1 **Crafting for a better MAGIC: systematic design and test for multiparental**

2 **advanced generation inter-cross population.**

3 Chin Jian Yang[*]

4 Rodney N. Edmondson[†]

5 Hans-Peter Piepho[‡]

6 Wayne Powell[*]

7 Ian Mackay[*,§]

8 [*]Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh,

9 EH9 3JG, UK.

10 [†]Rana House 18 Church Street, Wellesbourne, Warwick, CV35 9LS, UK.

11 [‡]Biostatistics Unit, Institute of Crop Science, University of Hohenheim, 70593,

12 Stuttgart, Germany.

13 [§]IMplant Consultancy Ltd., Chelmsford, UK.

14 **Short running title**

15 MAGIC population design

16

17 **Keywords**

18 Multiparental advanced generation inter-cross (MAGIC)

19 Multiparental population (MPP)

20 Population design

21 Quantitative genetics

22 Genetic diversity

23

24 **Corresponding author**

25 Ian Mackay

26 Scotland's Rural College (SRUC), Kings Buildings, West Mains Road, Edinburgh,

27 EH9 3JG, UK.

28 ian.mackay@sruc.ac.uk

Abstract

Multiparental advanced generation inter-cross (MAGIC) populations are valuable crop resources with a wide array of research uses including genetic mapping of complex traits, management of genetic resources and breeding of new varieties. Multiple founders are crossed to create a rich mosaic of highly recombined founder genomes in the MAGIC recombinant inbred lines (RILs). Many variations of MAGIC population designs exist; however, a large proportion of the currently available populations have been created empirically and based on similar designs. In our evaluations of five MAGIC populations, we found that the choice of designs has a large impact on the recombination landscape in the RILs. The most popular design used in many MAGIC populations has been shown to have a bias in recombinant haplotypes and low level of unique recombinant haplotypes, and therefore is not recommended. To address this problem and provide a remedy for the future, we have developed the "magicdesign" R package for creating and testing any MAGIC population design via simulation. A Shiny app version of the package is available as well. Our "magicdesign" package provides a unifying tool and a framework for creativity and innovation in MAGIC population designs. For example, using this package, we demonstrate that MAGIC population designs can be found which are very effective in creating haplotype diversity without the requirement for very large crossing programmes. Further, we show that interspersing cycles of crossing with cycles of selfing is effective in increasing haplotype diversity. These approaches are applicable in species which are hard to cross or in which resources are limited.

Introduction

The multiparental advanced generation inter-cross (MAGIC) population was initially proposed in crops by Mackay and Powell (2007) as a highly recombined population derived from multiple founders. The first MAGIC population was produced using 19 founders in *Arabidopsis thaliana* (Kover et al. 2009). The MAGIC pedigree described by Cavanagh et al. (2008) has served as a foundation for the design of many MAGIC populations in subsequent years. Briefly, the MAGIC pedigree shows a single funnel going from 8 founders to a recombinant inbred line (RIL). Starting with 8 founders labelled as A to H, two-way crosses are made as (A × B), (C × D), (E × F) and (G × H). Next, four-way crosses are made as ((A × B) × (C × D)) and ((E × F) × (G × H)). Lastly, eight-way crosses are made as (((A × B) × (C × D)) × ((E × F) × (G × H))) followed by several generations of selfing. Using this crossing scheme, the end of the funnel is a RIL with its genome composed of contributions from all 8 founders. Alternatively, a MAGIC population design may involve multiple funnels like the elite wheat MAGIC population by Mackay et al. (2014). Regardless of the designs, MAGIC RILs have diverse recombination landscape and rich mosaics of founder genomes (Scott et al. 2020).

Over the years, MAGIC populations have been used in various studies with great success. MAGIC populations are popular choices in mapping quantitative trait loci (QTLs), for examples, resistance QTLs in bread wheat (Stadlmeier et al. 2019), cold tolerance QTLs in maize (Yi et al. 2020) and high-throughput phenotype QTLs in rice (Ogawa et al. 2021). In addition to single trait analyses, multivariate analyses (multi-trait or multi-environment) have been demonstrated in MAGIC populations (Scutari et al. 2014, Verbyla et al. 2014). Diouf et al. (2020) used a tomato MAGIC population to dissect the underlying genetic-by-environment (G × E) and plasticity for

4

76   climate adaptation traits. Aside from QTL mapping, MAGIC populations are valuable

77   resources for genomic selection owing to their properties of highly recombined

78   genomes and large population size (Scott et al. 2021). Following that, there are

79   opportunities for using MAGIC RILs in breeding new varieties (Bandillo et al. 2013, Li

80   et al. 2013). With large numbers of founders, MAGIC populations also provide a

81   dynamic asset for the management of genetic resources (Thépot et al. 2015) and

82   may be used to improve our understanding of crop adaptation (Scott et al. 2021).

83   Given the longevity with a broad array of uses, MAGIC populations are an invaluable

84   community resource for creative and impactful research.

85   Considering the importance of MAGIC populations in crop research, we

86   sought to understand the relationship between MAGIC population designs and

87   population recombination landscape. We selected and analyzed five MAGIC

88   populations with publicly available marker genotypes for the founders and RILs,

89   genetic map positions and pedigrees. The selected populations comprise the UK

90   wheat 8-founder (Mackay et al. 2014), German wheat 8-founder (Sannemann et al.

91   2018), cowpea 8-founder (Huynh et al. 2018), tomato 8-founder (Pascual et al. 2015)

92   and UK wheat 16-founder (Scott et al. 2021) MAGIC populations. These MAGIC

93   populations were created from different designs. We contrasted the observed

94   recombinant haplotypes to expected (simulated) recombinant haplotypes in each

95   population. A comparable cross-population analysis can be challenging due to many

96   variables like genome sizes, marker genotyping platforms and numbers of founders.

