2	10	30.31 (16.12-37.32)	34.93	7.58	7.96
CI	1	15	4.19 (0.00-37.25)	34.31	7.45	7.77

<sup>a</sup> Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in t · ha<sup>-1</sup>, foliage yield (FYTHA) in t · ha<sup>-1</sup>, and commercial index (CI).

**Table S4 Allele additive effects from QTL mapped for eight yield-related traits in ‘Beauregard’ × ‘Tanzania’ (BT) full-sib family using random-effect multiple interval mapping (REMIM). ‘Beauregard’ ( $\{a, \dots, f\}$ ) and ‘Tanzania’ ( $\{g, \dots, l\}$ ) alleles represent the parental contribution to the trait mean.**

Trait <sup>a</sup>	QTL	‘Beauregard’						‘Tanzania’					
		<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>	<i>i</i>	<i>j</i>	<i>k</i>	<i>l</i>
NOCR	1	-0.0767	0.0998	-0.0025	0.0248	-0.0221	-0.0233	-0.0168	0.0216	0.0305	0.0523	-0.1057	0.0181
	2	0.0072	0.0782	-0.0600	0.0351	-0.0162	-0.0444	-0.0083	0.0966	-0.0420	0.0219	-0.0541	-0.0140
	3	0.0112	-0.0360	-0.1171	0.0791	0.0172	0.0456	-0.1475	0.0045	0.0302	0.0685	0.0272	0.0171
	4	-0.0446	-0.0138	0.0545	-0.0276	0.0031	0.0285	0.0407	-0.0534	0.0618	-0.0010	-0.0569	0.0088
	5	0.1042	-0.0060	-0.0487	-0.0179	-0.0287	-0.0029	0.0072	0.0802	-0.0385	-0.0659	0.0086	0.0083
	6	-0.0184	-0.0463	0.0944	0.0299	-0.0200	-0.0396	0.0406	-0.0627	-0.0084	-0.0149	-0.0005	0.0459
	7	-0.0979	0.0436	0.0795	-0.0238	0.0053	-0.0066	-0.0073	-0.0213	-0.0033	-0.0188	0.0254	0.0254
NONC	1	-0.0065	0.0898	-0.0631	-0.0086	0.0004	-0.0120	-0.0094	0.0480	0.0079	-0.0105	-0.0073	-0.0286
	2	0.0492	-0.0402	-0.0619	0.0578	-0.0399	0.0350	-0.0476	-0.0012	-0.0319	0.0295	-0.0109	0.0622
	3	-0.0601	0.0219	-0.0284	0.0097	-0.0053	0.0623	-0.0135	-0.0232	0.0733	0.0096	0.0048	-0.0509
	4	-0.0147	-0.0056	0.0199	-0.0666	0.0535	0.0134	0.0182	0.0024	-0.0133	-0.0574	0.0566	-0.0066
	5	-0.0355	-0.0245	0.1207	-0.0647	-0.0168	0.0207	0.0324	-0.0387	0.0096	0.0079	0.0051	-0.0163
	6	-0.0041	-0.0330	-0.0358	-0.0239	0.0225	0.0743	-0.0605	0.0054	0.0790	-0.0100	0.0174	-0.0314
	7	-0.0375	-0.0169	-0.0522	-0.0189	0.0433	0.0823	0.0026	0.0304	-0.0097	0.0073	0.0098	-0.0404
	8	0.0480	0.0233	-0.0488	-0.0325	0.0034	0.0065	-0.0718	0.0583	0.0063	-0.0527	0.0385	0.0214
	9	-0.0236	0.0210	0.0394	-0.0462	-0.0337	0.0431	0.0222	0.0696	-0.0431	-0.0368	0.0138	-0.0257
TNR	1	-0.1089	0.1489	0.0388	-0.0386	-0.0364	-0.0038	-0.0366	0.0279	-0.0089	0.1466	-0.1538	0.0248
