

# 1 Unraveling the hexaploid sweetpotato inheritance using 2 ultra-dense multilocus mapping

3  
4 Marcelo Mollinari<sup>1,2,\*</sup>, Bode A. Olukolu<sup>3</sup>, Guilherme da S. Pereira<sup>1,2</sup>, Awais Khan<sup>4</sup>, Dorcus  
5 Gemenet<sup>5</sup>, Craig Yencho<sup>2</sup>, Zhao-Bang Zeng<sup>1,2</sup>

6  
7 1 Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA

8 2 Department of Horticultural Science, North Carolina State University, Raleigh, North Carolina, USA

9 3 Department of Entomology and Plant Pathology, University of Tennessee, Knoxville, Tennessee, USA

10 4 Plant Pathology and Plant-Microbe Biology Section, Cornell University, Geneva, New York, USA

11 5 International Potato Center, ILRI Campus, Nairobi, Kenya

12

13 \* [mmollin@ncsu.edu](mailto:mmollin@ncsu.edu)

14

## 15 Abstract

16

17 The hexaploid sweetpotato (*Ipomoea batatas* (L.) Lam.,  $2n = 6x = 90$ ) is an important staple  
18 food crop worldwide and has a vital role in alleviating famine in developing countries. Due to its  
19 high ploidy level, genetic studies in sweetpotato lag behind major diploid crops significantly. We  
20 built an ultra-dense multilocus integrated genetic map and characterized the inheritance system  
21 in a sweetpotato full-sib family using our newly implemented software, MAPpoly. The resulting  
22 genetic map revealed 96.5% collinearity between *I. batatas* and its diploid relative *I. trifida*. We  
23 computed the genotypic probabilities across the whole genome for all individuals in the mapping

24 population and inferred their complete hexaploid haplotypes. We provide evidence that most of  
25 the meiotic configurations (73.3%) were resolved in bivalents, although a small portion of  
26 multivalent signatures (15.7%), among other inconclusive configurations (11.0%) were also  
27 observed. Except for low levels of preferential pairing in linkage group 2, we observed a  
28 hexasomic inheritance mechanism in all linkage groups. We propose that the hexasomic-bivalent  
29 inheritance promotes stability to the allelic transmission in sweetpotato.

30

## 31 **Introduction**

32

33 The cultivated hexaploid sweetpotato (*Ipomoea batatas* (L.) Lam.,  $2n = 6x = 90$ ) is an important  
34 staple food crop worldwide with an annual production of ~112.84 tons<sup>1</sup>. It has a vital role in  
35 alleviating famine, especially in developing countries in Africa and Southeast Asia<sup>2</sup>. Despite its  
36 undeniable social and economic importance, genetic studies in sweetpotato lag behind major  
37 diploid crops significantly due to its complex polyploid genome. Polyploids are organisms with  
38 more than two chromosome sets. They are grouped into two categories, *allopolyploids* or  
39 *autopolyploids*, when these chromosome sets are originated from either different or same  
40 species, respectively<sup>3</sup>. While in diploid organisms the study of allelic transmission and genetic  
41 linkage are rather straightforward, in polyploids these studies are greatly hindered due to the  
42 wide range of meiotic configurations these species undergo<sup>4-6</sup>. Moreover, current linkage  
43 analysis methods for complex polyploids (i.e., ploidy level > 4) are mostly based on pairwise (or  
44 two-point) marker analyses<sup>7-12</sup>. These methods rely on the assumption that the information in  
45 isolated pairs of markers is sufficient to detect recombination events between them accurately.  
46 However, in cases of complex polyploids, this is rarely true due to the limited mapping

47 population size and the incomplete information provided by biallelic markers. Here, we present a  
48 fully informative multilocus genetic map of a full-sib hexaploid sweetpotato population derived  
49 from a cross between the cultivars ‘Beauregard’ and ‘Tanzania’ (BT population) scored with  
50 more than 30,000 informative single nucleotide polymorphisms (SNPs) using our newly  
51 developed R package called MAPpoly. We also inferred the haplotypes of all individuals in the  
52 full-sib population, which provided novel insights into the multivalent formation and preferential  
53 pairing in the sweetpotato genome.

54

55 Our multilocus analysis considers multiple SNPs simultaneously and propagates their  
56 information through the linkage group (LG) to overcome the typical low informativeness of  
57 some two-loci combinations. This strategy is fundamentally important for complex polyploid  
58 genome analysis since pairs of biallelic markers carry very little information about the  
59 recombination process individually<sup>13,14</sup>. Moreover, the signal-to-noise (S/N) ratio in complex  
60 polyploid SNP data sets is considerably lower as compared to that in diploids and tetraploids<sup>15</sup>,  
61 thus making the genotype calling more prone to errors. The multilocus approach can more  
62 appropriately take into account these errors by using the probability distribution of genotypes  
63 provided by the genotype calling software<sup>14</sup>. Actually, multilocus methods are essential to use  
64 the information of multiple-dose markers to assess complex polyploid inheritance systems  
65 adequately.

66

67 Several studies attempted to elucidate the polyploidy nature in sweetpotato (allo vs.  
68 autopolyploid), including cytological and molecular marker analyses<sup>9,11,16–22</sup>, and more recently  
69 sequence-based studies<sup>23–25</sup>. Two polyploidization scenarios were proposed: the first suggests an

70 allopolyploid origin involving the hybridization of two sweetpotato wild diploid relatives, *I.*  
71 *trifida* and *I. triloba*<sup>20</sup>; the second, well supported by the literature, suggests an autopolyploid  
72 origin with *I. triloba* having a dual role in the polyploidization process<sup>21,23–25</sup>. Corroborating this  
73 scenario, the polysomic inheritance observed in several molecular marker studies<sup>9,11,19,22</sup> rules  
74 out the strict allopolyploid sweetpotato origin. Nevertheless, none of these studies presented a  
75 comprehensive profile of chromosomal pairing for all homology groups across the whole  
76 genome nor the potential formation of multivalents at a population level. Solving these missing  
77 pieces of information is essential to unravel the precise mode of inheritance in sweetpotato, and  
78 consequently, allow an efficient application of molecular techniques in this complex polyploid  
79 breeding system. The BT population coupled with high-coverage sequence genotyping has two  
80 essential characteristics that enabled high-quality mapping: 1) high and uniform sequence read  
81 depth across the genome, which allows for good quality genotype calling that includes high-dose  
82 markers, and 2) sufficiently large sample size to allow the detection of recombination events in a  
83 hexaploid scenario. Additionally, we considered the uncertainty in the genotype calling in  
84 filtering problematic SNPs and also in modeling the error during the map construction using a  
85 hidden Markov model (HMM)<sup>14</sup>. Moreover, all methods can be readily used in tetraploid and  
86 octoploid full-sib populations.

87

## 88 **Results**

### 89 **Genotype calling**

90 Next-generation sequencing produced several millions of barcoded reads, resulting in  
91 approximately 41 million tags which were aligned against the genomes of two sweetpotato  
92 diploid relatives, *I. trifida* and *I. triloba*,<sup>26</sup> resulting in 1,217,917 and 1,163,397 SNPs,

93 respectively. We used the software SuperMASSA<sup>27</sup> to call a total of 442,184 SNPs anchored to *I.*  
94 *trifida* genome and 438,808 anchored to *I. triloba* genome. After filtering out low-quality and  
95 redundant SNPs (Supplementary Fig. 1 A), we obtained 38,701 SNPs scored in 311 individuals.  
96 For all SNPs we obtained dosage-based calls and the associated probability distribution  
97 (exemplified in Fig. 1). From the total SNPs, 55.5% were classified as simplex (single-dose  
98 markers present in one parent) or double-simplex (single-dose markers present in both parents)  
99 and 44.5% were classified as multiplex (Supplementary Fig. 1 B).

100

101

## 102 **De novo map construction**

103 To build the genetic map, we implemented the R package MAPpoly  
104 (<https://github.com/mmollina/MAPpoly>). The software comprises routines to perform all steps  
105 involved in the map construction of autopolyploid species using a combination of pairwise  
106 recombination fraction and HMM-based map estimation. Firstly, we obtained the recombination  
107 fractions and associated likelihoods for each possible linkage phase for all SNP pairs (~749  
108 million pairs). Next, we selected the recombination fractions associated to the most likely  
109 linkage phase configuration for each SNP pair and applied the Unweighted Pair Group Method  
110 with Arithmetic Mean (UPGMA) hierarchical clustering. We formed 15 distinct clusters  
111 representing *I. batatas* homology groups (Supplementary Fig. 2). For the 15 groups, 93.4% of  
112 the SNPs were co-located on the same chromosomes in both references and LGs (Supplementary  
113 Table 1). These matched SNPs were selected to build the “de novo” map. Since each LG had the  
114 majority of their SNPs corresponding to a distinct chromosome in both references, LGs were  
115 numbered after the diploid references.