97   Fortunately, there are two elite wheat 8-founder MAGIC populations (Mackay et al.

98   2014, Sannemann et al. 2018) that were genotyped with the same 90k SNP array

99   (Wang et al. 2015), which allowed us to directly compare the two populations in

5

100   greater depth. We found that the recombination landscape varies across designs

101   and that this variation is consistent across species.

102   Following our results, we have created the "magicdesign" package in R (R

103   Core Team 2021) for the purpose of creating and testing different MAGIC population

104   designs. Three major steps are involved in the package pipeline: design creation,

105   population simulation and comparative analysis. Users can create a design by either

106   specifying input variables or providing a custom pedigree. Once a design is created,

107   "magicdesign" converts it into a crossing scheme from the founders to final RILs and

108   simulates a population based on the crossing scheme. After multiple iterations of

109   simulation, distributions of recombinant haplotypes and founder genomes are

110   summarized. Results from one or more designs can be combined and compared

111   visually in plots. In addition, "magicdesign" produces a pedigree in both text and plot

112   formats that can be used as a guide to support crossing work in practice. Aside from

113   the described roles, "magicdesign" serves as a tool to advance the use of MAGIC in

114   future multiparental populations.

115

116   **Materials and Methods**

117   **Evaluation of MAGIC population designs**

118   We surveyed all available MAGIC populations that have been published to

119   date (including pre-prints) and identified five populations with publicly available

120   marker data. These five MAGIC populations include wheat with 8 UK elite founders

121   (Mackay et al. 2014), wheat with 8 German elite founders (Sannemann et al. 2018),

122   cowpea with 8 founders (Huynh et al. 2018), tomato with 8 founders (Pascual et al.

123   2015) and wheat with 16 UK diverse founders (Scott et al. 2021). These populations

124   are referred to as wheat-UK8, wheat-DE8, cowpea, tomato and wheat-UK16,

125   respectively (Table 1). These datasets were chosen because the marker data for the

126   founders and recombinant inbred lines (RILs), genetic map positions, and pedigree

127   are publicly available. The wheat-UK16 population is an exception as it has founder

128   genotype probabilities to compensate for the lack of genetic map positions. Links to

129   the source datasets are listed in Supplementary Text. Other publicly available

130   datasets like the wheat with 8 German founders (Stadlmeier et al. 2018), wheat with

131   8 Australian founders (Shah et al. 2019), rice with 8 founders (Raghavan et al.

132   2017), maize with 9 founders (Dell'Acqua et al. 2015) and *Arabidopsis* with 19

133   founders (Kover et al. 2009) are excluded because at least one component of the

134   data needed for our purpose is not present.

135   All five chosen MAGIC populations vary in numbers of RILs and marker

136   density (Table 1). The original wheat-UK8 dataset is made of 643 RILs and 18,599

137   markers while the original wheat-DE8 dataset is made of 910 RILs and 7,579

138   markers. To maintain a fair comparison between these two populations, we kept only

139   5,138 markers that are common between wheat-UK8 and wheat-DE8. In wheat-DE8,

140   missing data were previously imputed to numerical mean (twice the allele

141   frequency). These imputed marker data cannot be used in "qtl2" (Broman et al.

142   2019) for calculating founder genotype probabilities, so we reverted the imputed

143   marker data by converting any non-integer marker data to missing. The cowpea

144   dataset is made of 305 RILs and 32,114 markers after removing 16 markers in the

145   original dataset where the marker data is missing in at least one founder. The tomato

146   dataset is made of 238 RILs and 1,345 markers. The wheat-UK16 dataset is made of

147   504 RILs and 1,065,178 markers.

7

148 **Identification of recombinant haplotypes in MAGIC populations**

149   To identify recombinant haplotypes, the biallelic marker data in the RILs need

150 to be converted into founder genotypes. For each dataset except wheat-UK16, we

151 determined the founder genotypes in each RIL using the "qtl2" package (Broman et

152 al. 2019) in R (R Core Team 2021). We first calculated the founder genotype

153 probabilities using *calc.genoprob* function with error probability of 0.01 (1%) and

154 Haldane map function. Next, we inferred the founder genotypes from the

155 probabilities using *maxmarg* function with minimum probability of 0.5001. We chose

156 a slightly higher threshold than the previously used minimum probability of 0.5 by

157 Gardner et al. (2016). Since the genotype probabilities for all founders at each RIL's

158 marker sum to 1, the threshold of 0.5001 eliminates the risk of the *maxmarg* function

159 picking a founder genotype at random when there are two or more with the same

160 probability above the threshold. For the wheat-UK16 dataset, the founder genotype

161 probabilities are readily available. These probabilities were calculated from STITCH

162 (Davies et al. 2016), which is a different software but uses the same underlying

163 hidden Markov model (HMM) as "qtl2". We inferred the founder genotypes in the

164 wheat-UK16 dataset using an equivalent threshold of 1 because the probabilities are

165 provided as a sum of two allele probabilities. Markers without any founder genotype

166 probabilities above the threshold were set to missing.

167   Using the inferred founder genotypes in each dataset, we identified the

168 recombinant haplotypes at each breakpoint. The recombinant haplotype is a

169 combination of flanking founder genotypes at each breakpoint. For example, 1_5 is

170 the recombinant haplotype at a breakpoint where the two flanking founder genotypes

171 are founder 1 and 5. For a population with n founders, there are $n^2 - n$ individual

172 recombinant haplotypes. Therefore, in an 8-founder MAGIC population, there are 56

8

173   individual recombinant haplotypes (1_2, 1_3, …, 8_6, 8_7). In order to summarize

174   the results, we summed the counts of each individual recombinant haplotype in

175   every RIL, and averaging the counts across all RILs to obtain the mean counts of

176   individual recombinant haplotypes.

177   As a control, we simulated a similar MAGIC population based on the original

178   pedigree for each dataset and calculated the true counts of recombinant haplotypes.

