

# 1 Sequencing depth and genotype quality: Accuracy and breeding 2 operation considerations for genomic selection applications in 3 autopolyploid crops

4 Dorcus C Gemenet<sup>1</sup>, Hannele Lindqvist-Kreuze<sup>2</sup>, Bode A Olukolu<sup>3</sup>, Bert De Boeck<sup>2</sup>, Guilherme  
5 da Silva Pereira<sup>4,5</sup>, Marcelo Mollinari<sup>4</sup>, Zhao-Bang Zeng<sup>4</sup>, G Craig Yencho<sup>4</sup>, Hugo Campos<sup>2</sup>

6 <sup>1</sup>International Potato Center, ILRI Campus, P.O. Box 25171-00603, Nairobi, Kenya

7 <sup>2</sup>International Potato Center, Av. La Molina 1895, Lima, Peru

8 <sup>3</sup>University of Tennessee, Knoxville, TN 37996, USA

9 <sup>4</sup>North Carolina State University, Raleigh, NC 27695, USA

10 <sup>5</sup>Current address: University of Sao Paulo, Brazil

11

12  International Potato Center, ILRI Campus, Old Naivasha Road, 25171-00603, Nairobi, Kenya;  
13 Email: [d.gemenet@cgiar.org](mailto:d.gemenet@cgiar.org); Telephone: 254 20 422 3637; ORCID: 0000-0003-4901-1694

## 14 Author Contributions

15 DCG, Z-BZ, GCY and HC led and managed different aspects of sweetpotato research, HLK and  
16 HC lead and managed different aspects of potato research, DCG, BAO, BDB, GSP, MM and  
17 HLK carried out data analysis, DCG wrote the manuscript, all co-authors read, edited and  
18 approved the manuscript.

## 19 Key words

20 Genomic selection, sequencing depth, genetic effects, allele dosage, sweetpotato, potato

## 21 Key message

22 Polyploid crop breeders do not need more investment for sequencing depth, dosage information  
23 and fewer highly informative SNPs recommended, non-additive models and QTL advantages on  
24 prediction dependent on trait architecture.

## 25 Abstract

26 The autopolyploid nature of potato and sweetpotato ensures a wide range of meiotic  
27 configurations and linkage phases leading to complex gene action and pose problems in  
28 genotype data quality and genomic selection analyses. We used a 315-progeny biparental  
29 population of hexaploid sweetpotato and a diversity panel of 380 tetraploid potato, genotyped  
30 using different platforms to answer the following questions: i) do polyploid crop breeders need to  
31 invest more for additional sequencing depth? ii) how many markers are required to make  
32 selection decisions? iii) does considering non-additive genetic effects improve predictive ability

33 (PA)? iv) does considering dosage or quantitative trait loci (QTL) offer significant improvement  
34 to PA? Our results show that only a small number of highly informative single nucleotide  
35 polymorphisms (SNPs;  $\leq 1000$ ) are adequate for prediction, hence it is possible to get this  
36 number at the current sequencing depth from most service providers. We also show that  
37 considering dosage information and additive-effects only models had the best PA for most traits,  
38 while the comparative advantage of considering non-additive genetic effects and including  
39 known QTL in the predictive model depended on trait architecture. We conclude that genomic  
40 selection can help accelerate the rate of genetic gains in potato and sweetpotato. However,  
41 application of genomic selection should be considered as part of optimizing the entire breeding  
42 program. Additionally, since the predictions in the current study are based on single populations,  
43 further studies on the effects of haplotype structure and inheritance on PA should be studied in  
44 actual multi-generation breeding populations.

#### 45 **Data availability**

46 All single nucleotide polymorphism (SNP) data used in the current manuscript are provided with  
47 the manuscript as **Online Resources 2-4** while all best linear unbiased estimators (BLUEs) are  
48 provided as **Online Resources 5 and 6**.

#### 49 **Acknowledgements**

51 The sweetpotato research, most co-authors and GBSpoly genotyping was supported by a Bill &  
52 Melinda Gates Foundation grant (Grant number OPP1052983). DArTSeq genotyping of  
53 sweetpotato was done through the collaboration of the Integrated Genotyping Service and  
54 Support platform (IGSS) and DArT<sup>TM</sup> in Australia. Potato research and genotyping at Cornell  
55 was supported by a GIZ grant for the project ‘Accelerating the Development of Early-Maturing-  
56 Agile Potato for Food Security through a Trait Observation and Discovery Network’. The work  
57 was done as part of the Consultative Group on International Agricultural Research (CGIAR)-  
58 Research Program on Roots, Tubers and Bananas (RTB) which is supported by CGIAR Fund  
59 Donors (<http://www.cgiar.org/about-us/our-funders/>). Research teams at CIP and collaborating  
60 partners in China are greatly appreciated for providing phenotypic data.

#### 61 **Conflict of interest**

62 On behalf of all co-authors, the lead author declares no conflict of interest

63

64

## 65 **Introduction**

66 Phenotyping under recurrent selection has been an important approach for variety development  
67 in plant breeding, with substantial success to date. However, this process may take a long time  
68 for most crops, particularly for clonally propagated crops (**Slater et al. 2016**). For example, in  
69 potato, it typically takes an entire year to develop enough tubers from botanical seed obtained  
70 from crossing nurseries, for experimental trial purposes. This is followed by at least two years of  
71 field evaluation for qualitative traits, with evaluation for most quantitative traits in replicated  
72 multi-environment trials beginning in around year four (**Endelman et al. 2018**). The same can be  
73 said for sweetpotato, although cycle times in sweetpotato are shorter by about a year due to the  
74 fact that the crop can be vegetatively propagated via stem cuttings (**Wolfgang et al. 2009**). This  
75 represents a stark contrast with what can be achieved in cereal and legume crops, where up to 6  
76 generations can be raised within a calendar year (**Watson et al. 2018**), or in private corn  
77 breeding programs based in the United States and Europe which can raise multiple generations  
78 per year through the coordinated use of winter nurseries located in both hemispheres such as  
79 United States, Puerto Rico, Hawaii and Chile. This therefore implies that the estimation of  
80 parental value based on genetic designs and phenotypic evaluation in potato and sweetpotato  
81 increases the selection cycle time, thereby reducing the rate of genetic gains and the speed of  
82 delivery of superior, novel genetics to farmers.

83 The use of genetic markers for selection offers potential to reduce the breeding cycle time as  
84 selection can be done at an earlier stage. Previously proposed methods have involved identifying  
85 quantitative trait loci (QTL) via QTL mapping and genome-wide association studies (GWAS),  
86 but they have had little practical application in the actual development of new cultivars through  
87 plant breeding to date, especially for complex quantitative traits, since identifying the causal  
88 genes underlying QTL needed to make their application practical is costly (**Xu and Crouch**  
89 **2008**). Genomic selection (GS) offers the ability to select parents within a shorter interval and  
90 increase selection intensity by predicting untested genotypes earlier and enhancing larger starting  
91 genetic variation. This approach uses genome-wide marker data to predict the performance of  
92 untested genotypes and estimate their breeding values (genomic estimated breeding values;  
93 GEBVs), based on a genotyped and phenotyped training population (**Meuwissen et al. 2001**).

94 Genomic selection is emerging as the approach of choice to circumvent the limitations associated  
95 with use of QTL for marker-assisted selection and to improve the efficiency of phenotypic  
96 selection (**Bernal-Vasquez et al. 2014**). Good genetic progress can be made using GS, as long as  
97 factors that affect its predictive ability (PA), i.e. the correlation between phenotypic best linear  
98 unbiased estimators (BLUPs) and GEBVs, are well understood. These include trait architecture,  
99 the size of the training population, the relationship between the training and validation  
100 populations, heritability of the trait, the quality of phenotypic efforts, the level of linkage  
101 disequilibrium (LD), marker density, environmental variances and covariance among traits  
102 (**Covarrubias-Pazaran et al. 2018**).

103 The application of GS is taking shape in plant breeding with more and more crops exploring its  
104 utility (**Spindel et al. 2016; Wang et al. 2018; Endelman et al. 2018; Covarrubias-Pazaran et**  
105 **al. 2018; Faville et al. 2018; Nyine et al. 2018; Bhandari et al. 2019**). For crops like rice and  
106 wheat that are normally self-pollinated and have a high incidence of high-effect QTL (**Spindel et**  
107 **al. 2016**), faster success is expected from applying GS as prediction accuracy depends primarily  
108 on the factors listed above. However, breeders of auto-polyploid, clonally propagated crops like  
109 potato and sweetpotato, which are normally heterogenous and heterozygous, have to ask  
110 themselves additional questions and identify trade-off points that enhance the success of GS-  
111 assisted breeding (**Slater et al. 2016; Endelman et al. 2018**). Potato and sweetpotato present a  
112 wide range of meiotic configurations and linkage phases (**Mollinari et al. 2020**). In addition to  
113 causing complex gene action effects, allelic and configuration diversity have consequences on  
114 genotyping and genotype data quality, which consequently affects downstream analysis for  
115 quantitative-genetic parameters required to make high quality breeding decisions. Genotyping-  
116 by-sequencing (GBS) has currently become a genotyping method of choice in plant breeding  
117 (**Poland and Rife 2012**) but it is also prone to genotyping errors and a high level of missingness  
118 at low depth of sequencing, while high sequencing depth has additional cost implications. Data  
119 from polyploid crops is more prone to low quality genotype calls at low sequencing depth when  
120 compared to diploid crops, because of uncertain allele dosages and possibility of non-random  
121 inheritance of alleles such as in preferential pairing or double reduction (**Blischak et al. 2016,**  
122 **2018**).

123 Public sector breeding programs like those conducted in centers which are part of the  
124 Consultative Group on International Agricultural Research (CGIAR), and in the individual  
125 National Agricultural Research Systems (NARS) existing in many countries, are currently  
126 undergoing breeding program optimization efforts in order to keep up with the challenges of  
127 climate change and population increase (Cobb et al. 2019). Application of GS is one such tool  
128 for breeding program optimization. In order to develop GS tools to make more effective breeding  
129 efforts in auto-polyploid crops such as potato and sweetpotato, we have taken a practical  
130 perspective within a plant breeding setting to address several pertinent questions related to  
131 application of GS in auto-polyploids. We used real data sets from a 380 training-panel made up  
132 of advanced tetraploid potato clones and a 315-full-sib family of hexaploid sweetpotato, both  
133 developed by the International Potato Center (CIP) and genotyped using different platforms, to  
134 address the following questions: i) do polyploid crop breeders need to invest more resources for  
135 additional sequencing depth? ii) how many genetic markers are required to make selection  
136 decisions? iii) does the consideration of non-additive genetic effects add value to predictive  
137 ability (PA) to enhance genetic gains either for population improvement or product development  
138 in polyploid crops? iv) given the multiple alleles at loci with diverse meiotic configurations and  
139 linkage phases, does considering dosage, haplotypic or QTL effects offer significant  
140 improvement to PA to enhance genetic advances? We also discuss other factors that need to be  
141 considered while adopting GS as a decision support tool in an optimized breeding program.

## 142 **Materials and Methods**

### 143 **Genetic Materials and Phenotyping**

#### 144 *Sweetpotato bi-parental population*

145 A wide genetic variability exists in sweetpotato in terms of yield, nutritional content and culinary  
146 aspects, abiotic stress tolerance, biotic stress tolerance, among other attributes (Low et al. 2017).  
147 Introgression of high  $\beta$ -carotene content into locally adapted varieties is a major breeding  
148 objective especially in sub-Saharan Africa where vitamin A deficiency is prevalent. A 315-  
149 progeny full-sib family was developed by crossing a US-bred high  $\beta$ -carotene variety,  
150 ‘Beauregard’, with an adapted, locally preferred, starchy, low  $\beta$ -carotene landrace variety,  
151 ‘Tanzania’, at CIP-Peru. These two parents differ in additional traits of interest and the  
152 population will henceforth be referred to as the BT population. The population was evaluated in

153 six environments of Peru, and six environments of Uganda for various quality-related and yield-  
154 related traits as described by **Gemenet et al. (2020)** and **Pereira et al. (2019)**, between 2016 and  
155 2017. The full 315-progeny population was also genotyped together with the parents using an  
156 optimized protocol for hexaploid sweetpotato, ‘GBSpoly’ at North Carolina State University  
157 (NCSU). To support genotyping protocol optimization for the hexaploid, a diploid relative of  
158 sweetpotato, *Ipomoea trifida*, was used to develop a full-sib family of about 200. The family was  
159 developed from two *I. trifida* lines referred to as M9 and M19, hence the M9 x M19 population,  
160 and also genotyped at NCSU. Additionally, a sub-sample of 292-progeny and the two parents of  
161 the BT population were genotyped by DArTSeq<sup>TM</sup> in Australia, under the collaboration between  
162 the Integrated Genotyping Service and Support (IGSS) platform at the Biosciences east and  
163 central Africa (BecA) hub in Nairobi, Kenya and DArT. The quality-related traits measured in  
164 the BT population include: dry matter (DM) content, measured as a percentage of the laboratory  
165 dried samples divided by the initial fresh weight of 100g; Starch and  $\beta$ -carotene (BC) content,  
166 estimated using near-infrared reflectance spectroscopy (NIRS), and flesh color (FC), measured  
167 using internal color scales developed by CIP and partners. All quality-related traits were  
168 measured in Peru, but only flesh color was measured in Uganda (FC\_U). Data is further  
169 described in **Gemenet et al. (2020)**. For yield-related traits, total number of storage roots (TNR),  
170 number of commercial storage roots (NOCR), weight of total storage roots (RYTHA), weight of  
171 commercial storage roots (CYTHA), and total weight of foliage (FYTHA), were measured in the  
172 six experiments of Peru. Data is further described in **Pereira et al. (2019)**. Trait abbreviations  
173 are further defined in **Table 1**.