116

117 To order the SNPs in each LG, we used the Multidimensional Scaling (MDS) algorithm<sup>28</sup>. The  
118 reordered recombination fraction matrix is shown in [Supplementary Fig. 3 A](#). With the proposed  
119 MDS order, the parental allelic variants were phased using the procedure presented by Mollinari  
120 and Garcia<sup>14</sup>. The algorithm is based on LOD scores of pair-wise markers as the first source of  
121 information to sequentially position the allelic variants in specific homologs. For situations  
122 where pairwise analysis had limited power, the algorithm used the likelihood of multiple markers  
123 in a Markov chain (see [Material and Methods](#) and [Supplementary Note](#)).

124

125 The “de novo” multilocus map is presented in [Supplementary Fig. 3 B](#). The length of the LGs  
126 ranges from 723.7 centimorgans (cM) in LG 8 to 2,037.0 cM in LG 4, with a total map length of  
127 20,201.8 cM and 32,200 SNPs (average inter-locus distance ~0.63 cM), with no considerable  
128 gaps between SNPs. Although MDS algorithm yielded adequate global marker orders for all LGs  
129 ([Supplementary Fig. 3 C](#)) the resulting map is considerably large. Two main reasons for this  
130 inflation are misplacement of closely linked SNPs and genotyping errors<sup>14,29-31</sup>, which will be  
131 systematically addressed in the next sections. The alignment of the “de novo” map against the  
132 reference genomes is shown in [Supplementary Fig. 4](#). Despite several chromosomal  
133 rearrangements, we observed high levels of collinearity between both reference genomes and the  
134 estimated map. The collinearity extended in blocks with few megabase pairs (Mb), as in LGs 2  
135 and 7, up to the whole chromosome in LGs 5, 9, 10, 11, 12, 14, and 15. In cases where the  
136 collinearity extended through the whole chromosome, we observed sites of suppressed  
137 recombination (plateaus in [Supplementary Fig. 4](#)), possibly indicating the location of  
138 centromeric regions.

139

## 140 **Genomic assisted map improvement and modeling of genotyping errors**

141 To reduce the effects of the local marker misplacement in map inflation, we used *I. trifida*  
142 genome to propose alternative SNP orders within collinearity blocks and evaluated the likelihood  
143 of the resulting maps, keeping the one with the higher likelihood (see [Material and Methods](#) and  
144 [Supplementary File 1](#)). We used *I. trifida* as the primary reference genome because the quality of  
145 the assembly is superior and more closely related to *I. batatas* when compared to *I. triloba*<sup>26</sup>.  
146 After the order improvement, ~97.0% of the *I. trifida* SNPs present in the map were locally  
147 reordered (see [Material and Methods](#) and [Supplementary File 1](#)). From the remaining *I. trifida*  
148 SNPs, ~1.3% were kept in their original “de novo” order and ~1.7% were eliminated since their  
149 inclusion caused map inflation higher than 2.00 cM. We then positioned the SNPs private from *I.*  
150 *triloba* reference genome into the resulting map using the constraints imposed by both genomes  
151 (see [Material and Methods](#) and [Supplementary Fig. 5, blue map](#)). The genomic-assisted  
152 reordering resulted in a map with 30,723 SNPs spanning 12,937.3 cM with an average inter-  
153 locus distance of ~0.42 cM, representing a reduction of 1.6-fold when compared to the de novo  
154 map. To address the effects of genotyping errors, we re-estimated the map using the probability  
155 distribution of the genotypes provided by SuperMASSA<sup>27</sup> as prior information in the HMM  
156 emission function<sup>14</sup>, as implemented in MAPpoly ([Supplementary Fig. 5, green map](#)). In this  
157 case, the map length was 4,764.1 cM with an average inter-locus distance of ~0.16 cM,  
158 representing a map reduction of 2.7-fold when compared to the genomic-assisted map.

159

## 160 **Probability distribution of multiallelic genotypes and final map estimation**

161 For all individuals in the BT offspring, we obtained the conditional probability distribution of the  
162 400 possible hexaploid genotypes in the whole genome given the estimated genetic map. We  
163 used the Markovian process to propagate the information throughout each LG (see [Material and](#)  
164 [Methods](#)). The genotypic probability distribution at each genome position was assessed by using  
165 the information of all markers in the LG in all individuals of the full-sib population  
166 ([Supplementary Table 2](#) and [Supplementary Fig. 6](#)). Next, we removed 13 individuals with  
167 inconsistent genotypic profiles ([Supplementary Figs. 7](#) and [8](#)) and, keeping the marker order, we  
168 re-estimated the final map considering 298 individuals. A comparison between the “de novo”  
169 and the final maps shows a length reduction of 7.5-fold due to the removal of spurious  
170 recombination events through the several steps of map improvement ([Supplementary Fig. 5](#)).  
171  
172 The final map contains 30,684 SNPs spanning 2,708.4 cM (average inter-locus distance of ~0.09  
173 cM), with 60.7% simplex and double-simplex markers, and 39.3% multiplex ([Table 1](#) and [Fig.](#)  
174 [2](#)). All homologs showed allelic variations along the LGs indicating that their inheritance pattern  
175 can be assessed in the full-sib population. However, several LG segments showed identical  
176 composition for a subset of homologs, as shown by the *Genotypic Information Content* (GIC)<sup>32</sup>.  
177 In our results, 81.8% of all map positions in ‘Beauregard’ and 77.3 % in ‘Tanzania’ had a GIC >  
178 80%, revealing that we can reliably trace back the inheritance of the most homologs from the  
179 offspring to the parents ([Supplementary Fig. 9](#)). A small number of homologs presented an  
180 identical allelic composition for certain segments, which is the case, for example, of homologs *i*  
181 and *j* for the most of LG 2 and *l* and *k* along the whole LG 11. The complete map can be  
182 interactively browsed at [https://gt4sp-genetic-map.shinyapps.io/bt\\_map/](https://gt4sp-genetic-map.shinyapps.io/bt_map/). For a selected segment,  
183 the browser provides the name of markers, dosages in the parents and the linkage phase



184 configuration of the allelic variants. [Supplementary File 2](#) shows more map information,  
185 including the linkage phase configuration in both parents.  
186  
187 [Supplementary Tables 3](#) and [4](#) summarize the results of collinearity blocks containing the  
188 identical SNP sequences between *I. batatas* genetic map and *I. trifida* and *I. triloba* genomes,  
189 respectively. Thirty-nine blocks were aligned to 326.5 Mb of *I. trifida* genome, covering 96.5%  
190 of the *I. batatas* map (2,614.8 cM), with an average density of one SNP/14.2 kb; 107 blocks were  
191 aligned to 258.8 Mb of *I. triloba* genome, covering 83.1% of the map (2,251.8 cM), with an  
192 average density of one SNP/13.4 kb. The averaged genetic to physical map ratios for these  
193 regions were of ~124.8 kb per cM for *I. trifida* and ~114.9 kb per cM for *I. triloba*.

194

### 195 **Haplotype reconstruction and multivalent formation**

196

197 To obtain the haplotype composition of all individuals in the full-sib population, we assessed the  
198 conditional probability distribution of the genotypes and appropriately combined them to build  
199 12 profiles (one for each homolog) indicating the probability of inheritance of a particular  
200 homolog across the whole chromosomes for all individuals in the BT population (see [Material](#)  
201 [and Methods](#)). The results can be accessed at [https://gt4sp-genetic-](https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/)  
202 [map.shinyapps.io/offspring\\_haplotype\\_BT\\_population/](https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/). By evaluating the recombination points  
203 and the homologs involved in the chromosomal exchange, we proposed a heuristic to obtain  
204 chains of homologs linked by recombination events (details in [Material and Methods](#)). These  
205 chains represent the inference of the meiotic process. Although they do not imply an exact  
206 pairing configuration, they can be classified according to the number of homologs involved in

207 the chain. The number of parental homologs that are present in a single homolog of a particular  
208 offspring individual indicates the minimum valency of the meiotic configuration involved in its  
209 gamete formation (see example in [Fig. 3](#)). Thus, recombination chains with two homologs  
210 indicate the formation of at least a bivalent; three homologous, at least trivalent, and so on. For  
211 each LG, we calculated the percentage of the maximum number of homologs involved in the  
212 same recombination chain ([Fig. 4](#)). Most of the configurations involve recombination of two  
213 homologs (~73.8% in ‘Beauregard’ and 72.8 % in ‘Tanzania’) indicating that, homologs can  
214 undergo multivalent formation during the pachytene, though with no consequences of a  
215 multivalent formation in the majority of gametes formed. However, we also observed 12.8% of  
216 gametes in ‘Beauregard’ and 15.2% in ‘Tanzania’ with haplotype configurations involving three  
217 or four parental homologs in a single offspring homolog (indicating trivalent or tetravalent  
218 formation), and less than 2% of the meiotic configurations with five or six homologs (indicating  
219 pentavalent trivalent or hexavalent formation; details per LG in [Supplementary Table 5](#)). We also  
220 detected a significant positive linear correlation ( $P < 10^{-3}$ ) between the number of individuals  
221 with meiotic configurations originated from multivalent formations and the length of LGs  
222 ([Supplementary Fig. 10](#)).