179   We first derived an approximated crossing scheme from the pedigree. Since the

180   information on replicates is not always present in the pedigree, we assumed that no

181   replicates and considered all funnels to be independent. Next, we used the

182   "AlphaSimR" package (Gaynor et al. 2021) in R (R Core Team 2021) to simulate the

183   MAGIC populations for a total of 100 iterations. In order to keep track of founder

184   genotypes, we expanded each marker into $n-1$ markers with the same exact

185   genetic map position to prevent recombination among these markers. Using $n=4$

186   founders as example, the founders are coded as 000, 100, 010 and 001 across the

187   three expanded markers. This expanded marker system tracked the true founder

188   genotypes from the start to the end of simulation, and therefore allowed us to

189   calculate the true counts of recombinant haplotypes using the same method

190   described in the previous paragraph. In all datasets except for wheat-UK16, we used

191   the same genetic map positions as the actual datasets. In wheat-UK16, we used

192   equally-spaced markers at 0.5 cM since the genetic map positions were not available

193   for this dataset.

194   We included an additional control using a hybrid approach of actual and

195   simulated datasets. Specifically, we converted the founder genotypes in simulated

196   RILs into biallelic marker data and inferred the founder genotypes using the same

197   procedures as we did in the actual datasets. This approach was applied to all four

9

198 MAGIC datasets except wheat-UK16. The counts of recombinant haplotypes

199 identified from this approach provide an upper limit to the inferred number of founder

200 genotypes using "qtl2" (Broman et al. 2019). There is a caveat, however, that since

201 our simulation is based on an approximated crossing scheme, the outcomes of this

202 approach may not precisely represent the upper limit.

203 **Determination of unique or identical recombinant haplotypes**

204   Following the identified recombinant haplotypes in wheat-UK8 and wheat-

205 DE8, we classified these recombinant haplotypes into unique or identical groups

206 based on their positional overlaps within an interval. Recombinant haplotypes of the

207 same founder pairs are considered identical if they fall within the same interval,

208 otherwise unique if they do not overlap. The intervals are arranged in non-

209 overlapping bins of approximately 1 or 10 cM from the start to end of chromosome.

210 Identical recombinant haplotypes exist due to replications of cross progeny in the

211 MAGIC population. While this classification cannot distinguish between identical and

212 independent recombinant haplotypes within the interval, the probability of

213 independent recombinant haplotypes is low and assumed equal between wheat-UK8

214 and wheat-DE8. Therefore, the results from this comparison can elucidate the effect

215 of MAGIC population designs on the proportions of unique against identical

216 recombinant haplotypes.

217 **Minimum probability in calling founder genotypes**

218   The minimum probability used in calling founder genotypes determines the

219 power in identifying the correct founder genotypes in the RILs. We selected 10

220 thresholds ranging from 0.1 to 1.0 with an increment of 0.1. We applied each

221 threshold to the *maxmarg* function in "qtl2" (Broman et al. 2019) in simulated

222 populations based on wheat-UK8 and wheat-DE8. These simulated populations are

223 similar to the previously described hybrid approach where the simulated founder

224 genotypes in RILs are converted to biallelic markers prior to calculating genotype

225 probabilities. For each threshold, we computed the proportions of correct, incorrect

226 or missing founder genotypes by comparing the inferred to true founder genotypes.

227 The ideal threshold should have a high proportion of correct founder genotypes with

228 low proportions of incorrect and missing founder genotypes.

229 **Marker density in MAGIC population**

230   We used the proportion of recombinant haplotype recovered (PRHR) to

231 quantify the recombinations in MAGIC population that is captured in the marker data.

232 We calculated the empirical PRHR in all five datasets as the counts of recombinant

233 haplotypes in the actual dataset divided by the counts in the simulated dataset.

234 Since the marker density is not constant along the genomes in actual datasets, we

235 sought to determine a clearer relationship between PRHR and marker density via

236 simulation. We simulated a single chromosome of 200 cM with 8 founders crossed

237 using the same design as wheat-UK8. The founder alleles were simulated based on

238 the correlations among founders in wheat-UK8 using the *rmvbin* function in "bindata"

239 package (Leisch et al. 2021) in R (R Core Team 2021). We simulated a total of 4,000

240 markers that are equally spaced at 0.05 cM. To test lower marker densities, we

241 thinned the same simulated marker data to 0.10, 0.20, 0.40, 0.80, 1.60, 3.20, 6.40,

242 12.80 cM respectively. For each marker density, we inferred the founder genotypes

243 in the RILs using "qtl2" (Broman et al. 2019) at a probability threshold of 0.5001 and

244 calculated the counts of recombinant haplotypes. Lastly, we obtained the PRHR by

245 taking the counts of recombinant haplotypes for each marker density divided by the

246 true counts.

Data Availability

248 The "magicdesign" package and its installation instructions are available for

249 download at https://github.com/cjyang-sruc/magicdesign. Detailed instructions are

250 available at https://cjyang-sruc.github.io/magicdesign_vignette. The Shiny app

251 "magicdesignee" can be found at https://magicdesign.shinyapps.io/magicdesignee/.

252 R scripts used in all analyses can be found at https://cjyang-

253 sruc.github.io/files/magicdesign_analysis.R.

254

Results

Classifications of MAGIC population designs

257 Variations in the MAGIC population designs can be described by the number

258 of founders and the crossing scheme (Figure 1A). It is convenient to first consider

259 two classes of designs based on the number of founders: power of two (P2) and

260 non-power of two (NP2). As the names suggest, the P2 class has $n = 2^i$ founders for

261 any $i > 1$ whereas the NP2 class has $n \neq 2^i$ and $n > 2$ founders. P2 designs are

262 generally easier to implement in practice because the numbers of individuals in a

263 funnel are halved in every crossing generation. For either P2 or NP2 classes, the

264 crossing scheme can be structured, unstructured and semi-structured (Figure 1A). A

265 structured design involves strictly defined crosses among the founders and

266 intermediaries such that the crossing scheme can be further classified into full,

267 partial balanced, partial unbalanced or basic designs. These designs are elaborated

268   further in subsequent paragraphs. On the other hand, an unstructured design

269   involves random crosses among the founders and intermediaries, while a semi-

270   structured design is a combination of structured and unstructured designs. Additional

271   features of a structured design include: (1) precise tracing of the ancestry of each

272   RIL back to its progenitors, (2) number of crossing generations is equal to $log_2 n$

273   rounded up to the nearest integer. These features may not hold true in unstructured

274   or semi-structured designs.