174 The quality-related traits were analyzed by fitting the following linear mixed model in  
175 ASREML:

$$176 \quad y_{ijkl} = \mu + g_i + e_l + r_{k(l)} + b_{jk(l)} + (ge_i)_l + \varepsilon_{ijk(l)},$$

177 where  $y_{ijkl}$  = the vector of phenotypes of the individual  $i$  in block  $j$  within replicate  $k$  of  
178 environment  $l$ ,  $\mu$  = population mean,  $g_i$  the fixed treatment (genotype) effect,  $e_l$  = the random  
179 effect of environment  $l$ ,  $r_{k(l)}$  = random effect of replicate  $k$  in environment  $l$ ,  $b_{jk(l)}$  = random  
180 effect of block  $j$  within replicate  $k$  of environment  $l$ ,  $(ge_i)_l$  = random effect of individual  $i$  in  
181 environment  $l$  ( $L=5$ ) environments,  $\varepsilon_{ijk(l)}$  = random error of the residuals, assuming  
182  $e_l \sim N(0, \sigma_e^2)$ ,  $r_{k(l)} \sim N(0, \sigma_r^2(l))$ ,  $b_{jk(l)} \sim N(0, \sigma_b^2(l))$ ,  $g_i \sim N(0, \Sigma)$  with  $\Sigma$  = variance-covariance

183 matrix across L environments, which varies according to the trait,  $\varepsilon_{ijk(l)} \sim N(0, \sigma_l^2)$  (**Gemenet et**  
184 **al. 2020**).

185 The yield-related traits were also analyzed with linear mixed models as described by **Pereira et**  
186 **al. (2019)** using GENSTAT 14 as:

$$y_{ijkl} = \mu + g_i + e_l + r_{k(l)} + b_{jkl} + ge_{il} + \varepsilon_{ijkl}$$

187 where  $y_{ijkl}$  = the vector of phenotypes as above,  $\mu$  = population mean,  $g_i$  the fixed treatment  
188 (genotype) effect,  $e_l$  = fixed effect of environment l,  $r_{k(l)}$  = fixed effect of replication k in  
189 environment l,  $b_{jkl}$  = random effect of block j within replication k in environment l;  
190  $b_{jkl} \sim N(0, \sigma_b^2)$ ,  $ge_{il}$  = the fixed interaction effect of individual i and environment l, and  
191  $\varepsilon_{ijk(l)} \sim N(0, \sigma^2)$  is the random residual error. The best linear unbiased estimators (BLUEs) as  
192 obtained by fitting the above models to the experimental data were then used to estimate GEBVs

### 193 ***Potato trait observation network population***

194 A 380-genotype panel made up of advanced clones from the potato breeding program and  
195 representing all breeding populations at CIP was assembled for a trait observation network  
196 (TON) in Peru, China and Ethiopia. Henceforth, we shall refer to this population as the TON  
197 panel. The evaluation of the panel was carried out in diverse agro-ecological zones, and in  
198 subsets of genotypes subject to participating NARS' partner capacity and/or ability to produce  
199 enough mini-tubers for experimentation. The experimental sites, experimental designs and the  
200 number of genotypes evaluated per experiment are summarized in **Table 2**. The TON panel was  
201 evaluated for maturity (bulking) by tuber characteristics at three harvest dates where average  
202 yield per plant (kg; AYP), weight of marketable tubers per plant (kg; WMT), were measured.  
203 Additionally, mature tuber weight was evaluated by measuring total tuber weight per plant  
204 (TTW; kg). In Peru, TTW was measured as the average total tuber weight across three drought-  
205 related treatments: terminal drought (irrigation stopped at flowering until harvest; TTW16\_TD),  
206 recovery (partially irrigated after drought stress; TTW16\_REC), and fully irrigated (normally  
207 irrigated throughout the growth period; TTW16\_NI), while random drought was used in China,  
208 with no controlled treatments. Resistance to potato virus Y (PVY) was evaluated after infection  
209 with virulent vectors and susceptible spreader rows using standard protocols at CIP, while late

210 blight resistance (LB) was evaluated by growing the population in endemic disease pressure and  
211 scored using standard protocols at CIP. Trait abbreviations are defined in **Table 1**. Additionally,  
212 the 380 genotypes were genotyped by GBS at Cornell University.

213 The experiments were analyzed as single trials, depending on the experimental design used as  
214 summarized in **Table 2**. A linear mixed model, taking into account the respective experimental  
215 design, was fitted to the phenotypic data. For those traits with different treatments like TTW in  
216 Peru, the joint adjusted means were additionally obtained across all treatments by fitting a linear  
217 mixed model. Genotype was considered as a fixed effect in these mixed models, so that BLUEs  
218 for the genotypic means were obtained for each trait and used to predict GEBVs.

## 219 **Genotyping and Variant Calling**

### 220 *DArTSeq<sup>TM</sup> for Sweetpotato*

221 DArTseq<sup>TM</sup> represents a combination of DArT complexity reduction methods and next  
222 generation sequencing platforms (**Kilian et al. 2012; Courtois et al. 2013; Raman et al. 2014;**  
223 **Cruz et al. 2013**). Therefore, DArTseq<sup>TM</sup> represents a new implementation of sequencing  
224 complexity reduced representations (**Altshuler et al. 2000**) and more recent applications of this  
225 concept on the next-generation sequencing platforms (**Baird et al. 2008; Elshire et al. 2011**).  
226 Similar to previous DArT methods based on array hybridizations, the technology is optimized for  
227 each organism and application by selecting the most appropriate complexity reduction method  
228 (both the size of the representation and the fraction of a genome selected for assays). Four  
229 methods of complexity reduction were tested in sweetpotato (data not presented) and the *PstI*-  
230 *MseI* method was selected. DNA samples were processed in digestion/ligation reactions  
231 principally as per **Kilian et al. (2012)** but replacing a single *PstI*-compatible adaptor with two  
232 different adaptors corresponding to two different Restriction Enzyme (RE) overhangs. The *PstI*-  
233 compatible adapter was designed to include Illumina flowcell attachment sequence, primer  
234 sequence and “staggered”, varying length barcode region, similar to the sequence reported by  
235 **Elshire et al. (2011)**. This reverse adapter contained a flowcell attachment region and a *MseI*-  
236 compatible overhang sequence. Only “mixed fragments” (*PstI-MseI*) were effectively amplified  
237 in 30 rounds of PCR using the following reaction conditions: i) 94 °C for 1 min, ii) 30 cycles of:  
238 94 °C for 20 sec, 58 °C for 30 sec, 72 °C for 45 sec, iii) 72 °C for 7 min. After PCR,  
239 equimolar amounts of amplification products from each sample of the 96-well microtiter plate

240 were bulked and applied to c-Bot (Illumina) bridge PCR followed by sequencing on Illumina  
241 HiSeq2000. The sequencing (single read) was run for 77 cycles. Sequences generated from each  
242 lane were processed using proprietary DArT analytical pipelines. In the primary pipeline the  
243 FastQ files were first processed to filter away poor-quality sequences, applying more stringent  
244 selection criteria to the barcode region compared to the rest of the sequence. This was to ensure  
245 reliability in the assignments of the sequences to specific samples carried in the “barcode split”  
246 step. Approximately 2,000,000 sequences per barcode/sample were identified and used in  
247 marker calling. Finally, identical sequences were collapsed into “fastqcoll files”. The fastqcoll  
248 files were “groomed” using DArT PL’s proprietary algorithm which corrects low quality base  
249 from singleton tag into a correct base using collapsed tags with multiple members as a template.  
250 The “groomed” fastqcoll files were used in the secondary pipeline for DArT PL’s proprietary  
251 SNP and SilicoDArT (presence/absence of restriction fragments in representation) calling  
252 algorithms (DArTsoft14). For SNP calling, all tags from all libraries included in the DArTsoft14  
253 analysis were clustered using DArT PL’s C++ algorithm at the threshold distance of 3, followed  
254 by parsing of the clusters into separate SNP loci using a range of technical parameters, especially  
255 the balance of read counts for the allelic pairs. Additional selection criteria were added to the  
256 algorithm based on analysis of approximately 1,000 controlled cross populations. Testing a range  
257 of tag counts parameters facilitated selection of true allelic variants from paralogous sequences.  
258 In addition, multiple samples were processed from DNA to allelic calls as technical replicates  
259 and scoring consistency was used as the main selection criteria for high quality/low error rate  
260 markers. Calling quality was assured by high average read depth per locus (>30X). The SNPs  
261 were coded as 0 = AA, 1 = BB, 2 = AB and “-“ = Missing. The sequences were not aligned to a  
262 reference genome because by the time of genotyping, the diploid references (**Wu et al. 2018**)  
263 had not been published.

#### 264 ***GBSPoly© for Sweetpotato***

265 GBSpoly is an optimized protocol for hexaploid sweetpotato developed at NCSU as part of a  
266 project focusing on developing genomic tools for sweetpotato improvement. The DNA was  
267 checked for quality on 1% agarose gel and quantified based on the PicoGreen fluorescence-  
268 based assay and the concentration was normalized to 50 ng/μl. Initially, several optimization  
269 efforts regarding restriction enzyme pairing was carried out (data not shown) and *CviAII-TseI*

270 was selected to be the best combination for hexaploid sweetpotato. Therefore, 1µg of DNA was  
271 double-digested using five units of *CviAII* for three hours at 25°C followed by digestion with  
272 *TseI* for another three hours at 65°C. A new England Biolabs (NEB) CutSmart buffer was used to  
273 make up a total volume of 30 µl. Purification of the digested samples was done using AMPure  
274 XP magnetic beads from ThermoFisher™ and quantified with PicoGreen assay. Barcodes were  
275 designed to account for substitution and indel errors and had an 8-bp buffer sequence to ensure  
276 that the barcode lay within high-quality base call regions of the sequence reads. Additional  
277 double digests on 64-plex pooled samples, purification, and size selection steps were carried out  
278 as described by **Wadl et al. (2018)** before performing 125 bp single-end sequencing on a total of  
279 40 sequencing lanes (8 lanes for each of the 5 libraries) of the Illumina HiSeq 2500 platform.  
280 The resultant FastQ files were aligned to reference genomes of two wild relatives of sweetpotato,  
281 *Ipomoea trifida* and *Ipomoea triloba*, and variant calling done using the GBSapp pipeline as  
282 described by **Wadl et al. (2018)**. The SNPs were coded according to the dosage of the alternative  
283 allele as 0 = AAAAAA, 1 = AAAAAB, 2 = AAAABB, 3 = AAABBB, 4 = AABBBB, 5 =  
284 ABBBBB, 6 = BBBBBB. The variant calling process is summarized in **Online Resource 1**.

### 285 ***GBS-Cornell for Potato***

286 The 380-genotype TON panel was genotyped by Cornell University using GBS in 2015. The  
287 DNA was digested with *EcoT221* restriction enzyme and 48-plex libraries were prepared for  
288 sequencing, using customized GBS protocols at Cornell. The resultant FastQ files were quality  
289 controlled and variant calling done using GATK HaplotypeCaller option (**Poplin et al. 2017**),  
290 disabling the duplicate read filter (this is recommended for GBS data) and using the joint  
291 genotyping -ERC GVCF mode. The reads were aligned to the potato genome reference  
292 sequenced from *S. tuberosum* group Phureja, line DM1-3 516 R44, a doubled monoploid (DM)  
293 via anther culture by the potato genome sequencing consortium (PGSC). Version  
294 PGSC\_DM\_v4.03 of the reference genome was used in alignment. The barcodes were removed  
295 using stacks and the ends were trimmed using trim-galore, followed by mapping to the reference  
296 using BWA. Resultant SAM files were processed using samtools and variants called using  
297 GATK Haplotype caller, targeting biallelic SNPs only. The SNPs were coded according to the  
298 dosage of the alternative allele as 0 = AAAA, 1 = AAAB, 2 = AABB, 3 = ABBB, and 4 =  
299 BBBB. The SNP filtration was done using bcftools allowing only for those SNPs with MAF of ≥

300 3%, missingness of  $\leq 2\%$ , average genotype quality (GQ)  $\geq 20$  and average allele sequencing  
301 depth (DP)  $\geq 16$ .