223

## 224 **Preferential pairing**

225 In a hexaploid organism, there are 15 possible pairing configurations for a chromosome segment  
226 during the prophase I of meiosis. To assess the level of preferential pairing among homologs, we  
227 calculated the probability profile for each of the 15 possible pairing configurations across all LGs  
228 for parents ([Fig. 5](#)). We did not observe significant preferential pairing across the whole  
229 sweetpotato genome, except LG 2 ( $P < 10^{-6}$ ) which showed a low but significant preferential

230 pairing. To further ascertain homolog preferential pairing, we evaluated the simplex marker  
231 information which confirmed our preferential pairing findings using the multilocus framework  
232 ([Supplementary Fig. 11](#)).

233

## 234 **Discussion**

235 We have built the first multilocus integrated genetic map of a hexaploid species, sweetpotato,  
236 using our newly implemented software MAPpoly. In the map, 90 homologs were densely  
237 represented in the 15 homology groups of cultivars ‘Beauregard and ‘Tanzania’ exhibiting high  
238 collinearity to two closely related diploid sweetpotato genomes, *I. trifida* and *I. triloba*. The high  
239 collinearity found by using our ultra-dense map corroborates with the levels of alignment (>  
240 90%) between the diploid genomes and the parent ‘Tanzania’ reported by Wu et al.<sup>26</sup>, suggesting  
241 that the diploid genome assemblies could be used as robust references for the hexaploid  
242 sweetpotato. We also have constructed the hexaploid haplotypes of all individuals in the  
243 offspring, estimating the level of preferential pairing and multivalent formation during the  
244 meiotic process at a population level.

245

246 Haplotype inference is the ultimate attainment in linkage analysis since it contains the complete  
247 information about genome transmission across generations. The challenge of performing such  
248 inference, both in parents and offspring, would require new approaches to model the multiallelic  
249 transmission in a very complex meiotic scenario. Here we accomplished this by propagating the  
250 incomplete information of dosage-based SNPs throughout the LG using a Markov chain. As a  
251 result of the efficient combination of multiple SNPs, several LGs displayed fully informative  
252 parental haplotypes in most of their extension ([Fig. 2](#) and [Supplementary Fig. 9](#)). Nevertheless,

253 LG11 had two homologs (*k* and *l*) carrying the same allelic variations across its entire extension,  
254 which leads us to speculate that these two homologs were formed by nondisjunction of sister  
255 chromatids in meiosis II in one of Tanzania's parent resulting in an unreduced gamete  
256 transmitted to the next generation<sup>33</sup>. Even though in some cases where not all homologs could be  
257 distinguished, we estimated their probability distribution, which can be readily used in further  
258 genetic studies, such as quantitative trait loci mapping performed for the BT population<sup>34</sup>. We  
259 also investigated how the pre-assembled parental homologs were transmitted to their offspring  
260 by assessing the probability distribution of the multiallelic genotypes across the whole genome  
261 for all individuals in the mapping population. Based on the inferred probability distributions, we  
262 presented a comprehensive probabilistic reconstruction of the haplotypes of all individuals in a  
263 full-sib hexaploid population. We found that ~15% of the offspring showed the evidence of  
264 multivalent formation, i.e., offspring homologs containing more than two parental homologs.  
265 This leads to intra-homolog variation, which could not be due to exclusive bivalent pairing.  
266  
267 Multivalent configurations often cause faulty chromosomal segregation leading to  
268 aneuploidy<sup>35,36</sup>. Such a phenomenon causes unbalanced gametes, and consequently the  
269 production of pollen and seeds with low viability, posing a significant hindrance to a stable  
270 genomic transmission throughout generations in polyploids<sup>37</sup>. Multivalents are usually observed  
271 in high numbers in recently formed polyploids, as in the case of the synthetic autopolyploid  
272 *Arabidopsis thaliana*<sup>38</sup>. Most of the established autopolyploids, however, show considerably  
273 fewer multivalents. In a survey involving 93 autopolyploid species, Ramsey and Schemske<sup>39</sup>  
274 showed that the average frequency of bivalents was 63.7% whereas the average frequency of  
275 quadrivalents was 26.8%, which are significantly different from the theoretically expected ( $1 \times$

276 two bivalents (II + II) to  $2 \times$  one quadrivalent (VI)<sup>4,40</sup>. For hexaploids, the theoretical proportion  
277 of bivalent to multivalent configurations is  $1 \times$  three bivalents (II + II + II) to  $6 \times$  one tetravalent  
278 plus one bivalent (IV + II) to  $8 \times$  one hexavalent (VI)<sup>40</sup>. However, in our work, the number of  
279 multivalent signatures observed was notably low, whereas the number of bivalents was relatively  
280 high (Fig. 4). These results corroborate the previous cytological study by Magoon and co-  
281 authors<sup>17</sup>, who found similar levels of multivalent configurations in sweetpotato pachytene cells.  
282 Nevertheless, our results provide population-level evidence to the prevalence of bivalent  
283 configurations in sweetpotato meiosis.

284

285 In a scenario of scarce multivalent formation, the double reduction (DR) phenomenon becomes a  
286 somewhat rare event. The DR of a given locus is a consequence of a series of events: the  
287 occurrence of a crossing-over event between a locus and its centromere in a multivalent and  
288 subsequent migration of sister chromatids carrying a duplicated region to the same pole of the  
289 cell at anaphase I<sup>41,42</sup>. Such events could generate genotypes which are not observable under  
290 random chromatid segregation, potentially producing new genotypes. Multivalent formation is a  
291 necessary, but not sufficient condition for the occurrence of DR, which is expected to occur in  
292 low frequency<sup>43</sup>. Consequently, the observed low frequency of multivalent formation would  
293 indicate that the occurrence of DR events is much less likely. Although we did not take into  
294 account DR events during the construction of genetic map, it would have little impact to our map  
295 as the algorithm used here was found to be robust under low levels of multivalent formation<sup>14</sup>.  
296 Nevertheless, even a rare event, it could generate transgressive phenotypes that can be inherited  
297 through the next generations.

298

299 All sweetpotato genetic maps published to date<sup>9,18,19,44–47</sup> have acknowledged the hexasomic  
300 segregation in sweetpotato. However, none of them systematically characterized this  
301 phenomenon using the information of multiple markers assembled in complete hexaploid  
302 homology groups. Here we used the multilocus map to assess this information generating  
303 preferential pairing profiles (Fig. 5). We showed that even though sweetpotato origin could be  
304 traced to an interspecific hybridization as suggested by some studies<sup>17,24,48,49</sup>, its inheritance  
305 pattern is vastly autopolyploid-like and random chromosome pairing would enable  
306 recombination between sub-genomes across generations.

307  
308 A variety of intrachromosomal rearrangements were observed between *I. batatas* map and *I.*  
309 *trifida* and *I. triloba* genomes. Rearrangements mapped to both diploid references, such as the  
310 chromosome inversion at the beginning of LG 6 (Fig. 2), represent structural changes exclusive  
311 to *I. batatas*. While the occurrence of such rearrangements could cause instability to meiotic  
312 process at some point of the evolutionary history of a polyploid species<sup>50</sup>, given the high level of  
313 bivalent signatures and the stable hexasomic segregation observed in our analysis, we concluded  
314 that these structural changes became fixed and do not cause major disturbances to the meiotic  
315 process in sweetpotato.

316  
317 More than a linear order of genetic markers positioned in LGs, a genetic map is a statement  
318 about the inheritance pattern involved in the transmission of genome from parents and their  
319 offspring. A full characterization of this process can be achieved if the mapping method allows  
320 the estimation of haplotypes in both generations. In diploid organisms, a hidden Markov model  
321 was proposed by Lander and Green for linkage analysis of multiple markers<sup>51</sup>. Later on, several

322 studies paved the way for a linkage map construction and haplotype inference in autotetraploid  
323 species<sup>52-55</sup>. However, for complex polyploids the map construction was restricted mostly to  
324 two-point marker analysis. We present the first integrated multilocus genetic map with fully  
325 phased haplotypes for both parents and offspring in a complex polyploid and, accompanied with  
326 it, the fully developed statistical methods and computational tool MAPpoly. This opens the door  
327 for detailed genetic analysis in complex polyploid species in general.