275   Within a structured design, there are two primary types based on the number

276   of funnels: full and partial (Figure 1A). A full P2 design has $n!/2^{n-1}$ funnels while a

277   partial P2 design has one or more funnels but less than $n!/2^{n-1}$ funnels. The

278   numerator is the total number of permutations of 1 to n founders, and the

279   denominator is the total number of equivalent permutations by the MAGIC definition.

280   Directions of crosses are disregarded in defining a funnel. The denominator can be

281   described as $2^{n-1} = \prod 2^{n/x}$ where $x = 2^i \leq n$ for $i > 0$. In an example with four

282   founders, 1234, 1243, 2134, 2143, 3412, 3421, 4312 and 4321 are all equivalent

283   funnels. For simplicity, ((1 × 2) × (3 × 4)) is written as 1234. Full and partial types

284   exist in NP2 designs although the number of funnels in a full design cannot be

285   generalized similarly (Table S1). From a practical perspective, a full P2 design is

286   achievable for four or eight founders but not for 16 or more founders as the number

287   of required funnels becomes unmanageable.

288   Within a partial design, the funnels can be chosen in either a balanced or an

289   unbalanced way (Figure 1A). A balanced design has an equal number of founders

290   among the funnels and equal frequency of founder pairs at each crossing

291   generation. In a four-founder MAGIC design, 1234, 1324 and 1423 form a set of

292   three balanced funnels. First, each founder occurs thrice in the set of funnels.

293   Second, each founder is paired once with another founder in the two-way crosses,

294   and twice in the four-way crosses. For example, founder 1 meets founder 2 once in

295   the two-way cross (first funnel) and twice in the four-way cross (second and third

296   funnels). Coincidentally, since $n = 4$ and $n!/2^{n-1} = 4!/2^3 = 3$, the set of three

297   balanced funnels is equivalent to a full design for four founders. Unlike the partial

298   balanced design where the number of funnels is restricted to set rules, the partial

299   unbalanced design is formed by funnels chosen randomly. Differences between

300   balanced and unbalanced designs are explored in a later section. Additionally, we

301   coin the special case of partial unbalanced design with one funnel as a basic design.

302   Examples of all of the designs are shown in Figure S1.

303   Based on our survey of 48 MAGIC populations in 15 crop species that have

304   been described in either published or pre-print literature to date, there are 39 P2 and

305   9 NP2 designs (Figure 1B and Table S2). The numbers of founders range from four

306   to 60, with 4 and 8 founders as the predominant numbers. Despite the ease of

307   handling required for crosses based on a full design with 4 founders, all 10 of the

308   populations were created using a basic design. Of the 26 MAGIC populations with 8

309   founders, there are 16 basic designs, nine partial designs and one semi-structured

310   design. The popularity of the basic design can be ascribed to Cavanagh et al.

311   (2008), who provided an illustrated pedigree of a basic design. We refrained from

312   classifying the partial designs into balanced and unbalanced designs due to the lack

313   of pedigree information in many MAGIC populations. Regardless of the number of

314   founders, there has not been any MAGIC population created with a full design. There

315   are several 8-founder populations that came close to a full design. The bread wheat

316   MAGIC population by Mackay et al. (2014) had 210 out of 315 required funnels for a

317   full design. The maize MAGIC population by Dell'Acqua et al. (2015) had mixed

318 funnels from pooling different four-way individuals and had to introduce an additional

319 founder due to a failed two-way cross. The three bread wheat MAGIC populations by

320 Shah et al. (2019) came closest to a full design with 311 to 313 funnels.

321 **Empirical evaluation of two bread wheat MAGIC populations**

322 Our evaluation on two bread wheat MAGIC populations derived from distinct

323 sets of 8 elite founders shows that the distributions of recombinant haplotypes differ

324 for each MAGIC design (Figure 2 and Figure S2). We used the wheat-UK8 and

325 wheat-DE8 populations, in which wheat-UK8 is an example of a partial design while

326 wheat-DE8 is an example of a basic design (Table S2). To maintain our cross-

327 population comparison as fair as possible, we reduced the original wheat-UK8 and

328 wheat-DE8 datasets to smaller subsets with common markers (Figure 2), although

329 the same analysis was performed on the original datasets too (Figure S2). The

330 subsets include all 643 RILs in wheat-UK8 and 910 RILs in wheat-DE8, and 5,138

331 common markers arranged in the same genetic map positions as Gardner et al.

332 (2016). This genetic map is chosen over the original genetic map in the wheat 90k

333 array (Wang et al. 2014) because of higher map quality.

334 The distribution of all recombinant haplotypes is less skewed in wheat-UK8

335 than in wheat-DE8 (Figure 2). In wheat-UK8, none of the recombinant haplotype

336 appears more frequently than others (Figure 2A). In any given RIL, there are 0.879 ±

337 0.227 (mean ± standard deviation) individual recombinant haplotypes. In wheat-DE8,

338 eight recombinant haplotypes appear about twice as frequently as the others (Figure

339 2B). There are 1.910 ± 0.313 of these eight recombinant haplotypes (1_2, 2_1, 3_4,

340 4_3, 5_6, 6_5, 7_8 and 8_7) instead of 0.845 ± 0.150 of the other recombinant

341 haplotypes. In addition, the mean count of recombinant haplotype is approximately

342   normally distributed in wheat-UK8 (Figure 2C) but is skewed to the right in wheat-

343   DE8 (Figure 2D). The eight skewed recombinant haplotypes match with all of the

344   founder pairs in two-way crosses in wheat-DE8. This is not a coincidence because

345   two-way crosses have the largest founder genomes to recombine. With every

346   generation of crosses, the founder genomes are halved and so there are less

347   recombinations between any two founders.