### 302 **Model comparison for predictive ability**

303 We used the AGHmatrix package (**Amadeu et al. 2016**) to develop kinship G-matrices  
304 partitioning genetic variation based on several gene-action models. For the BT population  
305 DArTSeq markers (sweetpotato) where we did not have dosage information, we developed an  
306 additive G-matrix according to **VanRaden (2008)**, here-in referred to as Add\_2x\_DArTseq, and  
307 a non-additive effects G-matrix according **Vitezica and colleagues (2013)**, herein referred to as  
308 NonAdd\_2x\_DArTSeq. For the BT population GBSpoly (sweetpotato) and TON population  
309 GBS-Cornell (potato) data where we had dosage information, we employed three models to  
310 develop the G-matrices: (i) modeling only additive effects, according to **VanRaden (2008)** here-  
311 in referred to as Add\_6x\_GBSpoly for sweetpotato and Add\_4x\_GBSCornell for potato, (ii)  
312 modeling additive plus non-additive effects, according to **Slater et al. (2016)** here-in referred to  
313 as Add+Non\_6x\_GBSpoly for sweetpotato and Add+Non\_4x\_GBSCornell for potato, and (iii) a  
314 pseudo-diploidized effect model according to **Slater et al. (2016)**, here-in referred to as  
315 Pseudo\_2x\_GBSpoly for sweetpotato and Pseudo\_2x\_GBSCornell. The pseudo-diploidization  
316 collapses all dosage classes between the nulliplex and the hexaplex (in sweetpotato), and  
317 between the nulliplex and tetraplex (in potato) into one heterozygous class, under the assumption  
318 that all heterozygotes have an equal effect which falls in between both homozygotes. In the case  
319 of potato, the design matrix coding for the pseudo-diploid, additive autotetraploid and full  
320 autotetraploid was as described by **Slater et al. (2016)**, while that for sweetpotato is shown in  
321 **Table 3**. During kinship matrix development, additional filters were applied to the genotype  
322 data, to have MAF  $\geq 30\%$ , and missing data  $\leq 10\%$ . We used genomic best linear unbiased  
323 prediction (G-BLUP; **Clark and van der Werf 2013**) to compare the predictive ability of the  
324 five models for sweetpotato and three models for potato using the kinship matrices as variance-  
325 covariance matrices to fit the compressed linear mixed model (**Zhang et al. 2010**) and estimate  
326 genomic best linear unbiased predictors (G-BLUPs). The software GAPIT (**Lipka et al. 2012**)  
327 was used in the G-BLUP prediction fitting the following general model:

$$y = 1_n\mu + Zu + e$$

328 where  $y$  = vector of phenotypic data,  $1_n$  is the vector of ones,  $\mu$  = population mean,  $Z$  = the  
329 known design matrix for genotypes,  $u$  = random genetic effects and  $\sim N(0, \sigma_a^2 K \text{ or } \sigma_{a+na}^2 K)$   
330 with  $K$  = kinship matrix,  $a$  = additive model,  $na$  = non-additive model,  $e$  = vector of residuals  
331  $\sim N(0, \sigma_e^2 I)$ .

332 Cross-validation was done by setting 20% of the population to missing phenotypes to be used as  
333 a validation set. We used 1,000 iterations to estimate the predictive ability of the models using  
334 both simple/oligo (quality traits in sweetpotato, disease traits in potato) and complex (storage  
335 root or tuber yield and yield component traits in both), as defined in **Table 1**.

336 Unlike in sweetpotato where phenotype and genotype data were balanced across  
337 experiments, (292 + Parents for DArTSeq and 315 + parents for GBSpoly), the potato  
338 experiments were unbalanced in terms of experimental genotypes. For the purposes of this study,  
339 we only selected the locations with the highest training population per trait. Consequently, we  
340 used AYP from Kunming (China; AYP\_K), WMT from Kunming (China; WMT\_K), LB from  
341 Oxapampa (Peru; LB2014\_O), LB from Yunnan (China; LB2016\_Y), PVY from Lima (Peru;  
342 PVY\_L), TTW averaged across three treatments of 2016 in Ica (Peru; TTW16\_Ica), and TTW in  
343 2016 from Heilongjiang (China; TTW16\_HLJ), all having number of genotypes indicated in  
344 **Table 2**. Differences in PA among models per trait were tested using t-tests. Quantitative-genetic  
345 parameters were tested for the additive model with or without dosage by obtaining the additive  
346 genetic variation ( $\sigma_a^2$ ) and random residual effects ( $\sigma_e^2$ ) from the mixed linear model and  
347 calculating narrow-sense heritability for each trait as:

$$348 \quad h^2 = \frac{\sigma_a^2}{(\sigma_a^2 + \sigma_e^2)}$$

349

350 Additionally, we calculated the estimated rate of genetic gains from genomic selection per  
351 additive model with or without dosage for each trait according to **Oliveira et al. (2019)** as:

$$\Delta GG = \frac{(i * \sigma_a * PA)}{L}$$

352 assuming  $L=5$  for sweetpotato following the accelerated breeding scheme currently implemented  
353 (**Mwanga et al. 2017**), and  $L= 8$  for potato.

## 354 **How many markers are adequate for prediction?**

355 For sweetpotato, we used the GBSpoly data, using different filtration criteria to end up with  
356 different number of markers. We used three criteria i) total number of SNPs filtered at 10% MAF  
357 and  $\geq 90\%$  call rate, ii) total number of SNPs filtered at 30% MAF and  $\geq 90\%$  call rate (used in  
358 the analyses above), and iii) A random sample of 15,000 SNPs from the total number of SNPs  
359 and filtered at 30% MAF and  $\geq 90\%$  call rate. In Potato, the total number of SNPs was filtered  
360 using two criteria: i) 30% MAF and  $\geq 90\%$  call rate, ii) 40% MAF and  $\geq 90\%$  call rate. The  
361 model considering only additive effects was used in comparing the effect of number of markers  
362 in sweetpotato, while all three models were tested between the two filtering criteria in potato.

## 363 **Incorporating haplotypic-QTL in prediction models for sweetpotato**

364 By taking advantage of the fully phased integrated linkage map from BT (Mollinari et al. 2020),  
365 we tested the predictive ability from QTL-informed models. Towards this end, we used the same  
366 cross-validation scheme as above, where 80%:20% random samples were used as training and  
367 testing populations, respectively, replicated 1,000 times. In order to detect QTL, we ran our  
368 random-effect multiple interval mapping (REMIM) using a sequential forward search (Pereira  
369 et al. 2019). We used score statistics to test map positions every 2 centiMorgans (cM) and added  
370 a QTL at a time using a relaxed genome-wide significance level threshold ( $\alpha = 0.20$ ). A window  
371 size of 20 cM was used to avoid that another position was selected very close to another QTL  
372 already in the model. For G-BLUP models, realized kinship matrices were based on the  
373 haplotype information from markers positioned every 2 cM in the genetic map. For QTL-BLUP  
374 (Q-BLUP), realized kinship matrices were based on the haplotypes from QTL peak marker; if  
375 there were more than one QTL, their kinship matrices were averaged out; if there were no QTL,  
376 we obtained the prediction as in G-BLUP. For Q+G-BLUP models, two terms were fitted, each  
377 with realized kinship matrices based on QTL peak markers (like for Q-BLUP) and the remaining  
378 markers in the linkage map but those selected as QTL.

379

## 380 **Results**

### 381 **SNP profiles from the genotyping platforms**

382 DArTseq sequencing of sweetpotato resulted in 13,504 biallelic SNPs (**Online Resource 2**). The  
383 call rates and polymorphic information content (PIC) are shown in **Fig. 1A&B** and ranged from  
384 about 0.4 - 1.0, with a mean of 0.96 for call rate and from 0 - 0.5 with a mean of 0.37 for PIC.  
385 Stringent filtering at a call rate  $\geq 0.8$  and PIC  $\geq 0.25$  left 9,649 SNPs that were used in  
386 AGHMatrix. Additional filtration in AGHMatrix at  $\leq 0.1$  missingness and  $\geq 30\%$  MAF resulted  
387 in 6,015 diploidized, biallelic SNPs being used to develop the matrices following additive  
388 (Add\_2x\_DArTSeq) and nonadditive (NonAdd\_2x\_DArTSeq) models.

389 Cornell GBS in potato resulted in 295,401 biallelic SNPs at the variant calling step that were  
390 then hard-filtered to 3,262 high confidence SNPs by setting MAF  $\geq 0.03$ , missing  $\leq 2\%$  and  
391 average read depth (DP)  $\geq 16$  (**Online Resource 3**). The 3,262 SNP profiles are shown in **Fig.**  
392 **1C&D** showing MAF ranging from 0.03 - 0.5, with a mean of 0.15 and PIC ranging from 0.0 -  
393 0.5, with a mean of 0.23. The 3,262 SNPs were used in the AGHMatrix relationship matrix  
394 development. For a relative comparison of models across crops for trait groups, we also filtered  
395 the Cornell GBS data in AGHMatrix at  $\leq 0.1$  missingness and  $\geq 30\%$  MAF as done for DArTSeq  
396 data above, which resulted in 411 SNPs used to develop the additive (Add\_4x\_GBSCornell),  
397 additive plus non-additive (Add+Non\_4x\_GBSCornell) and the pseudo-diploidized  
398 (Pseudo\_2x\_GBSCornell) models. Examining the relationship matrices indicated that at MAF  $\geq$   
399 0.3, the full model (Add+Non\_4x\_GBSCornell) was mainly monomorphic. For potato therefore,  
400 we also changed the MAF to  $\geq 0.4$ , which resulted in 178 SNPs that were used to develop a  
401 second set of relationship matrices. All PA comparisons among traits for potato are based on this  
402 matrix.

403 For GBSpoly in sweetpotato called according to **Wadl et al (2018)**, comparing diploid genotype  
404 data from M9 x M19 diploid population and hexaploid BT data showed that for the same level of  
405 genotype quality as for the diploid at about 25x depth of coverage, we needed  $\geq 100x$  depth of  
406 coverage (**Fig. 2**). Consequently, for sweetpotato, GBSpoly data was filtered to this high depth of  
407 coverage, with MAF  $\geq 0.05$ . This resulted in 34,390 high confidence SNPs (**Online Resource 4**)  
408 that were used in AGHMatrix to develop the additive (Add\_6x\_GBSpoly), additive plus non-  
409 additive (Add+Non\_6x\_GBSpoly) and the pseudo-diploidized (Pseudo\_2x\_GBSpoly)  
410 relationship matrices. The filters in AGHMatrix were set to  $\leq 0.1$  missing and  $\geq 0.3$  MAF as for  
411 the preceding data types and resulted in a final 2,883 SNPs that developed the matrices for model

412 comparison. For comparing the effects of number of markers on PA, the first filtration criteria of  
413 10% MAF and  $\geq 90\%$  call rate resulted in 10,358 SNPs, while the third criteria based on a  
414 random sample of 15,000 SNPs resulted in 1,291 SNPs that were used in PA comparison, based  
415 on the additive only model.

416 The comparison of models with(out) QTL and use of markers *per se* or haplotypes was carried  
417 out using 30,684 SNPs from the same genotyping platform and data set, filtered and processed as  
418 described by **Mollinari et al. (2020)**, which were used to develop a 2,708.4 cM phased genetic  
419 linkage map for sweetpotato, and subsequent QTL analyses (**Pereira et al. 2019; Gemenet et al.**  
420 **2019**). Sweetpotato BLUEs are provided in **Online Resource 5** while potato BLUEs are  
421 provided as **Online Resource 6**.

#### 422 **Genotyping platforms, genetic effects and predictive ability**

423 In sweetpotato, the diploidized additive model (Add\_2x\_DArTSeq) using data from DArTSeq  
424 performed equally highly as or sometimes better than the additive model using high confidence  
425 dosage data from GBSpoly (Add\_6x\_GBSpoly), depending on trait architecture, for simpler  
426 quality-related traits (**Fig. 4**). DM had 0.33 and 0.44, Starch had 0.32 and 0.38, BC had 0.43 and  
427 0.43, FC\_P had 0.44 and 0.45, while FC\_U had 0.41 and 0.38 average PA for Add\_2x\_DArTSeq  
428 and Add\_6x\_GBSpoly models, respectively (**Table 4**). For these traits, additive only models  
429 were the best and the full model (Add\_Non\_6x\_GBSpoly) always had negative PA due to a  
430 largely monomorphic relationship matrix. Nonetheless, the situation changed with yield-related  
431 traits as the effects of dosage and non-additive effects became more important. For these traits,  
432 the high-quality data with dosage from GBSpoly (Add\_6x\_GBSpoly) was always better in  
433 prediction when compared to the additive model with diploidized data (Add\_2x\_DArTSeq).  
434 NOCR had 0.19 and 0.31, TNR had 0.25 and 0.37, CYTHA had 0.18 and 0.22, RYTHA had  
435 0.18 and 0.23, FYTHA had 0.21 and 0.26 average PA for Add\_2x\_DArTSeq and  
436 Add\_6x\_GBSpoly additive-models, respectively (**Table 4**). However, the additive only model  
437 with dosage (Add\_6x\_GBSpoly) was not always the best in PA for all yield-related traits,  
438 especially not for storage roots traits CYTHA and RYTHA, where it performed similar to either  
439 or both of the models considering non-additive effects whether with dosage  
440 (Add+Non\_6x\_GBSpoly) or without dosage (NonAdd\_2x\_DArTseq) (**Fig. 3**). Nevertheless, the  
441 largely monomorphic relationship matrix from the full model (Add+Non\_6x\_GBSpoly) ensured

442 low predictive ability using this model for most yield-related traits as well, especially FYTHA  
443 which had the highest negative PA, (collapsed to zero in Fig.3, for plotting purposes). In general,  
444 pseudo-diploidizing data already called with dosage (Pseudo\_2x\_GBSpoly) drastically reduced  
445 PA even more than using data called as diploid (DArTseq). In potato, the situation was not very  
446 different as the diploidized additive model (Pseudo\_2x\_GBSCornell) was the second-best model  
447 after the additive only model with dosage (Add\_4x\_GBSCornell) for simpler disease traits and  
448 its comparative advantage significantly reduced with more complex traits (**Fig. 4**). LB2014\_O  
449 had 0.68 and 0.63, LB2016\_Y had 0.62 and 0.52, PVY\_L had 0.54 and 0.50, AYP\_K had 0.45  
450 and 0.34, WMT\_K had 0.48 and 0.34, TTW16\_Ica had 0.16 and 0.16, while TTW16\_HLJ had  
451 0.37 and 0.32 average PA for Add\_4x\_GBSCornell and Pseudo\_2x\_GBSCornell, respectively.  
452 As with the full model in sweetpotato, the model including non-additive effects  
453 (Add+Non\_4x\_GBSCornell) was the least performing in terms of PA (**Table 4**).