## 328 **Material and Methods**

### 329 **Plant material**

330 The mapping population consists of 315 full-sib individuals originated from a cross between the  
331 orange-flesh cultivar ‘Beauregard’ (CIP440132 - male) and the African landrace ‘Tanzania’  
332 (CIP440166 - female). These two cultivars were selected due to their agronomic importance and  
333 contrasting traits, such as, beta-carotene, dry matter, drought tolerance and resistance for viruses  
334 and nematodes<sup>11,56</sup>, for further QTL studies.

335

### 336 **Optimized genotyping-by-sequencing protocol - GBSpoly**

337 Next-Generation Sequencing (NGS) library preparation protocol was optimized for polyploids  
338 and highly heterozygous genomes to produce uniform coverage across samples and loci,  
339 GBSpoly<sup>57</sup> (details in [Supplementary Note](#)). The optimizations were based on re-engineered  
340 barcoded adapters that ensure accurate demultiplexing and base calls. The 6-9 bp variable length  
341 barcodes with designed to account for both substitution and indel errors (based on  
342 edit/levenshtein distance), minimizes phasing error and maintains nucleotide diversity at every  
343 position along the reads. A new feature, buffer sequences, upstream of the barcodes ensures that  
344 the barcodes lie in high-quality base regions by avoiding the elevated error rates at the ends of  
345 the reads. The adapters were ligated to fragments generated by double digests, *TseI* and *CviAII*,  
346 and then size selected to minimize PCR bias. By designing barcodes that did not reconstitute the  
347 restriction sites, ligated fragments were subjected to a secondary digest to eliminate chimeric  
348 fragments. Sequencing was performed on the Illumina HiSeq 2500.

349

### 350 **Genotype calling**



351 We used the software SuperMASSA<sup>27</sup> to perform the genotype calling of parents and offspring  
352 of the full-sib population. For quality control purposes, we eliminated SNPs with read depth < 20  
353 and estimated ploidy level different from six. We also filtered out SNPs with more than 25% of  
354 missing data and with segregation distortion ( $P < 5 \times 10^{-4}$ ). Additionally, we removed four  
355 individuals with less than 100 reads on average for the selected SNPs (see [Supplementary Note](#)).  
356 We obtained the physical positions of the selected markers in two diploid reference genomes of  
357 *I. trifida* and *I. triloba*<sup>26</sup> and classified them into shared between both genomes or private to a  
358 specific genome based on the full-sib population genotype calls.

359

## 360 **De novo map construction**

361 *Grouping and SNP ordering* – We computed recombination fractions for all marker pairs  
362 considering all possible linkage phase configurations. For each marker pair, we selected the  
363 recombination fraction associated to the most likely linkage phase and assembled a  
364 recombination fraction matrix for all marker pairs. Using UPGMA hierarchical clustering we  
365 generated a dendrogram representing 15 LGs corresponding to the 15 sweetpotato homology  
366 groups. To order the SNPs in each LG, we converted the recombination fractions to genetic  
367 distances using Haldane’s map function and applied the unconstrained MDS algorithm with the  
368 squared linkage LOD Scores to construct the stress criterion<sup>28</sup>.

369

370 *Phasing and multilocus map estimation* – The parental linkage phase configuration was obtained  
371 by serially adding markers to the map sequence and evaluating two-point likelihoods associated  
372 to possible configurations between the inserted markers and the ones already positioned. If the  
373 LOD Score between the two most likely configurations was less than ten for a subset of

374 configurations, we compared the multipoint likelihoods to proceed to the next marker insertion.  
375 When the last marker was inserted, we re-estimated the multipoint recombination fractions  
376 between all adjacent markers<sup>14</sup>. For more details see [Supplementary Note](#).

377

### 378 **Genome-assisted map improvement**

379 Using the *I. trifida* reference, we detected collinearity blocks within each LG by visually  
380 inspecting abrupt breakages in the scatter plots continuity ([Supplementary Fig. 4](#)). For each  
381 collinearity block, we evaluated the multilocus likelihood associated with the “de novo” order  
382 and the order provided by *I. trifida* reference. We selected the maximum likelihood order for  
383 each block, tested several orientations among them ([Supplementary File 1](#)) and chose the  
384 configuration that yielded the highest multilocus likelihood for the complete map. Next, we  
385 inserted the remaining private SNPs from *I. triloba* using the genomic position constraints  
386 imposed by SNPs shared by both genomes. We also eliminated SNPs that caused substantial map  
387 expansions (see [Supplementary Note](#)). Finally, we re-estimated the map by considering the  
388 probability distribution of the genotypes provided by SuperMASSA<sup>27</sup>. We also computed the  
389 GIC<sup>32</sup> for each homolog across the entire genome.

390

### 391 **Probability distribution of the offspring genotypes**

392 The probability distribution for all possible 400 hexaploid genotypes was calculated using the  
393 HMM framework detailed in [Supplementary Note](#). Briefly, if  $\mathcal{G}_{k,j}$  denote the  $j^{\text{th}}$  genotype,  
394  $j \in \{1, \dots, 400\}$  of an individual in a hexaploid full-sib population at locus  $k$ , the conditional  
395 probability distribution of  $\mathcal{G}_{k,j}$  is defined as

396

$$\Pr(\mathcal{G}_{k,j} | O_1, \dots, O_z, \lambda) = \frac{\alpha_k(j)\beta_k(j)}{\sum_{i=1}^{400} \alpha_k(i)\beta_k(i)} \quad (1)$$

where  $O_1, \dots, O_z$  is a sequence of observations of  $z$  markers,  $\lambda$  denotes the map parameters,  $\alpha_k(j)$  denotes the joint probability of the partial observation sequence to the left of marker  $k$  (including  $k$ ) and genotype  $\mathcal{G}_{k,j}$ , given the map parameters  $\lambda$ ; similarly,  $\beta_k(j)$  denotes the probability of the partial observation sequence to the right of the position  $k$  given the genotype  $\mathcal{G}_{k,j}$  and the map parameters  $\lambda$ . The quantities  $\alpha_k(j)$  and  $\beta_k(j)$  can be obtained using the classical *forward-backward* algorithm<sup>58,59</sup> and their derivation is presented in [Supplementary Note](#).

### Offspring haplotype reconstruction

The probability that an offspring individual carries a specific parental homolog  $\mathcal{H}$  at position  $k$  can be obtained using

$$\Pr(\mathcal{H}_k | O_1, \dots, O_z, \lambda) = \sum_{i=1}^{400} \Pr(\mathcal{H}_k | \mathcal{G}_{k,i}, O_1, \dots, O_z, \lambda) \Pr(\mathcal{G}_{k,i} | O_1, \dots, O_z, \lambda)$$

where,  $\mathcal{H}_k \in \{a, b, c, d, e, f, g, h, i, j, k, l\}$  is the inherited homolog at locus  $k$ ,

$\Pr(\mathcal{H}_k | \mathcal{G}_{k,i}, O_1, \dots, O_z, \lambda) = 1$  if  $\mathcal{H}_k \in \mathcal{G}_{k,i}$ , 0 otherwise (see [Supplementary Table 2](#)). We

obtained the haplotype probability profile for all 15 homology groups (one curve for each homolog, from  $a$  through  $l$ ) for all individual in the bi-parental cross population by computing

$\Pr(\mathcal{H}_k | O_1, \dots, O_z, \lambda)$  at every marker  $k$  across the genome.

418

### 419 **Heuristic algorithm to detect crossing-over events**

420 Given the probabilistic nature of the haplotype profiles, we proposed the following heuristic  
421 algorithm to detect crossing-over events:

422

423 1. Regions with haplotype probabilities greater than 0.8 are assumed to be 1.0,  
424 otherwise 0.0, forming a binary profile;

425 2. SNPs within a continuous segment of homolog or gaps flanked by crossing-overs  
426 smaller than 10 cM are removed.

427 3. If the remaining SNPs represent 20% or more of all SNPs in the analyzed LG, use  
428 Eq. (1) to re-estimate the 400 genotypes across the whole LG and compute a new  
429 homolog probability profile using Eq. (2). Otherwise, consider the probability profile  
430 inconclusive.

431 4. The crossing-over points are assessed by checking the points of probability  
432 transition across the LG. Homologs involved in the chromosomal exchange can be  
433 trivially assessed.

434 5. Exchange points closer than 0.5 cM are considered inconclusive since the  
435 haplotypes involved in the exchange could be erroneously assigned due to the lack of  
436 resolution in the mapping population.