348   While wheat-UK8 has a slightly lower number of recombinant haplotypes per

349   RIL than wheat-DE8 in both reduced (Table 2) and full (Table S3 and S4) datasets,

350   the proportion of unique recombinant haplotypes is higher in wheat-UK8 than in

351   wheat-DE8 (Figure 3 and Figure S3). Due to the imprecision of inferred

352   recombination breakpoints, we defined recombinant haplotypes with breakpoints

353   within any non-overlapping intervals as identical. We chose the intervals to be 1 cM

354   and 10 cM wide. With the interval width set to 1 cM, the number of identical

355   recombinant haplotypes in wheat-UK8 ranges from 1 (unique) to 21 (Figure 3A)

356   while the number in wheat-DE8 ranges from 1 (unique) to 128 (Figure 3B). In

357   addition, 11,602 out of 31,642 recombinant haplotypes in wheat-UK8 are considered

358   unique, and 9,567 out of 50,794 recombinant haplotypes in wheat-DE8 are

359   considered unique (Figure 3C). After discounting the identical recombinant

360   haplotypes, there are 17,786 distinct recombinant haplotypes distributed among 643

361   RILs in wheat-UK8, which is equivalent to 27.66 distinct recombinant haplotypes per

362   RIL. Similarly, there are 17,643 distinct recombinant haplotypes distributed among

363   910 RILs in wheat-DE8, which is equivalent to 19.39 distinct recombinant haplotypes

364   per RIL. When the interval is set to 10 cM, the counts and proportions of unique

365   recombinant haplotypes decrease and the differences between wheat-UK8 and

366   wheat-DE8 holds (Figure S3).

16

Empirical evaluation of three other MAGIC populations

367

368   While not directly comparable, the relationship between MAGIC population

369   designs and the distributions of recombinant haplotypes in three other datasets

370   remains consistent. Similar to wheat-DE8, the cowpea and tomato MAGIC

371   populations were created from a basic design and thus have a skewed distribution of

372   recombinant haplotypes (Figure 4). The recombinant haplotypes from two-way

373   founder pairs are higher than the other recombinant haplotypes. In cowpea, the two-

374   way recombinant haplotypes are 0.936 ± 0.171 (mean ± standard deviation) per RIL

375   while the other recombinant haplotypes are 0.384 ± 0.123 per RIL. In tomato, the

376   two-way recombinant haplotypes are 0.907 ± 0.095 per RIL and the other

377   recombinant haplotypes are 0.416 ± 0.116 per RIL. On the other hand, wheat-UK16

378   was created from a partial balanced design and does not have any skew in its

379   distribution of recombinant haplotypes (Figure S4). The recombinant haplotypes are

380   0.878 ± 0.102 per RIL.

Minimum probability for calling founder genotype

381

382   The minimum probability for calling founder genotype is important for the

383   identification of recombinant haplotypes, and our simulation results suggest that the

384   range of 0.4 to 0.6 gives a good balance of correct, incorrect and missing founder

385   genotype calls (Figure 5A and 5B). The results are similar between simulated wheat-

386   UK8 and wheat-DE8 populations, so only results from the simulated wheat-UK8

387   population are elaborated here. At a minimum probability of 0.4, the correct, incorrect

388   and missing founder calls are 69%, 16% and 15% of the total markers, respectively.

389   At a minimum probability of 0.5, the rates are 64%, 11% and 25%. At a minimum

390   probability of 0.6, the rates are 58%, 6% and 36%. As the minimum probability

391   increases, the rates of correct and incorrect founder calls decrease while the missing

392   rate increases. In order to avoid the issue of having two or more founder probabilities

393   above the threshold, the minimum probability can be set to 0.5 or higher.

394   **Marker density in MAGIC population**

395   In all five analyzed datasets, the proportion of recombinant haplotypes

396   recovered (PRHR) is higher in populations genotyped at higher marker density

397   (Table 1). PRHR is computed by taking the number of recombinant haplotypes in

398   actual dataset divided by the true number of recombinant haplotypes in simulated

399   dataset. In wheat-UK8 and wheat-DE8 with an average marker distance of 1.024 cM,

400   the PRHR is approximately one-third (Figure 2). Even though the tomato population

401   is genotyped at a lower marker density with an average marker distance of 1.603

402   cM, the PRHR is higher than in the two wheat populations (Figure 4B). This is likely

403   because the markers on the wheat D-genome are generally sparse due to its low

404   diversity (Akhunov et al. 2010). The cowpea population is genotyped at a high

405   marker density with an average marker distance of 0.030 cM, in which the PRHR is

406   approximately two-thirds (Figure 4A). Lastly, wheat-UK16 is genotyped at the highest

407   marker density of all analyzed datasets with an average marker distance of 0.005

408   cM, and it has the highest PRHR of almost 80% (Figure S4). Marker density is an

409   important factor in identifying fine-scale recombination breakpoints in MAGIC

410   populations.

411   Under ideal conditions, where the markers are evenly spaced, a marker

412   distance of 0.20 cM or less between two adjacent markers is sufficient to achieve a

413   PRHR of at least 90% (Figure 5C). We tested recombinant haplotype recovery rates

414   for markers that are evenly spaced across 0.05, 0.10, 0.20, 0.40, 0.80, 1.60, 3.20,

415 6.40 and 12.80 cM. At the smallest tested distance of 0.05 cM, approximately 97% of

416 the true recombinant haplotypes can be recovered. As the distance increases, the

417 recovery rate decreases. At the largest tested distance of 12.80 cM, approximately

418 11% of the true recombinant haplotypes can be recovered. These results are more

419 optimistic than the actual results (Table 1). In practice, more markers are required to

420 achieve the same recovery rate for any given marker density because markers are

421 not evenly distributed across the whole genome. In addition, the discrepancy

422 between simulated and actual results can also be attributed to marker quality. For

423 example, the markers on the wheat D-genomes are generally sparser than on the

424 others.