#### 454 **Number of markers and environments**

455 Our results in potato indicated that an increased number of markers by more than double did not  
456 have a significant effect on PA (411 vs 178 SNPs; **Fig. 4A**) considering the best predictive  
457 model (Add\_4x\_GBSCornell). LB2014\_O had 0.69 and 0.68, LB2016\_Y had 0.66 and 0.62,  
458 PVY\_L had 0.59 and 0.54, AYP\_K had 0.51 and 0.45, WMT had 0.51 and 0.48, TTW16\_Ica  
459 had 0.19 and 0.16, while TTW16\_HLJ had 0.40 and 0.37 average PA for 411 and 178 SNPs  
460 respectively. Similarly, in sweetpotato, comparing PA using 10,358 SNPs, 2,883 SNPs and 1,291  
461 SNPs using the best predictive model (Add\_6x\_GBSpoly) showed no effect of increasing marker  
462 density at the cost of marker informativeness on PA. PA based on 10,358 SNPs which had 10%  
463 MAF generally performed lower than 2,883 and 1,291 SNPs which both had 30% MAF (**Fig. 5**).  
464 Additionally, 2,883 SNPs did not have a clear comparative advantage over 1,291 SNPs (**Fig. 5**).  
465 Regarding traits in different locations, environmental effects on PA were observed, though the  
466 magnitude of such effects was also dependent on trait architecture. The PA based on the best  
467 model for FC\_P (0.45; Peru) and FC\_U (0.38; Uganda) in sweetpotato and LB2014\_O (0.68;  
468 Peru) and LB2016\_Y (0.62; China), in potato, though a bit different, were both relatively high to  
469 allow meaningful selections for the trait. For more complex yield traits, the PA for TTW16\_Ica  
470 (0.16; Peru) and TTW16\_HLJ (0.37; China) were significantly different.

#### 471 **Effects of quantitative trait loci, haplotypes and dosage on predictive ability**

472 We additionally tested three analysis models using BT sweetpotato data: i) Q-BLUP based on  
473 relationship matrices from QTL-peak haplotypes, ii) Q+G-BLUP fitting two terms based on  
474 QTL-peak haplotypes and the rest of the markers in the linkage map, iii) G-BLUP, predictions  
475 using markers spaced every 2 cM in the genetic map without considering QTL. The PA results  
476 are shown in **Fig. 6**. Considering QTL haplotypes either *per se* (Q-BLUP) or with G-BLUP  
477 (Q+G-BLUP) had a clear comparative advantage for PA in simpler traits. However, this  
478 comparative advantage faded with more complex yield-related traits. Our results therefore show  
479 that with genomic selection, the comparative advantage of using the linkage map information  
480 and QTL is dependent on trait architecture, hence the magnitude of QTL effects that can be  
481 mapped (**Fig. 6**).

#### 482 **Genetic variation, heritability and estimated rate of genetic gain**

483 Given that the additive effects only model with dosage performed better for most traits in both  
484 sweetpotato and potato (Add\_6x\_GBSpoly and Add\_4x\_GBSCornell, respectively), we  
485 evaluated quantitative genetic parameters for this model in comparison with the additive model  
486 without dosage for both crops (Add\_2x\_DArTseq for sweetpotato and Pseudo\_6x\_for potato).  
487 Narrow sense heritability ( $h^2$ ) ranged from 0.24-0.66 for the model with dosage  
488 (Add\_6x\_GBSpoly) and 0.13-0.62 for the model without dosage (Add\_2x\_DArTseq) in  
489 sweetpotato. In potato, ( $h^2$ ) ranged from 0.07 – 0.49 in the model with dosage  
490 (Add\_4x\_GBSCornell) and 0.10 – 0.46 in the model with pseudo-diploidized dosages (Pseudo-  
491 2x\_GBSCornell; **Table 4**). As expected, traits with simpler architecture (quality-related traits in  
492 sweetpotato; disease traits in potato had the highest ( $h^2$ ) compared to more complex yield-  
493 related traits. All models across crops resulted in positive estimated genetic gain considering L=  
494 5 years in sweetpotato and L= 8 in potato, which are the cycle lengths of current breeding  
495 schemes at CIP (**Table 4**). This implies that more genetic gains can be realized if such breeding  
496 cycle lengths are further significantly reduced.

497

#### 498 **Discussion**

499 **Low-cost, targeted amplicon sequencing platforms could realize faster genetic gains per**  
500 **unit time**

501 Having a reliable, cost-efficient genotyping platform that ensures faster data turn-around to  
502 breeding programs on time to impact selection and advancement decisions is a must for routine  
503 application of genomic selection in plant breeding programs. Here we have compared results  
504 based on data from three GBS-based platforms, two of which provide data at the commercial  
505 diploid sequencing depth level (DArTSeq and GBS-Cornell). About 100x read depth was  
506 required to confidently call all the five heterozygous dosage classes of sweetpotato, against 25-  
507 30x required for the diploid. These results agree with studies in potato where **Uitdewilligen et al.**  
508 **(2013)** reported that 60-80x depth was required to confidently call the three heterozygote classes.  
509 GBSpoly (**Wadl et al. 2018**) which had high quality dosage data in our study was developed as  
510 part of a project to understand optimal conditions for GBS in hexaploid sweetpotato and  
511 therefore not amenable to routine use in plant breeding. Other options for more precise  
512 genotyping such as SNP arrays, in addition to issues with ascertainment biases, are crop-specific  
513 and therefore do not benefit from economies of scale that drive costs down. Breeding programs  
514 of polyploid crops therefore have to weigh whether investing more for higher depth of  
515 sequencing is an efficient resource allocation strategy (**Endelman et al. 2018**). To this end,  
516 although our results show that genotype quality and consequently the number of realized SNPs is  
517 lower with low allele sequencing depth, we also show as described in the next sections that only  
518 a small number of highly informative SNPs are required to realize relatively high PA depending  
519 on the trait. These results agree with the findings of **Chang et al. (2019)** who showed that PA  
520 can be improved by prioritizing relevant SNP polymorphisms. This therefore implies that for  
521 practical plant breeding applications, using established genotyping platforms that ensure low-  
522 costs due to scale effects and faster data turn-around will have better likelihood of success in  
523 routine application of genomic selection in polyploids despite the low allele sequencing depths.  
524 Since both crops already have GBS-based SNPs at high density, the process can be fast-tracked  
525 by targeting the high informative segregating loci in amplicon sequencing. This is encouraging  
526 as polyploid crops in developing countries with limited access to expensive, high quality  
527 genotypic datasets could also deploy GS approaches.

528 **A few highly informative SNPs segregating in the population are adequate for prediction**  
529 **purposes**

530 **Guo et al. (2018)** found that at allele sequencing depth between 10x to 20x, between 80-100K  
531 SNPs would be required to accurately predict additive breeding values in tetraploid ryegrass. Our  
532 research in both hexaploid sweetpotato and tetraploid potato however shows a reduced number  
533 of realized SNPs after quality filtration, which can be attributed to the difficulty of genotyping  
534 polyploid crops. SNP calling in polyploids is further complicated by the presence of polymorphic  
535 positions across homologues within and among individuals in addition to the polymorphic  
536 positions within a single homologue among individuals (**Clevenger et al. 2015**). In our potato  
537 example, the initial filtration of SNPs to allele sequencing depth at  $DP \geq 16$  and MAF 3%  
538 resulted in only 3,262 SNPs. The same scenario was observed for sweetpotato. However, our  
539 results also show that if SNPs are highly informative ( $MAF \geq 30\%$ ), a number as low as 178  
540 SNPs could give relatively high PA comparable to a larger number of SNPs. In potato, 178 SNPs  
541 at  $MAF \geq 40\%$  performed relatively similar as 411 SNPs at  $MAF \geq 30\%$ . Not shown results from  
542 a preliminary analysis of the same dataset of potato using  $MAF \geq 10\%$  resulted in 1,710 SNPs  
543 whose PA did not differ significantly with the PA using either 411 or 178 SNPs. Additionally, in  
544 sweetpotato, 2,883 SNPs at  $MAF \geq 30\%$  gave the same or better PA as 10,358 SNPs at  $MAF \geq$   
545  $10\%$ , and 1,291 SNPs. Our results therefore agree with the findings of **Covarrubias-Pazaran et**  
546 **al. (2018)** using three biparental populations of the American cranberry, that addition of SNPs  
547 after 500 markers did not result in much increase in PA as only a few hundred SNPs were needed  
548 to reach PA plateau. Even though their study used a consensus map to intentionally distribute  
549 markers evenly across the genome, our random sampling method based on MAF and PIC came  
550 to the same conclusion. These results imply that breeding programs with limited resources for  
551 genotyping can target few highly informative regions within the genome that are segregating in  
552 their breeding populations via targeted genotyping methods following amplicon sequencing  
553 techniques, as a cost-effective way of incorporating genomic selection in their breeding  
554 programs. We propose the use of between 500-1000 highly informative SNPs for routine  
555 prediction purposes in a breeding program.

#### 556 **Modelling non-additive genetic effects has negligible contribution to predictive ability**

557 Our results both in potato and sweetpotato show that additive effects-only models, whether  
558 diploidized or with dosage, were comparatively better in PA than the models considering non-  
559 additive effects for all simple traits. This comparative advantage however lessened with more

560 complex traits, where non-additive effects and inclusion of dosage information became slightly  
561 more relevant, although in most cases the additive effects-only model with dosage still remained  
562 the best in terms of PA. This finding makes sense in quantitative genetic terms as the more the  
563 number of genes affecting a trait, the more the expected interaction among loci. In sweetpotato  
564 for example, issues of ‘missing’ heritability have been established for yield-related traits using  
565 the current BT population in multiple environments, where only a few QTL with very small  
566 effects were reported even though a very dense, well phased hexaploid genetic map was used  
567 (**Pereira et al. 2019; Gemenet et al. 2020**). According to **Varona et al. (2018)**, the contribution  
568 of non-additive effects to genetic variance depends on the allele frequency of the causative loci,  
569 and their consideration in breeding programs can improve the prediction accuracy for breeding  
570 values and inform cross-combinations that maximize non-additive variation in progeny. Several  
571 studies have however shown that inclusion of non-additive effects in the prediction models have  
572 negligible effects in improving the accuracy of predicting breeding (additive) values. For  
573 instance, **Endelman et al. (2018)** reported uncertainty in partitioning non-additive genetic  
574 variance in tetraploid potato, whereas **Crow (2010)**, suggested that variance due to epistasis  
575 would have little effects in plant breeding as additive variance and covariance effects quickly  
576 overshadow such contribution following selection. Non-additive effects are mainly considered  
577 important in genomic prediction (prediction for performance of different traits based on the  
578 genotype of the individual), while additive-only methods as important in genomic selection  
579 (prediction of parental value of an individual), because only additive effects can be passed from  
580 parents to progeny (**Varona et al. 2018**). However, our results, supported by previous findings in  
581 other crops, imply that in light of the large number of moving parts to consider, including  
582 concerns with genotyping platforms and genotype quality for polyploids, practical breeding  
583 programs for potato and sweetpotato, and perhaps other polyploid crops, will achieve more  
584 advances considering only the infinitesimal model (additive) for both genomic selection and  
585 genomic prediction.

586 **The relative importance of considering dosage, haplotypes and quantitative trait loci is**  
587 **dependent on trait architecture**

588 **Oliveira et al. (2019)** showed that the relative advantage of including dosage information to PA  
589 is dependent on trait architecture. Our results confirm this and show that for simple traits

590 diploidized data, especially when the genotypic data are directly called as diploid during variant  
591 calling e.g. the DArTSeq data in sweetpotato rather than pseudo-diploidizing data already called  
592 with dosage e.g. in GBS-Cornell data in potato, would just do fine. However, as the traits  
593 become more complex, considering dosage improves PA and therefore the rate of progress that  
594 can be made for such traits. **Endelman et al. (2018)** also showed that not considering allele  
595 dosage effects in potato reduced prediction accuracy by about 0.13 on average using data from  
596 the SolCAP potato SNP array, where they reported PA ranging from 0.06 to 0.63 for specific  
597 gravity, yield and fry color. Given that most traits are quantitative, we recommend the use of  
598 data with dosage even though they may come from sequencing platforms with low allele  
599 sequencing depth, that could benefit more with improved genotype calling methods, such as  
600 Bayesian genotype calling methods.

601 Our data also shows that for all traits, considering both QTL and haplotypes resulted in the best  
602 PA especially for simple traits, although this comparative advantage also faded with more  
603 complex yield traits. Having markers in complete LD with causative QTL for a given trait is a  
604 prerequisite for improving PA in genomic prediction (**Velasco et al. 2019**). The study of  
605 **Cuyabano et al. (2014)** showed that considering haplotype blocks rather than single markers  
606 improved PA for dairy traits in cattle. This is because haplotypes are supposed to be in tighter  
607 LD with QTL than single markers. This can be attributed to the fact that GS-only GBLUP  
608 methods use the average genome information relationship for model building and for prediction  
609 whereas incorporating QTL analysis gives different weights (QTL effects) to different  
610 “significant” genome positions (QTL positions) for model building and for prediction. Due to  
611 this, studies have proposed a combination of QTL mapping to explain trait architecture and  
612 genomic prediction, to improve PA (**Spindel et al. 2016; Lopes et al. 2017; Morgante et al.**  
613 **2018; Bhandari et al. 2019**). Our results however indicate that the relative advantage of  
614 considering QTL-based haplotypes is dependent on trait architecture and directly related to the  
615 number and effect size of the QTL in question. In this case, yield-related traits did not show  
616 much improvement in PA when QTL were considered. Despite this finding, additional efforts in  
617 studying the effect of haplotype structure on PA is recommended to increase the likelihood of  
618 fully recovering the polyploid genetic information, where the information from individual dosage  
619 markers can be rather limited. However, given that QTL mapping/GWAS methods require high  
620 density markers, the application of such a strategy should be considered in the context of the cost

621 of developing high density markers against available resources for genotyping in a given  
622 program. Additionally, such methods would be computationally demanding and should also be  
623 considered depending on available computational tools and analytical capacity of a given  
624 program.