437

438 We applied this procedure to the 15 LGs of all individuals in the population. We also present an  
439 interactive version of the heuristic algorithm at

440 [https://gt4sp-genetic-map.shinyapps.io/offspring\\_haplotype\\_BT\\_population/](https://gt4sp-genetic-map.shinyapps.io/offspring_haplotype_BT_population/)

441

## 442 **Preferential pairing profiles**

443

444 Considering that all homologs pair during a hexaploid meiosis, there are 15 possible pairing  
445 configurations for a chromosomal segment. Let  $\Psi = \{\psi_i\}, i = 1, \dots, 15$  denote a set containing  
446 all 15 possible configurations<sup>14</sup> (see [Supplementary Note](#)). The posterior probability distribution  
447 of the pairing configurations at any position  $k$  in the genome can be computed using

448

$$\Pr(\psi_i | O_1, \dots, O_z, \lambda) = \sum_{l=1}^n \sum_{j=1}^{400} \Pr(\psi_i | \mathcal{G}_{k,j}, O_1, \dots, O_z, \lambda) \Pr(\mathcal{G}_{k,j} | O_1, \dots, O_z, \lambda)_l \quad (3)$$

449

450

451 where  $n$  is the number of individuals in the population,  $\Pr(\psi_i | \mathcal{G}_{k,j}, O_1, \dots, O_z, \lambda) = \left(\frac{m}{2}!\right)^{-1}$ ,  
452 ( $m = 6$  for hexaploids) if  $\mathcal{G}_{k,j}$  is consistent with  $\psi_i$ , i.e., if genotype  $\mathcal{G}_{k,j}$  can be originated from  
453 the pairing configuration  $\psi_i$ , 0 otherwise<sup>14</sup>. To test whether the observed frequencies of the 15  
454 bivalent configurations differ from the expected under random pairing ( $\frac{1}{15}$ ), we used the  $\chi^2$  test  
455 with  $P < 0.001$  to declare significance. We also used the likelihood associated to recombination  
456 fractions of single-dose markers to assess preferential pairing, as suggested by Wu and co-  
457 authors<sup>60</sup>.

458

459 Further details of methods are given in the [Supplementary Note](#).

## 460 References

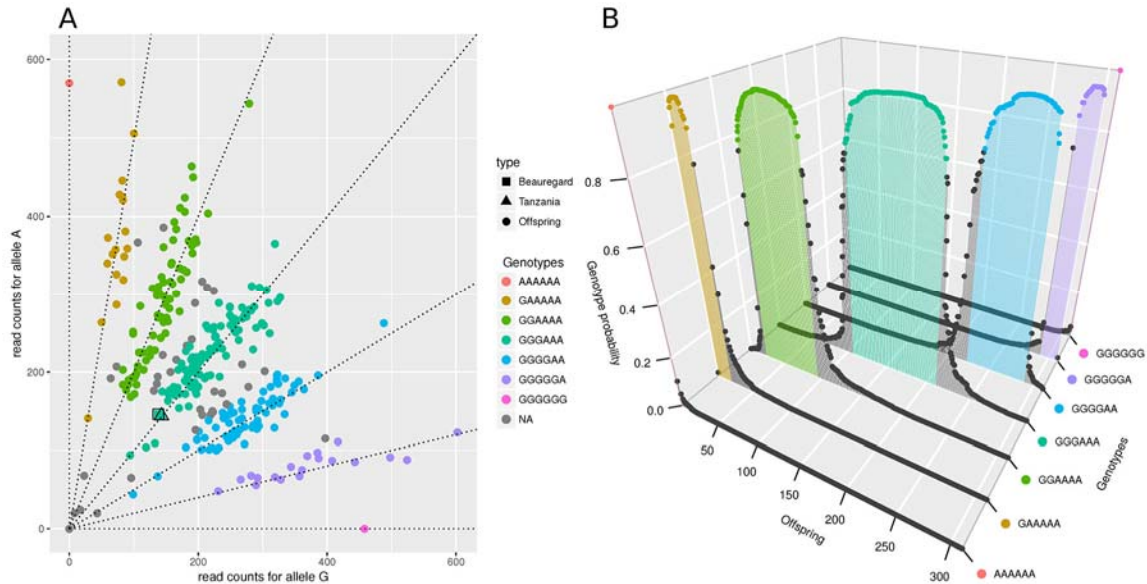
- 461 1. FAO. <http://www.fao.org/faostat/en>. (2017).
- 462 2. Loebenstein, G. Origin, Distribution and Economic Importance. in *The Sweetpotato* (eds.  
463 Loebenstein, G. & Thottappilly, G.) **49**, 9–12 (Springer, 2009).
- 464 3. Comai, L. The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–  
465 846 (2005).
- 466 4. Sybenga, J. *Meiotic configurations*. (Springer, 1975).
- 467 5. Gallais, A. *Quantitative genetics and breeding methods in autopolyploids plants*. (INRA,  
468 2003).
- 469 6. Zielinski, M.-L. & Scheid, O. M. Poliploidy and Genome Evolution. in (eds. Soltis, P. S.  
470 & Soltis, D. E.) 33–55 (Springer, 2012).
- 471 7. Fisher, R. A. The theoretical consequences of polyploid inheritance for the mid style form  
472 in *Lythrum salicaria*. *Ann. Eugen.* **11**, 31–38 (1941).
- 473 8. Ripol, M. I., Churchill, G. A., Silva, J. A. G. Da, Sorrells, M. & da Silva, J. A. G.  
474 Statistical aspects of genetic mapping in autopolyploids. *Gene* **235**, 31–41 (1999).
- 475 9. Kriegner, A., Cervantes, J. C., Burg, K., Mwanga, R. O. M. & Zhang, D. A genetic  
476 linkage map of sweet potato [*Ipomoea batatas* (L.) Lam.] based on AFLP markers. *Mol.*  
477 *Breed.* **11**, 169–185 (2003).
- 478 10. Aitken, K. S., Jackson, P. A. & McIntyre, C. L. Construction of a genetic linkage map for  
479 *Saccharum officinarum* incorporating both simplex and duplex markers to increase  
480 genome coverage. *Genome* **50**, 742–756 (2007).
- 481 11. Cervantes-Flores, J. C. *et al.* Development of a genetic linkage map and identification of  
482 homologous linkage groups in sweetpotato using multiple-dose AFLP markers. *Mol.*  
483 *Breed.* **21**, 511–532 (2008).
- 484 12. van Geest, G. *et al.* An ultra-dense integrated linkage map for hexaploid chrysanthemum  
485 enables multi-allelic QTL analysis. *Theor. Appl. Genet.* **130**, 2527–2541 (2017).
- 486 13. Luo, Z. W., Zhang, R. M. & Kearsey, M. J. Theoretical basis for genetic linkage analysis  
487 in autotetraploid species. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 7040–7045 (2004).
- 488 14. Mollinari, M. & Garcia, A. A. F. Linkage analysis and haplotype phasing in experimental  
489 autopolyploid populations with high ploidy level using hidden Markov models. *G3 -*  
490 *Genes, Genomes, Genet.* (2019).
- 491 15. Mollinari, M. & Serang, O. Quantitative SNP Genotyping of Polyploids with  
492 MassARRAY and Other Platforms. in *Methods in Molecular Biology* **1245**, 215–241  
493 (2015).
- 494 16. Gustafsson, Å. & Gadd, I. Mutations and crop improvement III. *Ipomoea batatas* (L.)  
495 Poir. (Convolvulaceae). *Hereditas* **53**, 77–89 (1965).
- 496 17. Magoon, M. L., Krishnan, R. & Vijaya Bai, K. Cytological evidence on the origin of  
497 sweet potato. *Theor. Appl. Genet.* **40**, 360–366 (1970).
- 498 18. Ukoskit, K. & Thompson, P. Autopolyploidy versus allopolyploidy and low-density  
499 randomly amplified polymorphic DNA linkage maps of sweetpotato. *Journal of the*  
500 *American Society of Agricultural Science* **122**, 822–828 (1997).
- 501 19. Zhao, N. *et al.* A genetic linkage map based on AFLP and SSR markers and mapping of  
502 QTL for dry-matter content in sweetpotato. *Mol. Breed.* **32**, 807–820 (2013).
- 503 20. Austin, D. F. Exploration, Maintenance and Utilization of Sweet Potato Genetic  
504 Resources.No Title. in *Report of the First Sweet Potato Planning Conference* 27–59