	2	0.0674	-0.0521	-0.2164	0.1942	-0.0873	0.0942	-0.2958	0.0116	-0.0017	0.1910	0.0015	0.0935
	3	-0.0910	0.0071	-0.0011	0.0510	-0.0154	0.0495	0.0869	-0.0924	0.1659	0.0022	-0.0802	-0.0824
	4	-0.0169	0.0362	0.0188	-0.0017	-0.0564	0.0201	-0.0109	-0.0672	0.0224	0.0496	-0.0830	0.0893
	5	0.0093	-0.0954	0.0393	-0.0886	0.0793	0.0561	0.0286	-0.0325	0.0097	-0.0735	0.0481	0.0197
	6	-0.0923	-0.0610	0.1614	-0.0188	-0.0513	0.0620	0.0677	-0.0757	-0.0191	-0.0217	-0.0048	0.0536
	7	0.0630	-0.0419	-0.0103	-0.0742	-0.0033	0.0667	-0.0921	0.0702	0.0999	-0.0205	0.0585	-0.1160
	8	0.0531	0.0498	-0.0964	-0.0560	0.0354	0.0141	-0.1263	0.0350	-0.0019	-0.0542	0.0885	0.0590
	9	-0.0183	0.0886	-0.0302	-0.0284	0.0634	-0.0751	0.0805	-0.1203	0.0930	-0.0222	-0.0296	-0.0014
	10	-0.0322	0.0995	0.0176	-0.0951	-0.0842	0.0944	-0.0457	0.1152	0.0623	-0.0294	-0.0839	-0.0186
CYTHA	1	0.3022	0.8168	-0.6094	-0.0872	-0.1778	-0.2447	0.0310	0.2950	1.0456	-0.3993	-0.4669	-0.5054
NCYTHA	1	-0.0293	0.0510	-0.0056	-0.0203	-0.0169	0.0212	0.0075	0.0358	0.0014	0.0273	-0.0984	0.0264
	2	-0.0021	0.0178	0.0192	-0.0118	-0.0464	0.0233	-0.0081	-0.0665	-0.0043	0.0406	-0.0114	0.0497
	3	-0.0346	0.0214	0.0028	-0.0366	0.0557	-0.0086	0.0044	0.0093	0.0535	0.0021	-0.0511	-0.0181
	4	-0.0092	-0.0401	-0.0260	0.0362	-0.0271	0.0662	-0.0336	-0.0396	0.0529	0.0296	0.0200	-0.0294
	5	0.0000	0.0234	-0.0298	-0.0756	0.0141	0.0679	-0.0543	0.0586	0.0442	0.0067	-0.0033	-0.0520
RYTHA	1	-0.0115	-0.2356	-0.3605	0.1004	0.4067	0.1005	0.4142	-0.2283	-0.5186	-0.3287	0.3150	0.3464
	2	0.3820	0.0903	-0.4735	0.1308	0.0187	-0.1483	0.3776	-0.1423	-0.7335	-0.2565	-0.0959	0.8506
	3	0.3013	0.8584	-0.6686	0.0513	-0.3013	-0.2410	0.0457	0.4922	1.1574	-0.5636	-0.4004	-0.7314
FYTHA	1	0.5933	0.2377	0.0278	-0.0235	-0.0610	-0.7743	0.3693	0.0962	-0.5706	-0.0128	0.3894	-0.2715
	2	0.3460	0.2159	0.0281	-0.6872	-0.0797	0.1769	0.6112	-0.5162	0.0262	0.1795	-0.3660	0.0653
	3	-0.2534	-0.0520	0.0282	0.1661	0.2388	-0.1277	0.5113	-0.3626	0.0336	0.4025	-0.0063	-0.5785
	4	-0.2832	-0.2848	0.2943	0.0672	-0.1976	0.4041	-0.2560	-1.0853	0.6058	0.2293	0.4539	0.0524
	5	0.1011	0.2950	0.0799	0.4343	-0.7205	-0.1898	0.0774	-0.7765	0.3509	0.0161	0.0354	0.2968
CI	1	0.0012	0.0120	-0.0060	-0.0027	0.0004	-0.0049	0.0012	0.0040	0.0104	-0.0048	-0.0056	-0.0052