425 **magicdesign: a tool to create and test MAGIC population designs**

426 Given that MAGIC population construction requires a lot of time and effort,

427 and that design choices can impact population attributes, there is a need for a "free

428 trial" before committing to create a MAGIC population. Here, we introduce an R

429 package called "magicdesign", which is specifically made for creating and testing

430 various MAGIC population designs via simulation. Alternatively, we also provide a

431 user-friendly Shiny app version called "magicdesignee" which implements the

432 "magicdesign" R package in its back-end. Therefore, minimal R knowledge required

433 for users to use "magicdesignee".

434 Briefly, the "magicdesign" package workflow can be described as: (1) design

435 creation, (2) population simulation, and (3) comparative analysis. In the design

436 creation step, the package creates a crossing scheme that spans from the founders

437 to the final RILs based on user inputs. In the population simulation step, the package

438 simulates a MAGIC RIL population constructed from the crossing scheme, and

439   repeats over multiple iterations. At this point, the first two steps may be repeated for

440   other MAGIC population designs. Finally, in the comparative analysis step, the

441   package extracts information from previously tested designs and summarizes the

442   results illustratively. Additional details on each step are described in subsequent

443   sections.

Design creation

445   In a structured design, the design creation step takes various user inputs to

446   create a crossing scheme. The major inputs include number of founders, number of

447   funnels or funnel sets, and a balanced design indicator. Based on how these inputs

448   are specified, one of the structured designs (Full, Partial Balanced, Partial

449   Unbalanced, Basic) as shown in Figure 1A is created. As defined previously, a

450   balanced design has an equal number of founders among the funnels and equal

451   frequency of founder pairs at each crossing generation. This design creation step

452   works for either power of 2 (P2) or non-power of 2 (NP2) number of founders.

453   Currently, the allowed range of number of founders is any integer between 3 and

454   128. The allowed number of funnels or funnel sets varies according to the number of

455   founders and the balanced design indicator, and the full list is provided in Table S1.

456   Finding a balanced design requires more computation power than finding an

457   unbalanced design. This is because the balanced design requires many funnel

458   permutations to be evaluated while the unbalanced design randomly sample the

459   required number of funnel permutations. To reduce the computational burden, we

460   have identified alternative methods that are less computationally intensive. In the

461   case of 8 founders, we have searched through all 315^7 possible combinations and

462   identified 720 partial balanced funnel sets. Each of these partial balanced funnel set

463 has 7 funnels, and can be combined with another non-overlapping partial balanced

464 funnel set to form a larger partial balanced funnel set. In the case of 16 founders, the

465 number of possible combinations is very large and so we opted for a different

466 approach. To start, we obtained the 15 funnels from Scott et al. (2021), which is a

467 partial balanced set for 16 founders. We searched through all 3^{15} possible

468 permutations of eight- and sixteen-way crosses in these funnels and identified 7,776

469 partial balanced funnel sets. More partial balanced funnel sets could be found by

470 searching through all 315^{15} possible permutations of four-way crosses, however, that

471 was beyond our available computational capacity. Unlike the case of 8 founders,

472 these funnel sets do overlap and thus cannot be combined to form a larger set.

473 Instead, by randomly swapping the founders from a starting partial balanced funnel

474 set, a non-overlapping partial balanced funnel set can be created and merged to

475 form a larger set. For other numbers of founders between 4 and 16, the balanced

476 design is created based on a nested incomplete block design (NIBD) generated

477 using the "blocksdesign" package (Edmondson 2020; Edmondson 2021). A MAGIC

478 funnel is analogous to a NIBD as the founders in two-way crosses (experimental

479 block of two plots) are nested within four-way crosses, founders in four-way crosses

480 are nested within eight-way crosses, and so on. Currently, a balanced design in

481 "magicdesign" is limited to 16 or less founders as there is not yet an efficient method

482 for larger number of founders.

483   In addition, "magicdesign" provides options to further modify the MAGIC

484 population design by specifying the number of replicates, number of selfing

485 generations, and an additional crossing indicator. The number of replicates

486 determines how many seeds from a cross are retained. This can help to increase the

487 haplotype diversity in the MAGIC population when the seeds are not genetically

488 identical. In the case of inbred founders, replicates of two-ways individuals are all

489 identical but not replicates of four-ways (or higher) individuals. The number of selfing

490 generations determine how many generations of selfing are required after each

491 cross. Typically, the selfing step is only applied after the last crossing generation as

492 a way to reduce heterozygosity in the RILs. However, selfing prior to that may be

493 beneficial in increasing recombinant haplotypes. Lastly, the additional crossing

494 indicator allows for an extra crossing generation to further increase recombinant

495 haplotypes. This is similar to the approach taken by Stadlmeier et al. (2018) and

496 Shah et al. (2019).

497   Alternatively, any MAGIC population design that is not available directly in

498 "magicdesign" can be created by supplying a complete pedigree. The only

499 requirement for the pedigree is that it must detail all crosses involved from the

500 founders to the final RILs. This option provides a greater flexibility to accommodate

501 for semi-structured or unstructured designs. Furthermore, it is also possible to modify

502 a design created from "magicdesign" and provide the pedigree of the modified

503 design.

504 **Population simulation**

505   Once a MAGIC population design is created, "magicdesign" simulates a

506 population based on the design and other user inputs. The major inputs include

507 distance between markers, chromosome genetic lengths, number of simulations and

508 recombinant haplotype interval size. The simulation step will create evenly-spaced

509 markers based on the distance between markers and chromosome genetic lengths.

510 All founders are considered unique and so each of these markers is used to encode

511 for the founder genotypes. The desired number of simulations is selected. In

22

512   addition, the recombinant haplotype interval size determines the distance between

513   two markers to look for recombinant haplotypes.

514   **Comparative analysis**

515   After simulating one or more designs, the final step is to compare the design

516   qualities in terms of recombinant haplotype proportions and distribution of founder

517   genomes in the RILs. In general, a good MAGIC population design should yield

518   consistently higher recombinant haplotype proportions as well as an even distribution

519   of founder genomes compared to other designs.