### 625 **Further considerations for optimized breeding programs using genomic selection**

626 The PA of genomic selection is influenced by several factors including trait architecture, the size  
627 of the training population, the relationship between the training and validation populations,  
628 heritability of the trait, the level of linkage disequilibrium (LD), marker density, environmental  
629 variances and covariance among traits (**Nakaya and Isobe 2012**). In addition to the already  
630 discussed factors, our results indicate that environment plays a significant role in determining PA  
631 as can be seen in the same traits measured across several environments. Additionally, PA  
632 magnitude even for simple traits were lower in sweetpotato where we used BLUEs across six  
633 environments, than in potato where predictions were made per single environment. Models  
634 incorporating genotype-x-environment interaction are important and more realistic when  
635 predicting performance of untested genotypes across environments (**Burgueno et al. 2012;**  
636 **Heslot et al. 2014; Wang et al. 2018**). Furthermore, PA for complex yield-related traits were  
637 always lower than for simpler quality-related or disease traits. PA for such complex traits have  
638 been shown to benefit from multi-trait selection models incorporating simpler, correlated traits  
639 with the primary trait (**Covarrubias-Pazaran et al. 2018; Michel et al. 2019**). Additionally,  
640 **Bernal-Vasquez et al. (2014)** alluded to the fact that phenotypic data analysis contributed to  
641 improved PA, which speaks to the necessary precision and accuracy of the phenotype in training  
642 populations. Taken together, the current results show that genomic selection will contribute  
643 towards increased genetic gains, especially via reduced breeding cycle time in potato and  
644 sweetpotato. However, the effectiveness of genomic selection will have to be considered from  
645 the perspective of optimizing the entire breeding program (**Cobb et al. 2019**). This refers to the  
646 assembly and deployment of a package of technological tools that allow a specific program to  
647 realize maximum genetic gains within its current context in terms of time and resources, by  
648 exploiting all components of the breeder's equation. Therefore, given the diversity existing from  
649 program to program in terms of resources and human capacity, no 'one size fits all' scenario is  
650 anticipated.

651 Finally, it does not escape to our attention that the predictions here-in are based on single  
652 populations. However, plant breeding requires several levels of allele recombination through  
653 generations. We cannot estimate from the current data, how such recombination complexity will  
654 affect the efficiency of GS in breeding programs. Additional studies estimating PA in actual  
655 multigeneration breeding populations therefore need to be carried out to reliably estimate the  
656 value of GS to potato and sweetpotato, and perhaps other polyploid breeding programs.

## 657 **Figure Captions**

658 **Fig. 1** Quality attributes of the SNP profiles from DArTSeq (call rate and polymorphic  
659 information content; PIC) data in sweetpotato and GBSCornell (minor allele frequency; MAF  
660 and PIC) in potato

661 **Fig. 2** Comparison of genotype quality at different allele sequencing depths in diploid *I. trifida*  
662 (M9xM19) and hexaploid sweetpotato (*I. batatas*; BT)

663 **Fig. 3** Boxplots comparing predictive ability of additive-effects-only models without dosage  
664 (Add\_2x\_DArTseq) and with dosage (Add\_6x\_GBSpoly); models considering also non-additive  
665 effects (NonAdd\_2x\_DArTSeq; Add+Non\_6x\_GBSpoly); and pseudo-diploidized dosage data  
666 (Pseudo\_2x\_GBSpoly) for quality related traits (A; DM = dry matter, starch, BC =  $\beta$ -carotene,  
667 FC\_P = flesh color in Peru; FC\_U = flesh color in Uganda); and yield related traits (B; NOCR =  
668 number of commercial storage roots, TNR = total number of storage roots, CYTHA = weight of  
669 commercial storage roots, RYTHA = weight of total storage roots, FYTHA = total weight of  
670 foliage) in a full-sib family of sweetpotato.

671 **Fig. 4** Box plots comparing predictive ability of additive-effects-only model  
672 (Add\_4x\_GBSCornell); additive and non-additive effects (Add+Non\_4x\_GBSCornell); and  
673 pseudo-diploidized dosage data (Pseudo\_2x\_GBSCornell); using minimum allele frequency  
674 (MAF)  $\geq$  30% (A; 411 SNPs) and MAF  $\geq$  40% (B; 178 SNPs). LB2014\_O = late blight in  
675 Oxapampa (Peru) in 2014, LB2016\_Y = late blight in Yunnan (China) in 2016, PVY\_L = potato  
676 virus Y in Lima (Peru), AYP\_K = average yield per plant in Kunming (China), WMT\_K =  
677 weight of marketable tubers in Kunming, TTW16\_Ica = total tuber weight in Ica (Peru) in 2016  
678 across three drought treatments, TTW16\_HLJ = total tuber weight in Heilongjiang (China) in  
679 2016, single treatment, in potato.

680 **Fig. 5** Box plots comparing the effect of number of markers on predictive ability using additive-  
681 effects only model (Add\_6x\_GBSpoly) with 10,358 SNPs, 2,883 SNPs and 1,291 SNPs in  
682 sweetpotato. A; DM = dry matter, starch, BC =  $\beta$ -carotene, FC\_P = flesh color in Peru; FC\_U =  
683 flesh color in Uganda; and yield related traits: B; NOCR = number of commercial storage roots,  
684 TNR = total number of storage roots, CYTHA = weight of commercial storage roots, RYTHA =  
685 weight of total storage roots, FYTHA = total weight of foliage in a full-sib family of  
686 sweetpotato.

687 **Fig. 6** Boxplots comparing predictive ability of models using QTL haplotypes only in prediction  
 688 (Q-BLUP); QTL combined with prediction based on markers per se, (Q+G-BLUP); prediction  
 689 using markers per se without QTL (G-BLUP) for quality-related traits (A; DM = dry matter,  
 690 starch, BC =  $\beta$ -caroten, FC\_P = flesh color in Peru; FC\_U = flesh color in Uganda); and yield  
 691 related traits (B; NOCR = number of commercial storage roots, TNR = total number of storage  
 692 roots, CYTHA = weight of commercial storage roots, RYTHA = weight of total storage roots,  
 693 FYTHA = total weight of foliage) in a full-sib family of sweetpotato.

#### 694 **Online Resource Captions**

695 **Online Resource 1** Variant calling pipeline used in the GBSapp for calling GBSpoly data in  
 696 sweetpotato

697 **Online Resource 2** DArTSeq SNP data for the Beaugard x Tanzania (BT) sweetpotato full-sib  
 698 family

699 **Online Resource 3** GBS-Cornell SNP data for the trait observation network (TON) potato  
 700 population

701 **Online Resource 4** GBSpoly SNP data for the Beaugard x Tanzania (BT) sweetpotato full-sib  
 702 family

703 **Online Resource 5** Best linear unbiased estimators for sweetpotato traits used in genomic  
 704 prediction in the current study

705 **Online Resource 6** Best linear unbiased estimators for potato traits used in genomic prediction  
 706 in the current study

707 **Table 1. Trait abbreviations and their description as used in the current study**

Crop	Trait Abbreviation	Trait Description
Sweetpotato	DM	Dry matter content
	Starch	Starch content
	BC	Beta-carotene
	FC_P	Flesh color in Peru
	FC-U	Flesh color in Uganda
	NOCR	# commercial storage roots
	TNR	# total storage roots
	CYTHA	Commercial storage root weight
	RYTHA	Total storage root weigh
	FYTHA	Total foliage yield weight
Potato	LB2014_O	Late blight in 2014 in Oxapampa, Peru
	LB2016_Y	Late blight 2016 in Yunnan, China
	PVY_L	Potato virus Y in Lima, Peru
	AYP_K	Average yield per plant in Kunming, China
	WMT_K	Weight of marketable tubers in Kunming, China
	TTW16_lca	Total tuber weight in 2016 in Ica-Peru
	TTW16_HLJ	Total tuber weight in 2016 in Heilongjinag, China

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

**Table 2. Locations, designs and traits measured in the trait observation network (TON) panel of potato**

Country		Location	Agroecology	
Peru		Lima, La Molina 12.0820° S, 76.9282° W	Lowland sub-tropics	
		Ica, Ica 14.0755° S, 75.7342° W		
		Pasco, Oxapampa 10.5853° S, 75.4053° W	Highland tropics	
China		Yunnan, Kunming 24.8801° N, 102.8329° E	Mixed agriculture systems, lowland & highland	
		Heilongjian, Harbin 45.8038° N, 126.5350° E	Temperate (long day)	
Trait Group	Trait	Location, Country, Year	Trial Design	#Genotype
Late Blight resistance	LB2014_O	Oxapampa, Peru, 2014	RCBD	241
	LB2016_Y	Yunnan, China, 2016	RCBD	336
Virus resistance	PVY_L	Lima, Peru, 2016-2018	RCBD	341
Bulking-based maturity	AYP_K	Kunming, China, 2016	RCBD	317
	WMT_K	Kunming, China, 2016	RCBD	317
Mature tuber weight	TTW16_Ica	Ica, Peru, 2016	Augmented	269
	TTW16_HLJ	Heilongjiang, China, 2016	Augmented	300

725

726

**Table 3** Proposed design matrix coding for auto-hexaploid sweetpotato as adapted from Slater et al. 2016.

	Pseudo_2x	Add_6x	Add+Non_6x						
Effects/Marker	1	1	1	2	3	4	5	6	7
AAAAAA	0	0	1	0	0	0	0	0	0
AAAAAB	1	1	0	1	0	0	0	0	0
AAAABB	1	2	0	0	1	0	0	0	0
AAABBB	1	3	0	0	0	1	0	0	0
AABBBB	1	4	0	0	0	0	1	0	0
ABBBBB	1	5	0	0	0	0	0	1	0
BBBBBB	2	6	0	0	0	0	0	0	1

727

728

729

730

731

732 **Table 4** Summary quantitative-genetic parameters derived from genomic selection with cross  
 733 validation applying different genetic effects models in sweetpotato and potato.  $\sigma_a^2$  is the additive  
 734 genetic variation,  $\sigma_e^2$  the residual variance,  $h^2$  the narrow-sense heritability,  $PA$  the predictive  
 735 ability,  $\Delta GG$  the estimated rate of genetic gain considering the current breeding cycle length

Crop	Trait <sup>a</sup>	Model <sup>b</sup>	$\sigma_a^2$	$\sigma_e^2$	$h^2$	$PA$	$\Delta GG$
Sweetpotato	DM	Add_2x_DArTSeq	1.6935	2.9536	0.36	0.33b	0.085889
		Add_6x_GBSpoly	4.0035	2.0762	0.66	0.44a	0.176077
	Starch	Add_2x_DArTSeq	6.1716	13.7616	0.31	0.32b	0.158993
		Add_6x_GBSpoly	12.1683	11.5424	0.53	0.38a	0.265111
	BC	Add_2x_DArTSeq	150.2336	113.1697	0.57	0.43a	1.0541
		Add_6x_GBSpoly	225.1431	152.0581	0.60	0.43a	1.29041
	FC_P	Add_2x_DArTSeq	0.5416	0.3304	0.62	0.44a	0.064762
		Add_6x_GBSpoly	0.8168	0.4189	0.66	0.45a	0.081339
	FC_U	Add_2x_DArTSeq	12.9633	10.9257	0.54	0.41a	0.295238
		Add_6x_GBSpoly	16.1489	16.777	0.49	0.38ab	0.305411
	NOCR	Add_2x_DArTSeq	50134102	2.93E+08	0.15	0.19b	269.0607
		Add_6x_GBSpoly	1.36E+08	2.43E+08	0.36	0.31a	722.4699
	TNR	Add_2x_DArTSeq	1.86E+08	7.40E+08	0.20	0.25b	681.9091
		Add_6x_GBSpoly	4.71E+08	5.83E+08	0.45	0.37a	1606.149
	CYTHA	Add_2x_DArTSeq	8.6149	27.7157	0.13	0.18b	0.105664
		Add_6x_GBSpoly	8.6149	26.6061	0.24	0.22a	0.129145
	RYTHA	Add_2x_DArTSeq	4.6249	31.4849	0.13	0.18b	0.07742
		Add_6x_GBSpoly	10.9811	29.4989	0.27	0.23a	0.152434
FYTHA	Add_2x_DArTSeq	7.678	26.0083	0.23	0.21b	0.116379	
	Add_6x_GBSpoly	12.8721	26.6023	0.33	0.26a	0.186564	
Potato	LB2014_O	Add_4x_GBSCornell	0.0189	0.0193	0.49	0.68a	0.011686
		Pseudo_2x_GBSCornell	0.0195	0.023	0.46	0.63b	0.010997
	LB2016_Y	Add_4x_GBSCornell	0.0191	0.0259	0.42	0.62a	0.010711
		Pseudo_2x_GBSCornell	0.0166	0.0323	0.34	0.52b	0.008375
	PVY_L	Add_4x_GBSCornell	0.0419	0.0738	0.36	0.54a	0.013817
		Pseudo_2x_GBSCornell	0.0364	0.0818	0.31	0.50ab	0.011924
	AYP_K	Add_4x_GBSCornell	0.0118	0.0327	0.27	0.45a	0.00611
		Pseudo_2x_GBSCornell	0.0066	0.0389	0.15	0.34b	0.003453
	WMT_K	Add_4x_GBSCornell	0.0132	0.0322	0.29	0.48a	0.006893
		Pseudo_2x_GBSCornell	0.0069	0.0392	0.15	0.34b	0.00353
	TTW16_Ica	Add_4x_GBSCornell	2.00E-04	0.0028	0.07	0.16a	0.000283
		Pseudo_2x_GBSCornell	3.00E-04	0.0027	0.10	0.16a	0.000346
	TTW16_HLJ	Add_4x_GBSCornell	0.0061	0.018	0.25	0.37a	0.003612
		Pseudo_2x_GBSCornell	0.0049	0.0192	0.20	0.32b	0.0028

736 <sup>a</sup>Traits as defined in Table 1, <sup>b</sup>Models: Add\_2x\_DArTseq = additive model using data from DArTseq called as  
 737 diploid; Add\_6x\_GBSpoly= additive model using data with dosage from GBSpoly; Add\_4x\_GBSCornell = additive

738 model using data with dosage from GBS at Cornell, Pseudo\_2x\_GBSCornell = additive model using data from GBS  
739 Cornell with three heterozygote classes collapsed into one.