- 505 (International Potato Center, 1988).
- 506 21. Shiotani, I. & Kawase, T. Synthetic hexaploids derived from wild species related to sweet  
507 potato. *Japan. J. Breed.* **37**, 367–376 (1987).
- 508 22. Monden, Y. & Tahara, M. Genetic linkage analysis using DNA markers in sweetpotato.  
509 *Breed. Sci.* **51**, 41–51 (2017).
- 510 23. Roullier, C. *et al.* Disentangling the Origins of Cultivated Sweet Potato (*Ipomoea batatas*  
511 (L.) Lam.). *PLoS One* **8**, (2013).
- 512 24. Yang, J. *et al.* Haplotype-resolved sweet potato genome traces back its hexaploidization  
513 history. *Nat. Plants* **3**, (2017).
- 514 25. Muñoz-Rodríguez, P. *et al.* Reconciling Conflicting Phylogenies in the Origin of Sweet  
515 Potato and Dispersal to Polynesia. *Curr. Biol.* **28**, 1246–1256.e12 (2018).
- 516 26. Wu, S. *et al.* Genome sequences of two diploid wild relatives of cultivated sweetpotato  
517 reveal targets for genetic improvement. *Nat. Commun.* **9**, 1–12 (2018).
- 518 27. Serang, O., Mollinari, M. & Garcia, A. A. F. Efficient exact maximum a posteriori  
519 computation for bayesian SNP genotyping in polyploids. *PLoS One* **7**, e30906 (2012).
- 520 28. Preedy, K. F. & Hackett, C. A. A rapid marker ordering approach for high-density genetic  
521 linkage maps in experimental autotetraploid populations using multidimensional scaling.  
522 *Theor. Appl. Genet.* **129**, 2117–2132 (2016).
- 523 29. Cartwright, D. A., Troggio, M., Velasco, R. & Gutin, A. Genetic mapping in the presence  
524 of genotyping errors. *Genetics* **176**, 2521–2527 (2007).
- 525 30. Cheema, J. & Dicks, J. Computational approaches and software tools for genetic linkage  
526 map estimation in plants. *Brief. Bioinform.* **10**, 595–608 (2009).
- 527 31. Bilton, T. P. *et al.* Accounting for Errors in Low Coverage High-Throughput Sequencing  
528 Data When Constructing Genetic Maps Using Biparental Outcrossed Populations.  
529 *Genetics* **209**, 65–76 (2018).
- 530 32. Bourke, P. M. *et al.* Multi-environment QTL analysis of plant and flower morphological  
531 traits in tetraploid rose. *Theor. Appl. Genet.* **131**, 2055–2069 (2018).
- 532 33. Burnham C. R. *Discussions in Cytogenetics. Biometrical Journal* **8**, (Burgess Publishing,  
533 1964).
- 534 34. Pereira, G. da S. *et al.* Multiple QTL mapping in autopolyploids: a random-effect model  
535 approach with application in a hexaploid sweetpotato full-sib population. *bioRxiv* (2019).  
536 doi:10.1101/622951
- 537 35. Arana, P. & Nicklas, R. B. Orientation and segregation of a micromanipulated  
538 multivalent: Familiar principles, divergent outcomes. *Chromosoma* **101**, 399–412 (1992).
- 539 36. Hollister, J. D. Polyploidy: Adaptation to the genomic environment. *New Phytol.* **205**,  
540 1034–1039 (2015).
- 541 37. Mwathi, M. W. *et al.* Segregation for fertility and meiotic stability in novel Brassica  
542 allohexaploids. *Theor. Appl. Genet.* **130**, 767–776 (2017).
- 543 38. Santos, J. L. *et al.* Partial diploidization of meiosis in autotetraploid *Arabidopsis thaliana*.  
544 *Genetics* **165**, 1533–1540 (2003).
- 545 39. Ramsey, J. & Schemske, D. W. Neopolyploidy in Flowering Plants. *Annu. Rev. Ecol. Syst.*  
546 **33**, 589–639 (2002).
- 547 40. Jackson, R. C. & Casey, J. Cytogenetic Analyses of Autopolyploids: Models and Methods  
548 for Triploids to Octoploids. *Am. J. Bot.* **69**, 487–501 (1982).
- 549 41. Butruille, D. V & Boiteux, L. S. Selection-mutation balance in polysomic tetraploids:  
550 Impact of double reduction and gametophytic selection on the frequency and