520   To demonstrate comparative analysis with "magicdesign", the five designs in

521   Table 3 are used as examples. These designs are all applied to a fictitious species

522   with five chromosomes of 1.0, 1.5, 2.0, 2.5, 3.0 centiMorgans (cM) length. All five

523   designs are created based on a MAGIC population of 8 founders. Design 1 is a full

524   design and so it has all 315 funnels. Design 2 and 4 are both partial balanced design

525   with one funnel set (7 funnels), and the only difference between them is that the four-

526   way individuals in design 4 are selfed once before making eight-way crosses. Design

527   3 is similar to design 2 except it is a partial unbalanced design with 7 funnels. Lastly,

528   design 5 is a basic design with 1 funnel inspired by the design used in Stadlmeier et

529   al. (2018). The numbers of replicates are varied for each design to achieve similar

530   final RIL population size close to 1,000. Aside from design 1 which has the highest

531   number of crosses, the other designs have fairly similar numbers of crosses. Design

532   1, 2 and 3 require 7 generations from founders to RILs, while design 4 and 5 require

533   8 generations because of the additional selfing and crossing generation respectively.

534   First, we investigated the designs' effects on recombinant haplotypes within a

535   5 cM interval. In term of total recombinant haplotypes, a good design should have

23

536 high mean with low variance. For design 1 to 5 respectively, the means are 0.167,

537 0.167, 0.169, 0.186 and 0.202 while the variances are 0.000158, 0.000244,

538 0.000293, 0.000452 and 0.003000 (Figure 6A). The means are similar in design 1, 2

539 and 3, slightly higher in design 4 and highest in design 5. However, the variances are

540 lowest in design 1, similar in design 2 and 3, slightly larger in design 4, and

541 substantially larger in design 5.

542   In any RIL derived from an 8-founder MAGIC population, there are 56 distinct

543 recombinant haplotypes and a good design should have high mean with low

544 variance. The means for the number of unique recombinant haplotypes are 52.20,

545 49.93, 50.29, 47.81 and 34.51 for design 1 to 5 respectively while the variances are

546 3.31, 4.29, 4.21, 4.68 and 34.78 for design 1 to 5 respectively (Figure 6B). The

547 means are highest in design 1, similar in design 2 and 3, slightly lower in design 4

548 and lowest in design 5. The variances follow a similar but reverse trend as the

549 means except for design 5 where the variance is over seven times greater.

550 Equivalently, the coefficients of variation (CVs) are 0.035, 0.041, 0.041, 0.045 and

551 0.171 for design 1 to 5 respectively. That for design 5 is approximately four times

552 that for the other designs.

553   The distributions of individual recombinant haplotype should be consistent

554 across all recombinant haplotypes with minimal variability across simulations in a

555 good design. With the exception of design 5, all other designs have similar

556 distributions of individual recombinant haplotype (Figure 6C and Table S5). Similar to

557 wheat-DE8 (Figure 2B), cowpea and tomato (Figure 4), design 5 has more two-ways

558 recombinant haplotypes than other recombinant haplotypes. Furthermore, the

559 spreads of two-ways recombinant haplotypes in design 5 are much higher than the

560 others, which imply low consistency.

561   Similar to the previous criterion, the proportions of founder genomes should

562   be consistent across all founders with low variability across simulations in a good

563   design. With 8 founders, the expected proportion of each founder genome in a

564   population is 0.125. The proportions are within 0.01 of expectation for all designs

565   except for design 5, which has 5 out of 8 proportions exceeding the range (Figure

566   7A). On the other hand, the variances are lowest in design 1, slightly higher in design

567   2, 3 and 4, and highest in design 5 (Figure 7A).

568   In any single chromosome, a RIL can carry tracts of 1 to 8 unique founder

569   genomes and it is generally better to have more unique founder genomes. In the

570   shortest chromosome (chromosome 1), design 1, 2 and 3 frequently produce 3

571   unique founders, while design 4 and 5 frequently produce 4 unique founders (Figure

572   7B). In the longest chromosome (chromosome 5), design 1, 2, 3 and 4 frequently

573   produce 6 unique founders while design 5 frequently produces 7 unique founders

574   (Figure 7B).

575   Lastly, a good design should have short non-recombinant segments. Across

576   all chromosomes, design 5 has the most short non-recombinant segments, followed

577   by design 4, and design 1, 2 and 3 being undiscernible (Figure 7C).

578   Of all the five designs considered here, each has its own advantages and

579   disadvantages. Design 1, 2 and 3 are highly similar except that design 1 tends to

580   show smaller variability than the other two at the cost of more crossing work

581   required. Design 4 is slightly better than the first three in most occasions, although it

582   is slightly more variable and requires one additional generation. Design 5 is generally

583   poor and should be avoided if possible, although the additional crossing generation

584   helps in increasing the number of unique founders and reducing non-recombinant

585 segment lengths. Of all designs considered, design 4 appears to be the best option if

586 the additional generation is acceptable, otherwise either design 2 or 3 is a good

587 option too. Across all the metrics used for comparisons in "magicdesign", there is no

588 observable difference between design 2 (balanced) and 3 (unbalanced).

589

590 **Discussion**

591 Ease of design appears as a major factor in driving the design choices in

592 currently available MAGIC populations. These MAGIC populations are predominantly

593 made of 4 or 8 founders crossed using a basic design (Figure 1B). There are several

594 possible explanations to the choice popularity. First, P2 designs are easier to handle

595 than NP2 designs since the individuals in every funnel are halved at every crossing

596 generation. Besides, 4 and 8 founders are effectively the lowest numbers of founders

597 available in P2 designs, and higher numbers of founders require more generations of

598 crossing and may increase the design complexity. Of all the explored designs, the

599 basic design likely requires the least amount of effort in population construction. The

600 only other option that may rival a basic design is the unstructured design with

601 random mating, which often relies on segregation of male sterility loci. Unfortunately,

602 this system is not always readily available in every species and may restrict the

603 founder choices.