740

## 741 **References**

742 Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, et al (2000) An SNP map of the  
743 human genome generated by reduced representation shotgun sequencing. *Nature* 407:  
744 513-516

745 Amadeu RR, Cellon C, Olmstead JW, Garcia AA, Resende MF, Muñoz PR (2016) AGHmatrix:  
746 R package to construct relationship matrices for autotetraploid and diploid species: A  
747 blueberry example. *Plant Genome* doi: 10.3835/plantgenome2016.01.0009

748 Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al (2008) Rapid SNP discovery and  
749 genetic mapping using sequenced RAD markers. *PLoS One* 3: e3376

750 Bernal-Vasquez A-M, Möhring J, Schmidt M, Schönleben M, Schön C-C, Piepho H-P (2014)  
751 The importance of phenotypic data analysis for genomic prediction - a case study  
752 comparing different spatial models in rye. *BMC Genomics* 15:646

753 Bhandari A, Bartholome J, Cao- Hamadoun T-V, Kumari N, Frouin J, Kumar A, et al (2019)  
754 Selection of trait-specific markers and multi-environment models improve genomic  
755 predictive ability in rice. *PLoS ONE* 14(5): e0208871

756 Blischak PD, Kubatko LS, Wolfe AD (2016) Accounting for genotype uncertainty in the  
757 estimation of allele frequencies in autopolyploids. *Mol Ecol Resources* 16:742–754

758 Blischak PD, Kubatko LS, Wolfe AD (2018) SNP genotyping and parameter estimation in  
759 polyploids using low-coverage sequencing data. *Bioinformatics* 34(3):407–415

760 Burgueño J, de los Campos G, Weigel K, Crossa J (2012) Genomic prediction of breeding values  
761 when modeling genotype x environment interaction using pedigree and dense molecular  
762 markers. *Crop Sci* 52:707–719

763 Chang L-Y, Toghiani S, Aggrey SE, Rekaya R (2019) Increasing accuracy of genomic selection  
764 in presence of high-density marker panels through the prioritization of relevant  
765 polymorphisms. *BMC Genetics* 20:21

- 766 Clark SA, van der Werf J (2013) Genomic best linear unbiased prediction (gBLUP) for the  
767 estimation of breeding values. *Methods Mol Biol* 1019:321-330
- 768 Clevenger J, Chavarro C, Pearl SA, Ozias-Akins P, Jackson SA (2015) Single nucleotide  
769 polymorphism identification in polyploids: A review, example, and recommendations.  
770 *Mol Plant* 8:831–846
- 771 Cobb JN, Juma RU, Biswas PS, Arbelaez JD, Rutkoski J, Atlin G, Hagen T, Quinn M, Ng EH  
772 (2019) Enhancing the rate of genetic gain in public-sector plant breeding programs:  
773 lessons from the breeder’s equation. *Theor Appl Genet* 132:627-645
- 774 Courtois B, Audebert A, Dardou A, Roques S, Ghneim-Herrera T, et al (2013) Genome-wide  
775 association mapping of root traits in a japonica rice panel. *PLoS ONE* 8(11): e78037
- 776 Covarrubias-Pazaran G, Schlautman B, Diaz-Garcia L, Grygleski E, Polashock J, et al (2018)  
777 Multivariate GBLUP improves accuracy of genomic selection for yield and fruit weight  
778 in biparental populations of *Vaccinium macrocarpon* Ait. *Front. Plant Sci.* 9: 1310
- 779 Crow, J. F. (2010). On epistasis: why it is unimportant in polygenic directional selection. *Philos.*  
780 *Trans. R. Soc. Lond. B Biol. Sci.* 365, 1241–1244
- 781 Cruz VM, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic  
782 diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related  
783 species. *PLoS One* 8(5): e64062
- 784 Cuyabano BCD, Su G Lund MS (2014) Geneomic prediction of genetic merit using LD-based  
785 haplotypes in the Nordic Holstein population. *BMC Genomics* 15:1171
- 786 Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al (2011) A robust, simple  
787 genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*,  
788 6(5):e19379
- 789 Endelman JB, Carley CAS, Bethke BC, Coombs JJ, Clough ME et al (2018). Genetic variance  
790 partitioning and genome-wide prediction with allele dosage information in tetraploid  
791 potato. *Genetics* 209:77-87

- 792 Faville MJ, et al (2018) Predictive ability of genomic selection models in a multi-population  
793 perennial ryegrass training set using genotyping-by-sequencing. *Theor Appl Genet*  
794 131:703-720
- 795 Gemenet DC, Pereira GDS, De Boeck B, Wood JC, Mollinari M, Olukolu BA et al (2020)  
796 Quantitative trait loci and differential gene expression analyses reveal the genetic basis  
797 for negatively-associated  $\beta$ -carotene and starch content in hexaploid sweetpotato  
798 [*Ipomoea batatas* (L.) Lam.]. *Theor App Genet* 133:23-36
- 799 Grüneberg W, Mwangi R, Andrade M, Espinoza J (2009) Breeding clonally propagated crops.  
800 In FAO, selection methods: chapter 13, part 5
- 801 Guo X, Cericola F, Fè D, Pedersen MG, Lenk I, Jensen CS, Jensen J and Janss LL (2018)  
802 Genomic prediction in tetraploid ryegrass using allele frequencies based on genotyping  
803 by sequencing. *Front Plant Sci* 9:1165
- 804 Heslot N, Akdemir D, Sorrells ME, Jannink JL (2014) Integrating environmental covariates and  
805 crop modeling into the genomic selection framework to predict genotype by environment  
806 interactions. *Theor Appl Genet* 127:463–480
- 807 Kilian A et al (2012) Diversity Arrays Technology: A generic genome profiling technology on  
808 open platforms. *Methods Mol Biol* 888:67-89
- 809 Lipka AE, Tian F, Wang Q, Peiffer J, Li M et al (2012) GAPIT: genome association and  
810 prediction integrated tool. *Bioinformatics* 28:2397–2399
- 811 Lopes MS, Bovenhuis, H, van Son M, Nordbø Ø, Grindflek EH, Knol EF, Bastiaansen, JWM  
812 (2017) Using markers with large effect in genetic and genomic predictions, *J Animal Sci*  
813 95(1):59–71
- 814 Low JW, Mwangi ROM, Andrade M, Carey E, Ball A (2017) Tackling vitamin A deficiency  
815 with biofortified sweetpotato in sub-Saharan Africa. *Global Food Secur* 14:23–30
- 816 Meuwissen TH, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using  
817 genome-wide dense marker maps. *Genetics* 157:1819–1829

- 818 Michel S, Löschenberger F, Ametz C, Pachler B, Sparry E, and Bürstmayr H (2019)  
819 Simultaneous selection for grain yield and protein content in genomics-assisted wheat  
820 breeding. *Theor Appl Genet* 132:1745–1760
- 821 Mollinari M, Olokulu B, Pereira GDS, Khan A, Gemenet DC, Yencho C, Zeng Z-B (2020)  
822 Unraveling the hexaploid sweetpotato inheritance using ultra-dense multilocus mapping.  
823 *G3: Genes, Genomes, Genetics* 10(1):281-292
- 824 Morgante F, Huang W, Maltecca C, Mackay TFC (2018) Effect of genetic architecture on the  
825 prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity*  
826 120:500-514
- 827 Mwanga ROM, Andrade MI, Carey EM, Low JW, Yencho GC, Grüneberg WJ (2017)  
828 Sweetpotato (*Ipomoea batatas* L.). In: *Genetic Improvement of Tropical Crops*. Campos  
829 H, Caligari PDS (Eds). Springer. p 181-218
- 830 Nakaya A, Isobe N. (2012). Will genomic selection be a practical method for plant breeding?  
831 *Ann Bot* 110(6):1303-1316
- 832 Nyine M, Uwimana B, Blavet N, Hřibová E, Vanrespaille H, Batte M, Akech V, Brown A,  
833 Lorenzen J, Swennen R, Doležel J (2018) Genomic prediction in a multiploid crop:  
834 Genotype by environment interaction and allele dosage effects on predictive ability in  
835 banana. *Plant Genome* 11:170090
- 836 Oliveira IDB, Resende Jr MFR, Ferrão LFV, Amadeu RR, Endelman JB, Matias Kirst M,  
837 Coelho ASG, Munoz PR (2019) Genomic prediction of autotetraploids; influence of  
838 relationship matrices, allele dosage and continuous genotype calls in phenotype  
839 prediction. *G3: Genes Genomes Genetics* 9:1189-1198
- 840 Pereira GDS, Gemenet DC, Mollinari M, Olukolu BA, Diaz F, Mosquera V, Gruneberg WJ,  
841 Khan A, Yencho GC, Zeng Z-B (2019) Multiple QTL mapping in autopolyploids: a  
842 random-effect model approach with application in a hexaploid sweetpotato full-sib  
843 population. *BioRxiv Preprint* doi: <https://doi.org/10.1101/622951>
- 844 Poland JA, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant*  
845 *Genome J* 5:92–102

- 846 Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, Kling  
847 DE, et al (2017) Scaling accurate genetic variant discovery to tens of thousands of  
848 samples. BioRxiv Preprint. DOI:10.1101/201178
- 849 Raman H, Raman R, Kilian A, Detering F, Carling J, et al (2014) Genome-wide delineation of  
850 natural variation for pod shatter resistance in *Brassica napus*. PLoS ONE 9(7): e101673
- 851 Slater AT, et al (2016) Improving genetic gain with genomic selection in autotetraploid potato.  
852 The Plant Genome 9(3):1-15
- 853 Spindel JE, Begum H, Akdemir D, Collard B, Redoña E, Jannink J-L, McCouch S (2016)  
854 Genome-wide prediction models that incorporate de novo GWAS are a powerful new  
855 tool for tropical rice improvement. Heredity 116:395–408
- 856 Uitdewilligen JG, AM Wolters BB, D'hoop TJ, Borm RG, Visser et al (2013) A next-generation  
857 sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid  
858 potato. PLoS One 8: e62355
- 859 VanRaden PM (2008) Efficient methods to compute genomic predictions. Journal of Dairy Sci  
860 91(11):4414-4423
- 861 Varona L, Legarra A, Toro MA, Vitezica ZG (2018) Non-additive Effects in Genomic Selection.  
862 Front Genet 9:78
- 863 Velazco JG, Malosetti M, Hunt CH, Mace ES, Jordan DR, van Eeuwijk FA (2019) Combining  
864 pedigree and genomic information to improve prediction quality: an example in sorghum.  
865 Theor Appl Genet 132:2055-2067
- 866 Vitezica ZG, et al (2013) On the additive and dominant variance and covariance of individuals  
867 within the genomic selection scope. Genetics 195:1223–1230
- 868 Wadl PA, Olukolu BA, Branham SE, Jarret RL, Yencho GC, et al (2018) Genetic Diversity and  
869 Population Structure of the USDA Sweetpotato (*Ipomoea batatas*) Germplasm  
870 Collections Using GBSpoly. Front Plant Sci 9:1–13
- 871 Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: Current  
872 status and prospects. The Crop Journal 6:330-340

873 Watson A, Ghosh S, Williams M. *et al* (2018) Speed breeding is a powerful tool to accelerate  
874 crop research and breeding. *Nature Plants* 4:23–29 (2018)

875 Wu S, Lau KH, Cao Q, Hamilton JP, Sun H, Zhou C *et al* (2018) Genome sequences of two  
876 diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement.  
877 *Nature communications* 9:4580

878 Xu Y, Crouch JH (2008) Marker-assisted selection in plant breeding: from publications to  
879 practice. *Crop Sci* 48:391–407

880 Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK *et al* (2010) Mixed linear model approach  
881 adapted for genome-wide association studies. *Nat Genet* 42:355–360