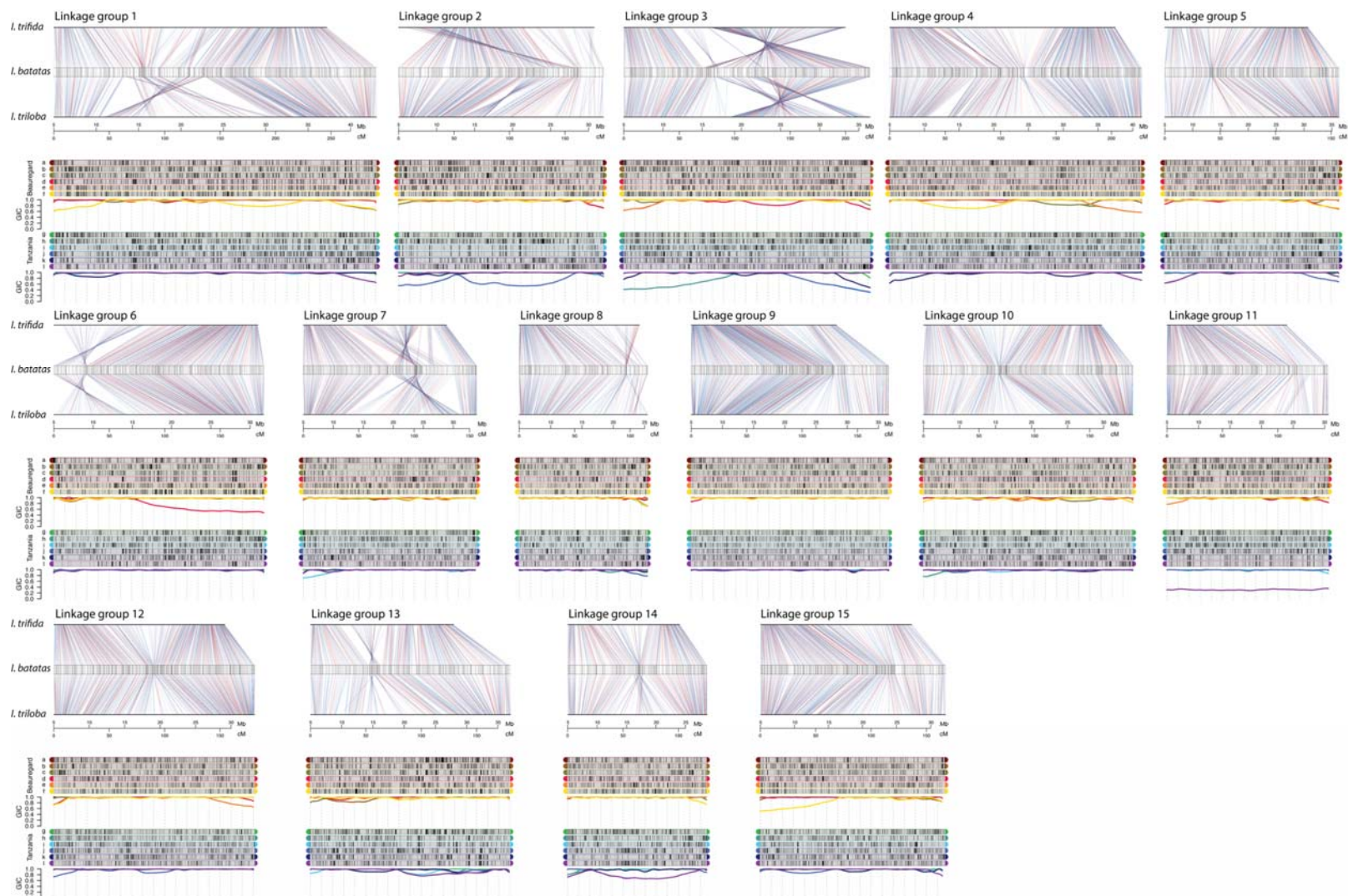


- 551 subchromosomal localization of deleterious mutations. *Proc. Natl. Acad. Sci.* **97**, 6608–  
552 6613 (2000).
- 553 42. Stift, M., Berenos, C., Kuperus, P. & van Tienderen, P. H. Segregation Models for  
554 Disomic, Tetrasomic and Intermediate Inheritance in Tetraploids: A General Procedure  
555 Applied to Rorippa (Yellow Cress) Microsatellite Data. *Genetics* **179**, 2113–2123 (2008).
- 556 43. Voorrips, R. E. & Maliapaard, C. a. The simulation of meiosis in diploid and tetraploid  
557 organisms using various genetic models. *BMC Bioinformatics* **13**, 248 (2012).
- 558 44. Shirasawa, K. *et al.* A high-density SNP genetic map consisting of a complete set of  
559 homologous groups in autohexaploid sweetpotato (*Ipomoea batatas*). *Sci. Rep.* **7**, 44207  
560 (2017).
- 561 45. Cervantes-Flores, J. C. *et al.* Development of a genetic linkage map and identification of  
562 homologous linkage groups in sweetpotato using multiple-dose AFLP markers. *Mol.*  
563 *Breed.* **21**, 511–532 (2008).
- 564 46. Ai-xian, L. *et al.* Establishment of Molecular Linkage Maps Using SRAP Markers in  
565 Sweet Potato. *Acta Agron. Sin.* **36**, 1286–1295 (2010).
- 566 47. Monden, Y. *et al.* Construction of a linkage map based on retrotransposon insertion  
567 polymorphisms in sweetpotato via high-throughput sequencing. *Breed. Sci.* **65**, 145–153  
568 (2015).
- 569 48. Shiotani, I. & Kawase, T. Genomic Structure of the Sweet Potato and Hexaploids in  
570 *Ipomoea trifida* (H.B.K) Don. *Japan. J. Breed.* **39**, 57–66 (1989).
- 571 49. Gao, M. *et al.* Wx intron variations support an allohexaploid origin of the sweetpotato  
572 [*Ipomoea batatas* (L.) Lam]. *Euphytica* **177**, 111–133 (2011).
- 573 50. Lenormand, T., Engelstädter, J., Johnston, S. E., Wijnker, E. & Haag, C. R. Evolutionary  
574 mysteries in meiosis. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).
- 575 51. Lander, E. S. & Green, P. Construction of multilocus genetic linkage maps in humans.  
576 *Proc. Natl. Acad. Sci. U. S. A.* **84**, 2363–2367 (1987).
- 577 52. Hackett, C. A., Pande, B. & Bryan, G. J. Constructing linkage maps in autotetraploid  
578 species using simulated annealing. *Theor. Appl. Genet.* **106**, 1107–15 (2003).
- 579 53. Hackett, C. A., McLean, K. & Bryan, G. J. Linkage analysis and QTL mapping using SNP  
580 dosage data in a tetraploid potato mapping population. *PLoS One* **8**, e63939 (2013).
- 581 54. Leach, L. J., Wang, L., Kearsey, M. J. & Luo, Z. Multilocus tetrasomic linkage analysis  
582 using hidden Markov chain model. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 4270–4274 (2010).
- 583 55. Zheng, C. *et al.* Probabilistic Multilocus Haplotype Reconstruction in Outcrossing  
584 Tetraploids. *Genetics* **203**, 119–131 (2016).
- 585 56. Gemenet, D., Pereira, G. D. S. & *et al.* Translating Genomic Research to Address  
586 Development and Adoption Bottlenecks of Nutritious Sweetpotato [*Ipomoea batatas* (L.)  
587 Lam.] in sub-Saharan Africa. *Theor. Appl. Genet.* **Submitted**, (2019).
- 588 57. Wadl, P. A. *et al.* Genetic Diversity and Population Structure of the USDA Sweetpotato  
589 (*Ipomoea batatas*) Germplasm Collections Using GBSpoly. *Front. Plant Sci.* **9**, 1–13  
590 (2018).
- 591 58. Rabiner, L. R. A Tutorial on Hidden Markov Models and Selected Applications in Speech  
592 Recognition. *Proc. IEEE* **77**, 257–286 (1989).
- 593 59. Jiang, C. & Zeng, Z.-B. Mapping quantitative trait loci with dominant and missing  
594 markers in various crosses from two inbred lines. *Genetica* **101**, 47–58 (1997).
- 595 60. Wu, K. K. *et al.* The detection and estimation of linkage in polyploids using single-dose  
596 restriction fragments. *Theor. Appl. Genet.* **83**, 294–300 (1992).



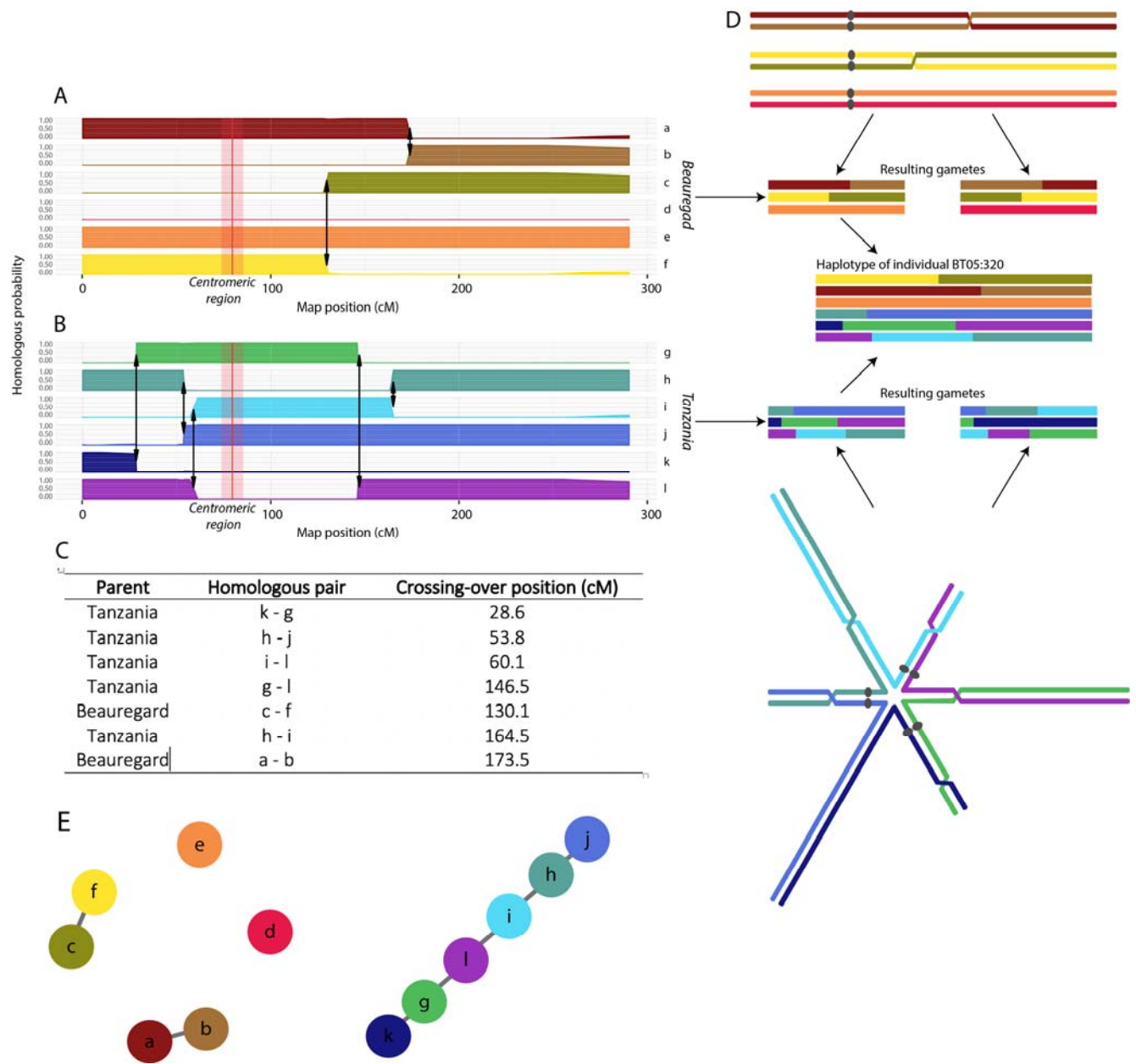


597  
 598 **Fig. 1** Example of genotype call of SNP *Tf\_S1\_30010438*. (A) Scatter plot of the read counts for  
 599 the two allelic variants A and G. The axes represent the read counts of both allelic variants.  
 600 Squared and triangle dots represent parents ‘Beaugard’ and ‘Tanzania’ respectively, and  
 601 regular dots represent the offspring. Dashed lines indicate seven possible dosages in a hexaploid  
 602 individual. The different colors indicate the dosages assigned to the individuals by  
 603 SuperMASSA. The low number of individuals observed between genotypic classes (gray dots,  
 604 with genotype probability smaller than 0.8), outlines a data set with low noise, producing a clear  
 605 classification. The genotypes of both parents were estimated as three doses of the allelic variant  
 606 A three doses of G. The genotype calling model also considered the expected Mendelian  
 607 segregation ratio, which under random chromosome pairing is 1:18:99:164:99:18:1. (B) Inferred  
 608 probability distribution of genotypes for each individual in the offspring. The colored dots  
 609 correspond to individuals with the same genotypic classes in panel A. Loci where the highest  
 610 posterior probability was smaller than 0.8 were assigned as missing data (gray dots).