<sup>a</sup> Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in t · ha<sup>-1</sup>, foliage yield (FYTHA) in t · ha<sup>-1</sup>, and commercial index (CI).

**Table S5 Map (in centiMorgans) and genome (in base pairs) positions and support intervals (inside the parentheses) for QTL mapped in 'Beauregard' × 'Tanzania' full-sib family using random-effect multiple interval mapping (REMIM) for eight yield-related traits.**

Trait <sup>a</sup>	QTL	LG	Map position (cM)	<i>I. trifida</i> genome (bp)	<i>I. triloba</i> genome (bp)
NOCR	1	1	139.24 (99.06-152.48)	22077567 (15917835-23281986)	26692560 (10387193-28126436)
	2	2	139.30 (98.83-152.5)	16126352 (11398424-3350738)	11822379 (6161757-13945807)
	3	3	20.18 (0.00-37.44)	1591872 (32030-3185578)	1802096 (22131-3524655)
	4	4	76.48 (57.01-100.14)	4780118 (3520761-7038218)	5570824 (4107181-11896279)
	5	7	38.27 (2.18-99.12)	2838315 (239482-15160871)	3005905 (302031-21622505)
	6	9	31.15 (22.24-73.45)	2036465 (1357520-3997473)	2537129 (1804182-4799470)
	7	11	98.39 (83.07-123.22)	7711459 (6002238-16415359)	10309151 (8138260-18201378)
NONC	1	2	88.62 (72.44-95.72)	10660510 (9568479-11099215)	5346836 (4151689-5850309)
	2	3	12.36 (0.00-21.01)	790052 (32030-1645526)	898334 (22131-1870158)
	3	4	65.09 (54.35-70.05)	4039479 (3226706-4328193)	4701624 (3769011-5022282)
	4	6	3.12 (0.00-25.13)	10827111 (11507438-7099549)	8590172 (9365714-4034190)
	5	9	15.11 (13.48-51.17)	963597 (777490-2666908)	1377719 (1161653-3269774)
	6	10	145.40 (137.16-164.81)	22087839 (21754871-23272183)	26130519 (25776790-27501919)
	7	12	137.16 (125.03-147.31)	21521557 (20738759-22197168)	25320356 (24372733-26179595)
	8	13	65.36 (50.68-98.41)	12302173 (4360140-17295029)	19828339 (8041557-25122550)
	9	15	46.19 (32.15-119.04)	2779657 (1939509-18430098)	3166284 (2238681-22171700)
TNR	1	1	140.03 (130.13-148.23)	22212737 (20422516-23040691)	26810094 (6478813-27858718)
	2	3	20.18 (18.43-25.28)	1591872 (1428660-1971958)	1802096 (1628157-2218567)
	3	4	62.03 (43.43-86.11)	3832284 (2567373-5837890)	4469117 (2988152-9198140)
	4	4	200.09 (164.01-205.15)	30146741 (27537111-30647356)	33756387 (30687469-34246129)
	5	6	3.12 (0.00-14.33)	10827111 (11507438-9128510)	8590172 (9365714-6585862)
	6	9	73.45 (13.48-82.10)	3997473 (777490-4928686)	4799470 (1161653-6032732)
	7	10	149.31 (144.46-185.31)	22370642 (22001105-24392989)	26474885 (26021589-28734008)
	8	13	71.22 (44.23-120.57)	13078300 (3404827-18987490)	20738566 (6812279-27040045)
	9	14	66.19 (59.27-79.27)	13866336 (6439913-15641815)	12047246 (13374139-19755290)
	10	15	109.10 (65.09-118)	11529245 (4456148-18121488)	13540937 (4998050-21802867)
CYTHA	1	15	5.27 (0.00-32.15)	477772 (21822-1939509)	570552 (56000-2238681)
NCYTHA	1	1	131.50 (113.05-152.48)	20574531 (19617725-23281986)	6295866 (8560921-28126436)
	2	4	180.36 (43.43-215.18)	28612876 (2567373-31472749)	32007820 (2988152-35147251)
	3	8	0.00 (0.00-76.43)	19730 (19730-5572337)	67559 (67559-6976839)
	4	10	60.24 (35.15-114.14)	5641312 (2981461-19734636)	7167078 (3609708-23944890)
	5	15	105.02 (75.28-125.3)	10277885 (5578641-18573061)	11851576 (6377917-22348321)
RYTHA	1	8	115.54 (40.19-115.54)	15710677 (2729207-15710677)	17988854 (3362265-17988854)
	2	13	128.02 (90.20-145.54)	19774175 (16115515-20717332)	27997761 (24485922-29160603)
	3	15	4.19 (0.00-8.02)	452966 (21822-656169)	541410 (56000-772513)
FYTHA	1	4	63.08 (0.00-112.02)	3949338 (42497-11940181)	4583079 (27376-15353793)
	2	7	56.40 (34.12-109.48)	4151772 (2632557-10523783)	4391711 (2793744-15591910)
	3	8	43.00 (27.20-54.33)	2990041 (2088002-3564360)	3816702 (2602036-4509723)
	4	10	29.09 (18.40-37.32)	2508929 (1725045-3121782)	2930617 (1764312-3779221)
	5	13	42.60 (31.67-68.18)	3275253 (2251209-12530096)	6492727 (3912768-20079912)
CI	1	15	5.27 (0.00-36.02)	477772 (21822-2185604)	570552 (56000-2490117)

<sup>a</sup> Trait abbreviations: number of commercial (NOCR), noncommercial (NONC) and total (TNR) roots per plant, commercial (CYTHA), noncommercial (NCYTHA) and total (RYTHA) root yield in t · ha<sup>-1</sup>, foliage yield (FYTHA) in t · ha<sup>-1</sup>, and commercial index (CI).