882

883

884

885

886

887

888

889

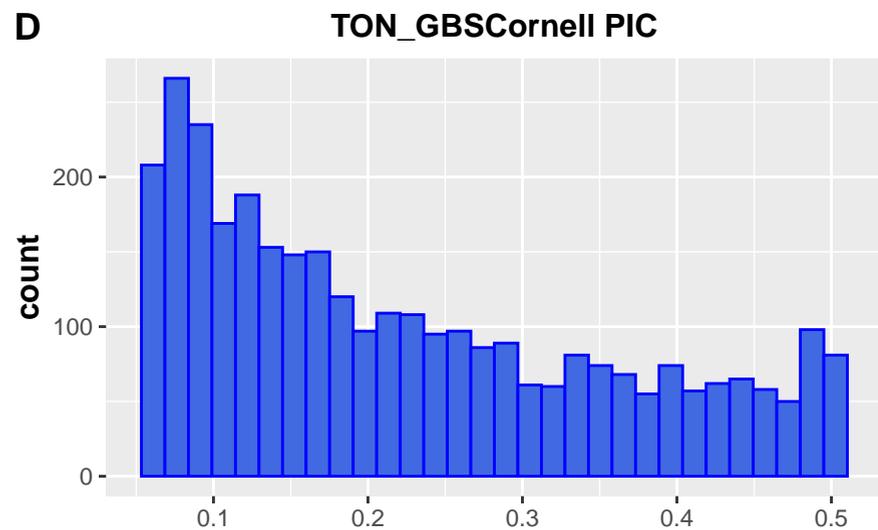
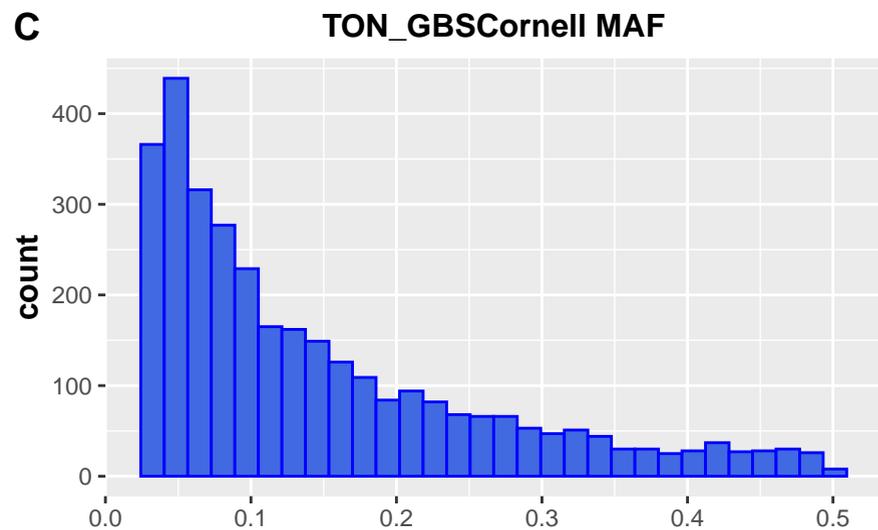
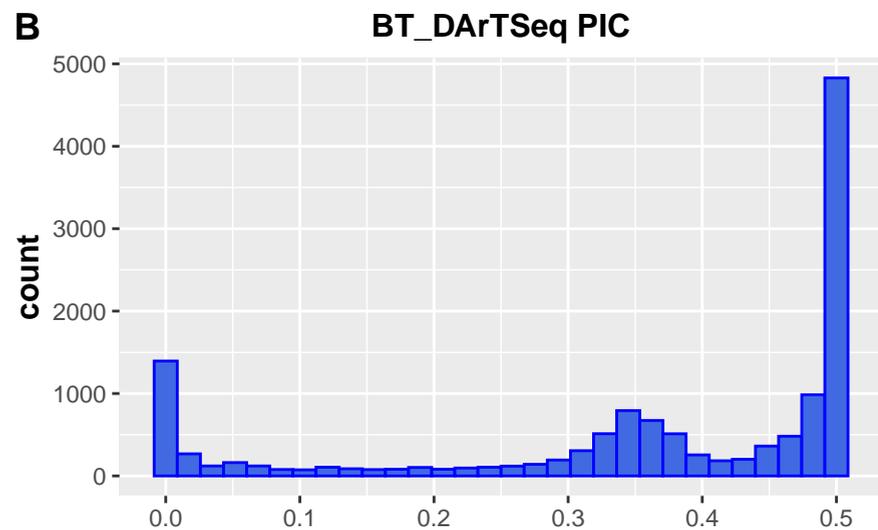
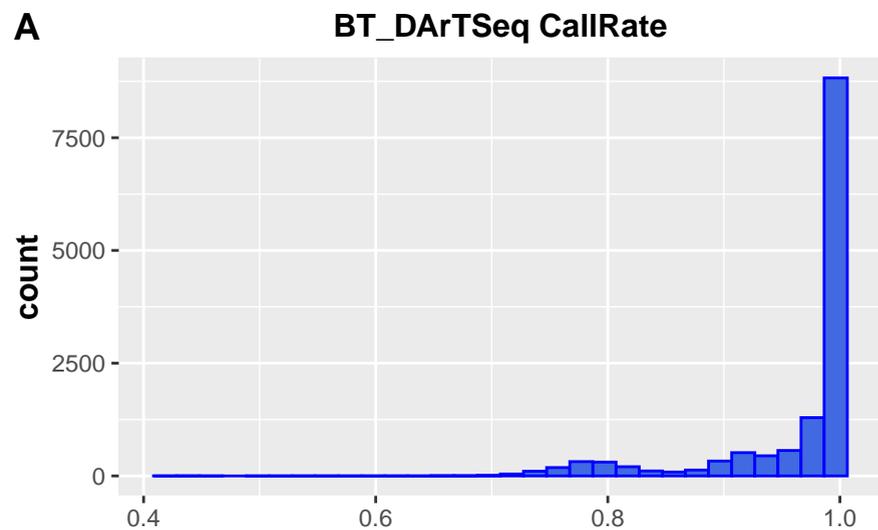
890