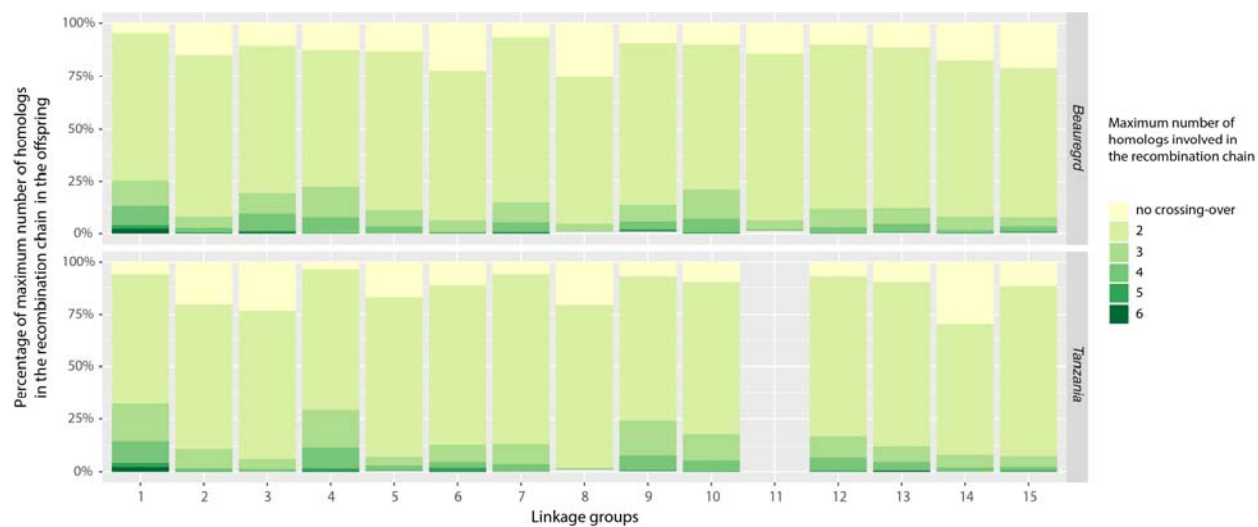
611

612 **Fig. 2** Sweetpotato genetic map. For each of the 15 LGs, we present the *I. batatas* genetic map with its SNPs anchored in both diploid  
 613 reference genomes. Blue lines connecting the map and reference genomes indicate SNPs shared between *I. trifida* and *I. triloba*  
 614 reference genomes and red lines indicate private SNPs. Above each map, we present a graphical representation of the parental linkage  
 615 phase configuration of the homology groups for parents ‘Beauregard’ and ‘Tanzania’. Black and gray rectangles indicate two allelic  
 616 variants in each marker in all 12 parental homologs (6 × in ‘Beauregard’ and 6 × in ‘Tanzania’). The Genotypic Information Content  
 617 (GIC), is presented below each homology group.



618

619 **Fig. 3** Example of haplotype reconstruction and distribution of meiotic configurations for  
 620 individual BT05.320, linkage group 1. A) and B) Probability profiles for 12 homologs indicating  
 621 the segments inherited from parents ‘Beauregard’ and ‘Tanzania’, respectively. The red line  
 622 indicates the approximated centromeric region obtained using the *I. trifida* reference genome.  
 623 The arrows indicate recombination points; C) Recombination signature table indicating the  
 624 homolog pairs involved in each crossing-over and their position in the map; D) Possible meiotic  
 625 configuration that originated gametes for individual BT05.320 in ‘Beauregard’ and ‘Tanzania’  
 626 and resulting gamete. Each chromosome is represented by one chromatid; E) Representation of  
 627 the meiotic results as a graph: nodes represent the homologs and the edges represent  
 628 recombination events between them.

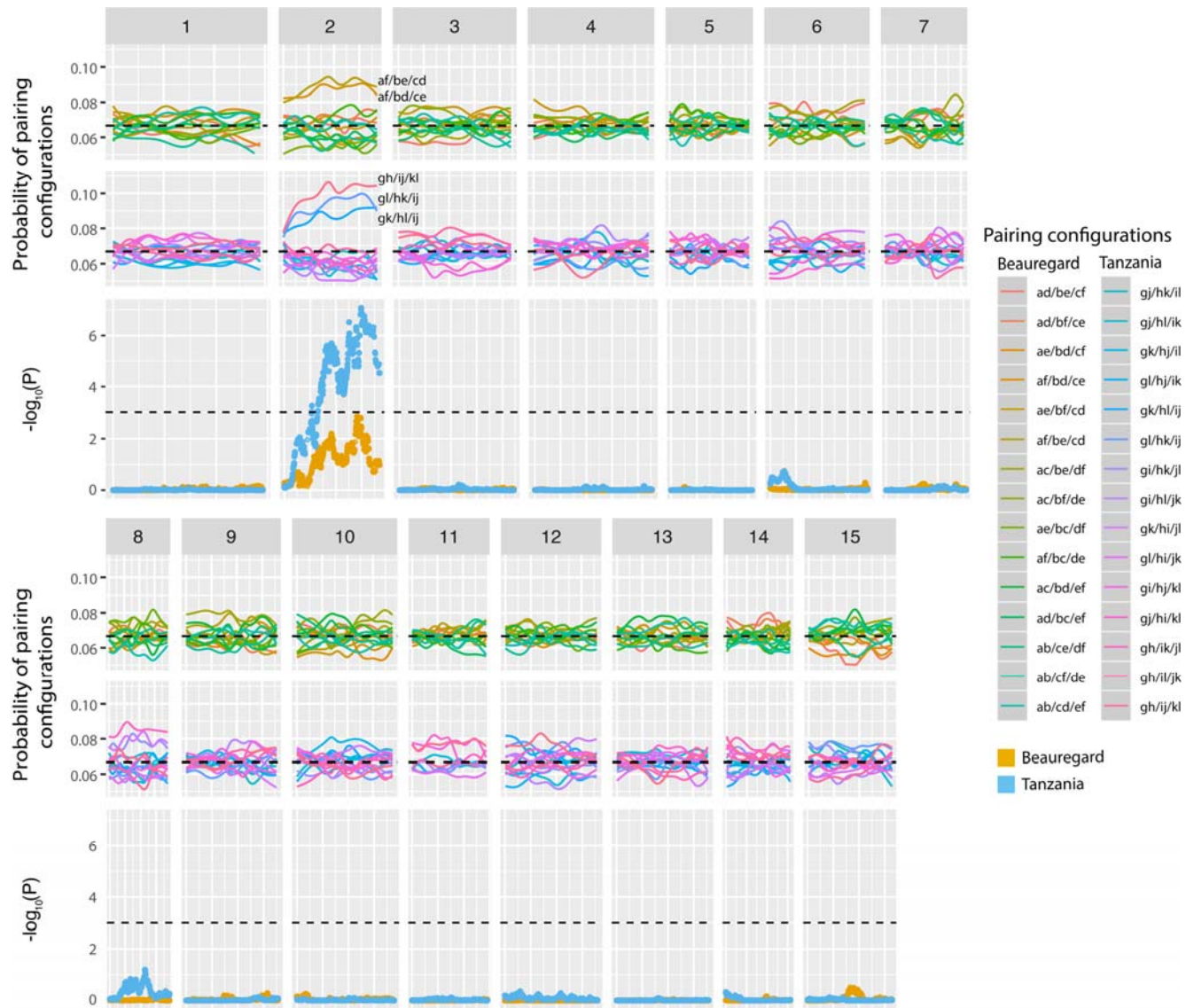


629

630 **Fig. 4** Percentage of maximum number of homologs connected in the same recombination chain  
631 during metaphase I in 'Beaugard' and 'Tanzania' for all 15 LGs. LG 11 for 'Tanzania' was  
632 mostly inconclusive and is not shown.

633





635 **Fig. 5** Preferential pairing profiles for 15 pairing configurations in parents 'Beauregard' and  
 636 'Tanzania' across 15 LGs. Notation  $ab/cd/ef$  indicates a configuration where homolog  $a$  paired  
 637 with  $b$ ,  $c$  with  $d$  and  $e$  with  $f$ . The dashed line in the probability profiles indicate the pairing  
 638 probability expected under random pairing ( $\frac{1}{15} \sim 0.067$ ). Orange and blue dots represent  $-\log_{10} P$   
 639 of a chi-square independence test for 'Beauregard' and 'Tanzania', respectively. Dashed line  
 640 indicates  $P = \frac{0.05}{15}$ . LG 2 presented a low, but significant preferential pairing involving three  
 641 bivalent configurations ( $gh/ij/kl$ ,  $gl/hk/ij$ , and  $gk/hl/ij$ ). Homologs  $i$  and  $j$  appear in all three  
 642 configurations, indicating a preferential association between these chromosomes.

643 **Table 1. Summary of sweetpotato genetic map**

LG	Length (cM)	Number of Markers			Total	SNPs/cM
		Simplex	Double-simplex	Multiple-dose		
1	290.9	1216	318	1211	2745	9.4
2	184.6	857	197	673	1727	9.4
3	222.1	1085	285	1052	2422	10.9
4	227.1	1372	377	1287	3036	13.4
5	157.1	892	193	816	1901	12.1
6	189.3	970	266	656	1892	10.0
7	156.3	1005	234	612	1851	11.8
8	115.5	712	140	312	1164	10.1
9	178.1	1403	261	715	2379	13.4
10	1188.7	1106	234	822	2162	11.5
11	145.6	724	177	729	1630	11.2
12	181.0	1367	246	1048	2661	14.7
13	180.1	761	174	743	1678	9.3
14	125.3	667	96	590	1353	10.8
15	166.6	1019	265	799	2083	12.5
Total	2708.3	15156	3463	12065	30684	11.3

644