891

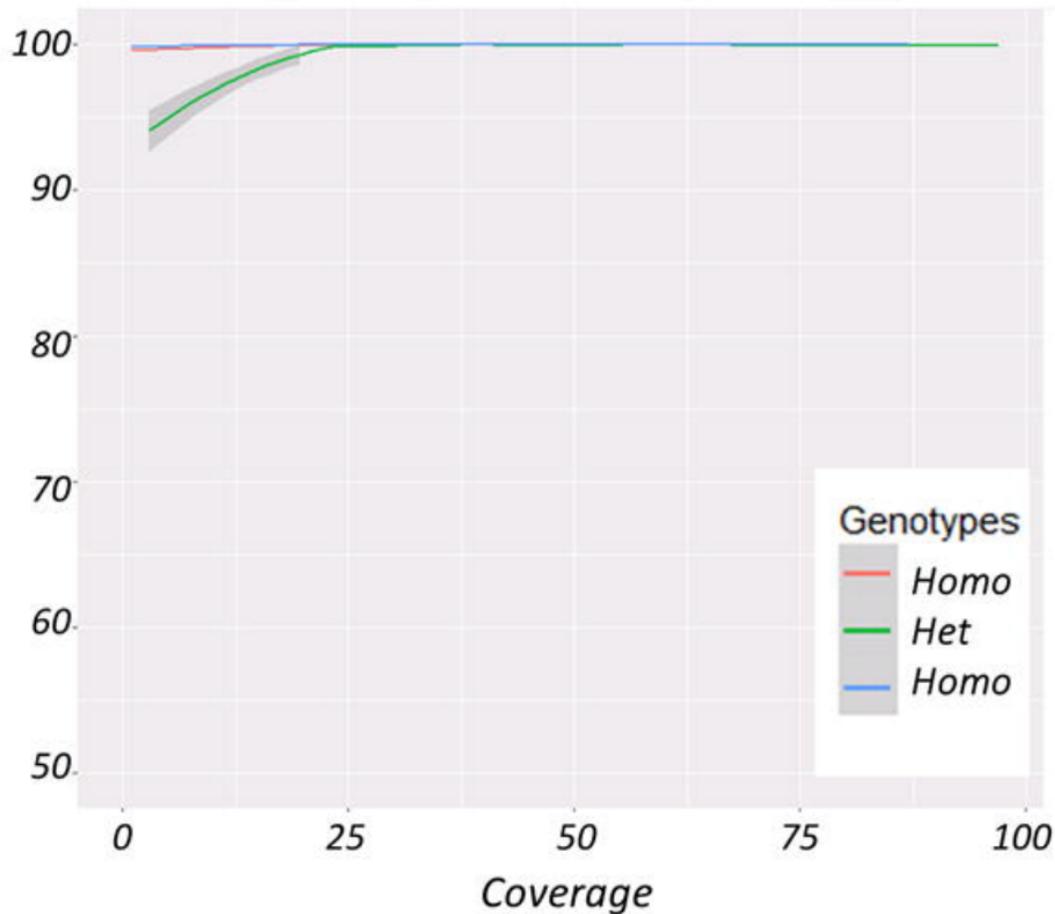
892

893

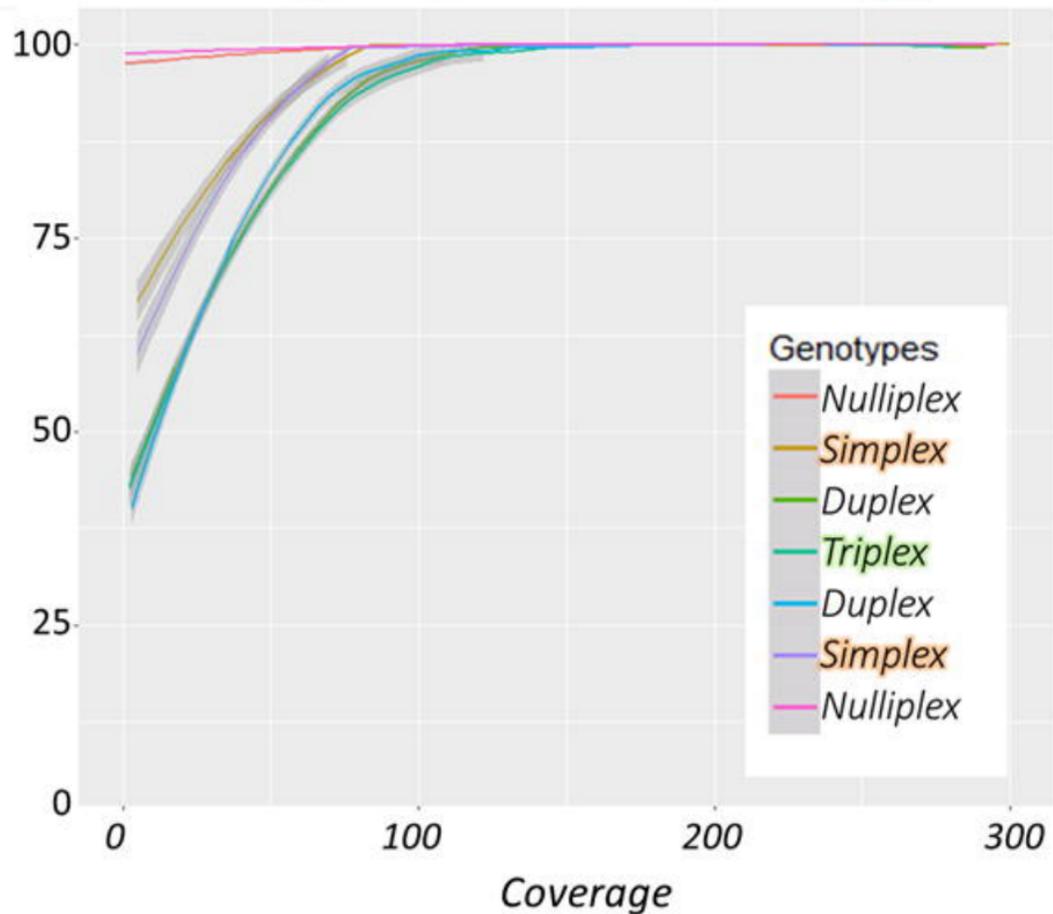
894

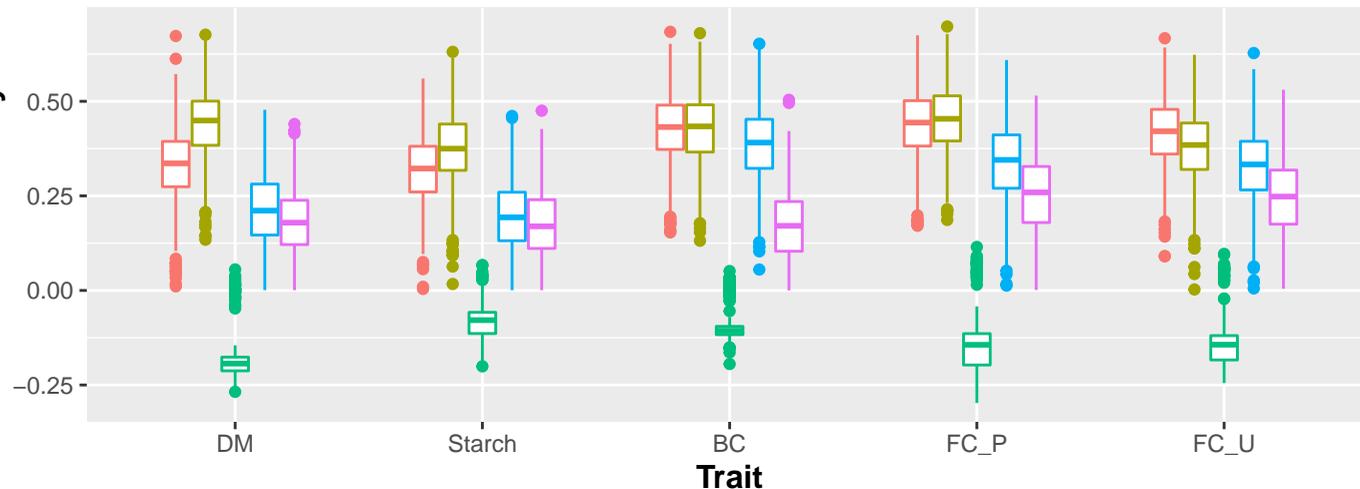


*Diploid (18.8 % are paralogs)*

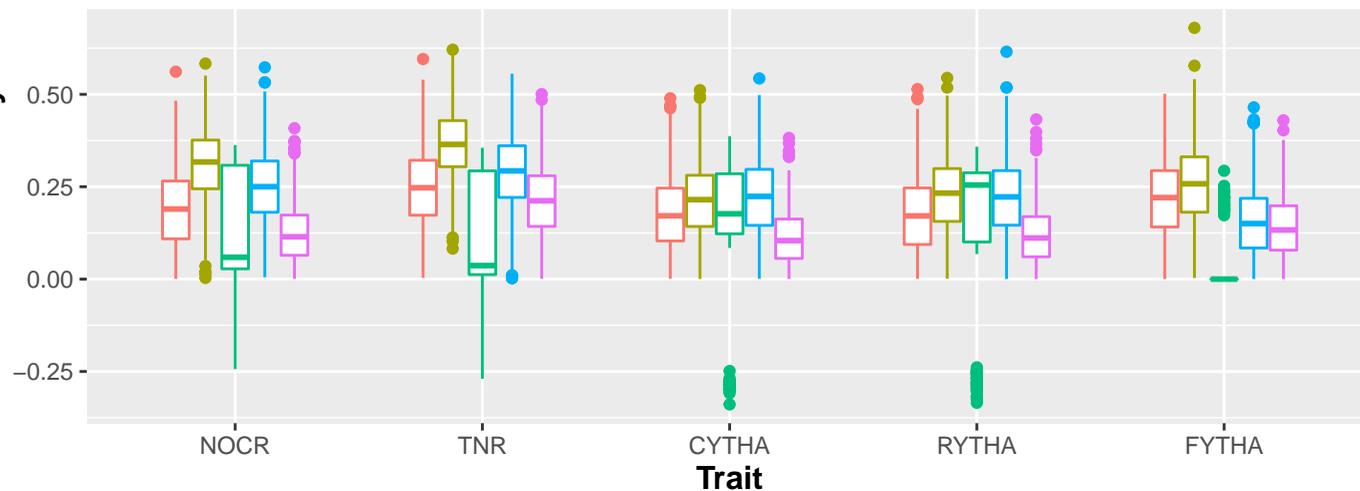


*Hexaploid (27.5 % are paralogs)*

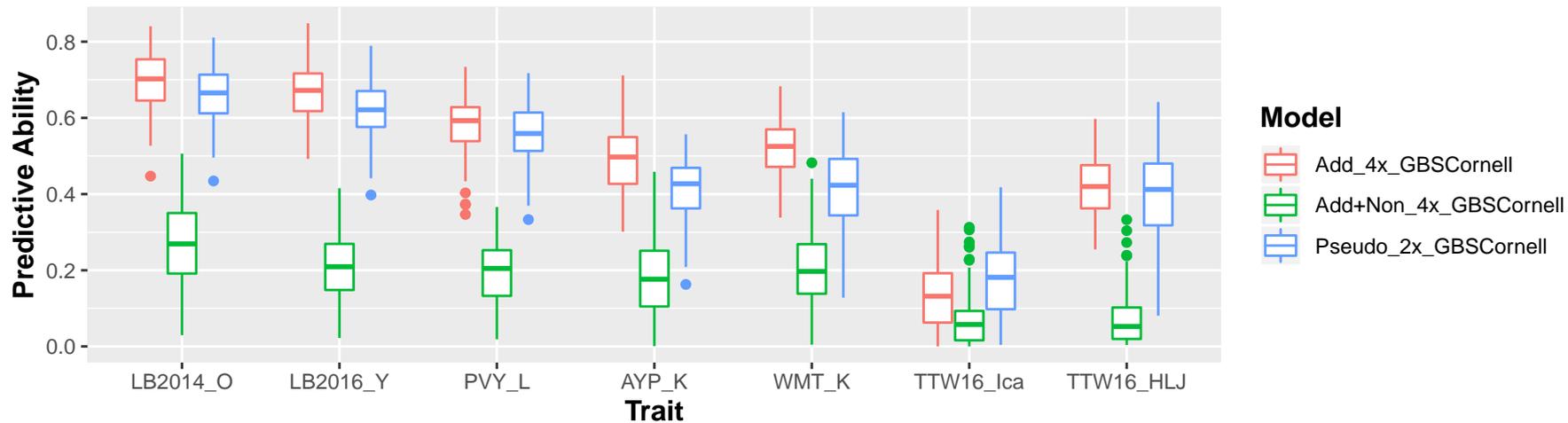
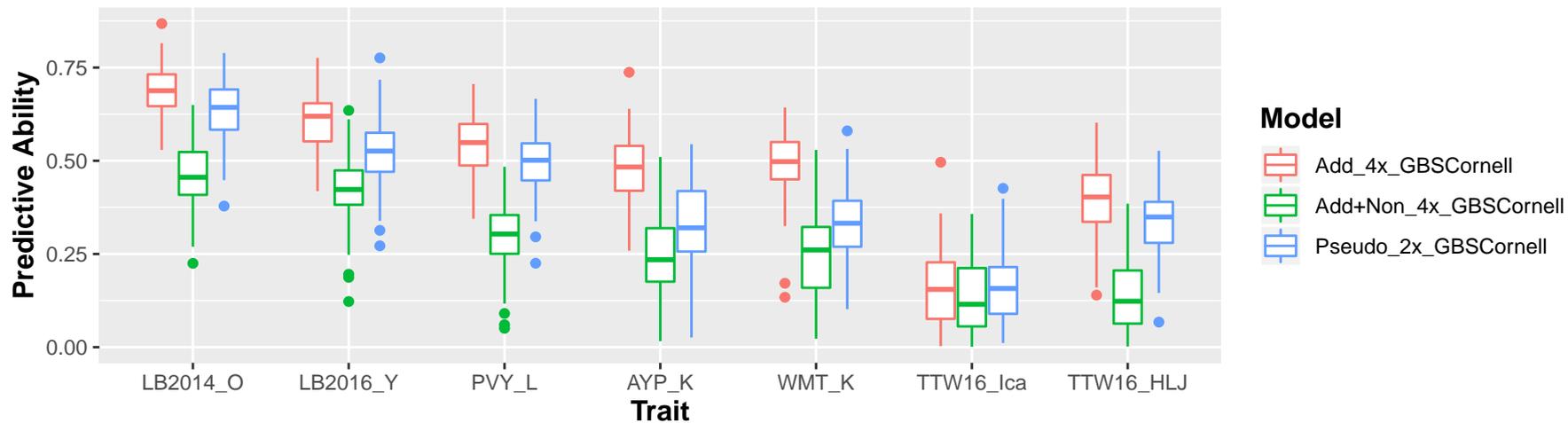


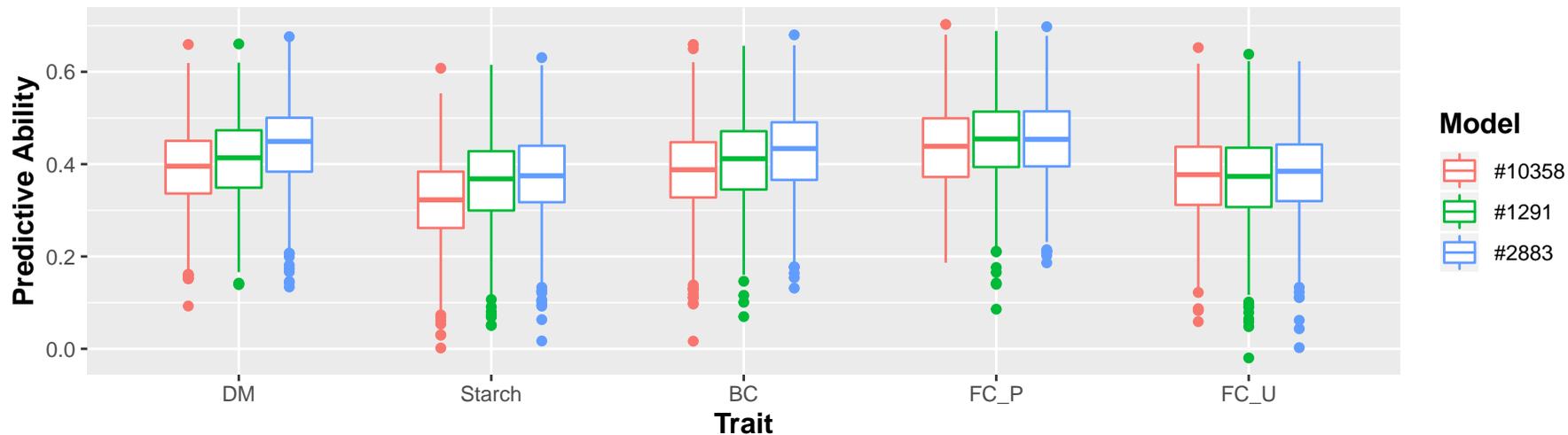
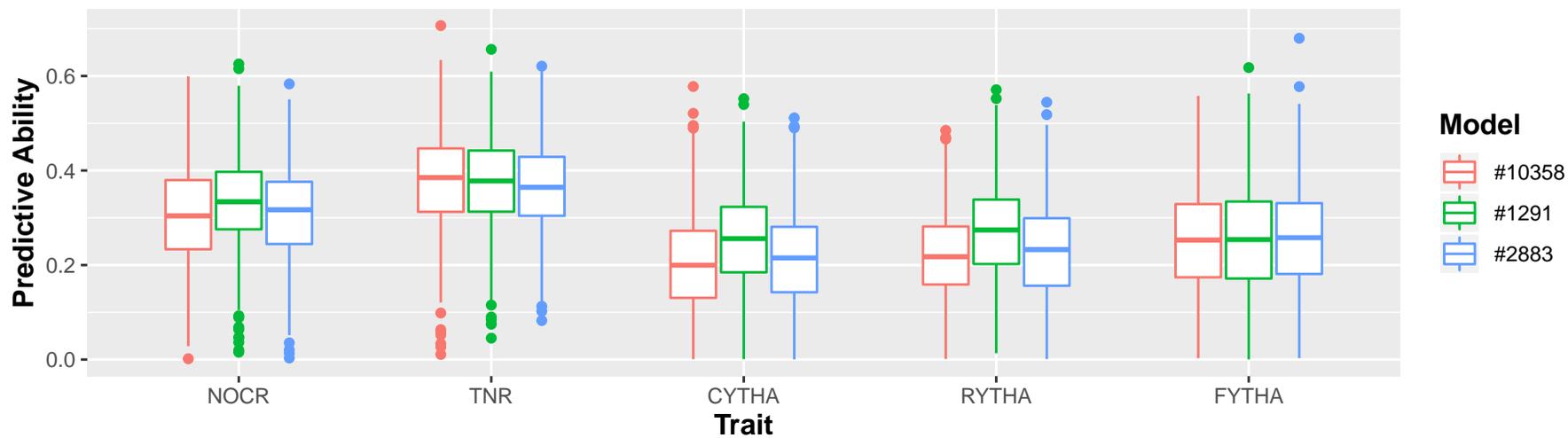
**A****Quality-related traits****Model**

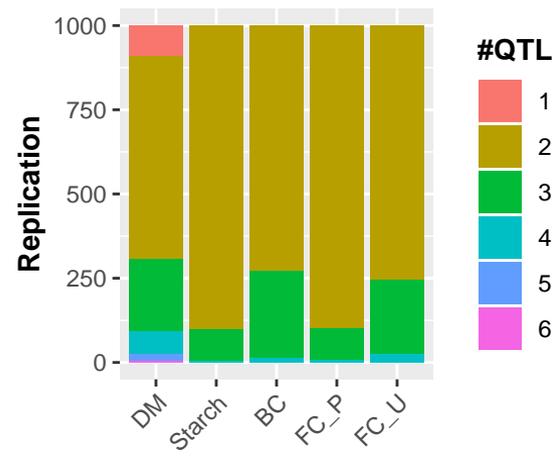
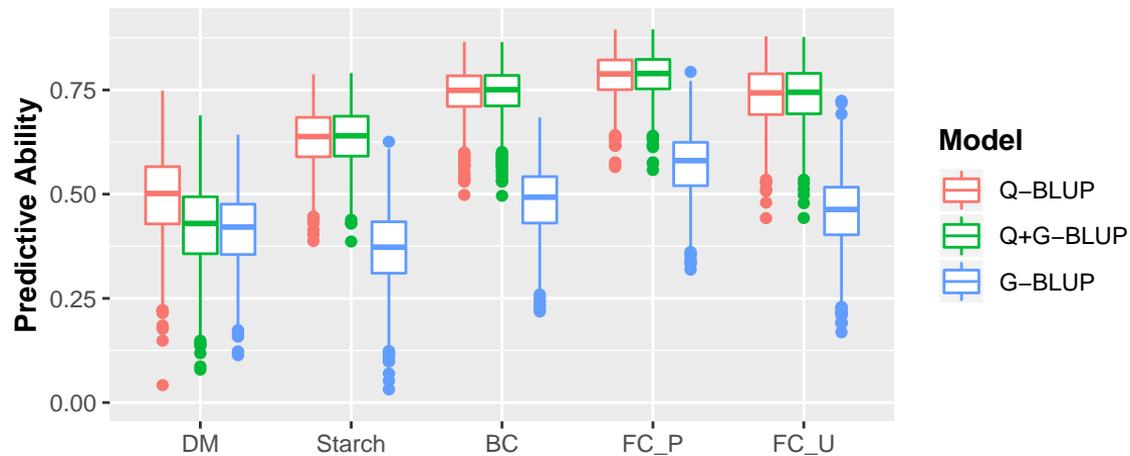
- ▭ Add\_2x\_DArTSeq
- ▭ Add\_6x\_GBSpoly
- ▭ Add+Non\_6x\_GBSpoly
- ▭ NonAdd\_2x\_DArTSeq
- ▭ Pseudo\_2x\_GBSpoly

**B****Yield-related traits****Model**

- ▭ Add\_2x\_DArTSeq
- ▭ Add\_6x\_GBSpoly
- ▭ Add+Non\_6x\_GBSpoly
- ▭ NonAdd\_2x\_DArTSeq
- ▭ Pseudo\_2x\_GBSpoly

**A****TON Potato traits\_MAF0.3 411 SNPs****B****TON Potato traits\_MAF0.4 178 SNPs**

**A****Quality-related traits****B****Yield-related traits**

**A****Quality-related traits****B****Yield-related traits**