604 Choice of MAGIC population design plays a critical role in determining the

605 recombination landscape in the RILs. In the comparison between wheat-UK8 and

606 wheat-DE8, we identified a bias in individual recombinant haplotypes in the basic

607 design but not the partial design (Figure 2). The bias resulted in more two-way

608 recombinant haplotypes than other recombinant haplotypes. The bias might be

26

609   exacerbated if the pairs of founders in two-ways are genetically more similar than

610   others, which can happen if the founders stratify into two or more groups. It is

611   possible to avoid pairing the founders of the same groups in two-ways if the grouping

612   is known. For example, Pascual et al. (2015) made the two-way crosses by crossing

613   tomato founders with large fruits to founders with small fruits, and Ogawa et al.

614   (2018) followed similarly by crossing *indica* rice founders to *japonica* rice founders.

615   This countermeasure is only possible if the numbers of founders are equal between

616   groups, but not if the founders cannot be subdivided equally like the barley

617   (Sannemann et al. 2015), cowpea (Huynh et al. 2018) and wheat (Stadlmeier et al.

618   2018) MAGIC populations.

619   In addition to the bias, the basic design also resulted in a lower proportion of

620   unique recombinant haplotypes than the partial design (Figure 3). Since a basic

621   design always has less funnels than any other designs, high replication of cross

622   progeny is required to bring the number of RILs up. In general, replicates reduce the

623   amount of crossing work required in prior generation by keeping more than one

624   progeny from a single cross to advance. The recombination landscape in these

625   replicated individuals is non-independent because any prior recombinations are

626   passed down from their parents. The detriments from replication can be minimized

627   by replicating in earlier generations as subsequent crosses will reduce the non-

628   independence among replicates. In a MAGIC population with 8 inbred founders, the

629   earliest meaningful replication would be the four-way individuals. However, replicates

630   prior to the final crosses do increase the amount of downstream crossing work, and

631   so it is important to consider the trade-offs between available work resources and

632   uniqueness of recombinant haplotypes.

633   High marker density is needed to capture the highly recombined genomes of

634   MAGIC RILs. We used the proportion of recombinant haplotypes recovered (PRHR)

635   as a measure of how well the markers capture recombinant haplotypes. PRHRs in

636   the five analyzed datasets correlate well with the marker density. Even with the high

637   marker density in wheat-UK16, the PRHR is only 0.799 (Table 1), which suggests

638   that one-fifth of the recombinant haplotypes is still missing. Some explanations

639   include the sparser marker density in the D-genomes, uneven marker density along

640   the genomes and segregation distortions of the founder genomes. To generalize the

641   relationship between marker density and recombinant haplotypes further, our

642   simulation results showed that marker distance of 0.80 cM or less is sufficient to

643   recover over three-quarters of the recombinant haplotypes. Despite the results from

644   actual datasets being less optimistic than the results from simulation, the importance

645   of high marker density in MAGIC populations still holds.

646   Given that the advantages and disadvantages of different MAGIC population

647   designs are largely unexplored, the "magicdesign" package serves as an important

648   tool to create and test different designs. Specifically, "magicdesign" provides the

649   opportunity to evaluate the options before committing to years of effort in

650   constructing MAGIC populations. In our examples, an additional selfing generation

651   offers a simple path to improvement (Figure 6 and 7), especially in inbreeding

652   species. When used in combination with speed breeding (Watson et al. 2018), the

653   additional time due to selfing can be minimized. In addition, "magicdesign" also acts

654   as a bridging tool for researchers who are new to MAGIC populations by providing a

655   starting point to creating a MAGIC population. The opportunity to create and test

656   different designs will encourage innovation in MAGIC population designs rather than

657   relying on previously used designs. Overall, "magicdesign" is a valuable resource for

658   unifying the process of creating and testing MAGIC population designs, and

659   providing the flexibility for additional features to be included in future updates as the

660   package grows.

661

662   **Acknowldgements**

663   We thank Rajiv Sharma, Ian Dawson and David Marshall for helpful discussion. We

664   also thank the UK Crop Diversity group (https://www.cropdiversity.ac.uk/) for

665   providing the High Performance Computing (HPC) resource used in evaluating large

666   permutations.

Literature Cited

Bandillo, N., C. Raghavan, P. A. Muyco, M. A. L. Sevilla, I. T. Lobina *et al.*, 2013 Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6: 11.

Broman, K. W., D. M. Gatti, P. Simecek, N. A. Furlotte, P. Prins *et al.*, 2019 R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. Genetics 211: 495-502.

Cavanagh, C., M. Morell, I. Mackay, and W. Powell, 2008 From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr. Opin. Plant Biol. 11: 215-221.

Davies, R. W., J. Flint, S. Myers and R. Mott, 2016 Rapid genotype imputation from sequence without reference panels. Nat. Genet. 48: 965-969.

Dell'Acqua, M., D. M. Gatti, G. Pea, F. Cattonaro, F. Coppens *et al.*, 2015 Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol. 16: 167.

Diouf, I., L. Derivot, S. Koussevitzky, Y. Carretero, F. Bitton *et al.*, 2020 Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. J. Expt. Bot. 71: 5365-5376.

Edmondson, R. N., 2020 Multi-level block designs for comparative experiments. J. Agr. Biol. Envir. St. 25: 500-522.

Edmondson. R. N., 2021 blocksdesign: Nested and crossed block designs for factorial and unstructured treatment sets. R package version 4.9. https://cran.r-project.org/package=blocksdesign.

690   Gardner, K. A., L. M. Wittern and I. J. Mackay, 2016 A highly recombined, high-

691   density, eight-founder wheat MAGIC map reveals extensive segregation distortion

692   and genomic locations of introgression segments. Plant Biotechnol. J. 14: 1406-

693   1417.

694   Gaynor, R. C., G. Gorjanc and J. M. Hickey, 2021 AlphaSimR: an R package for

695   breeding program simulations. G3: Genes, Genomes, Genetics 11: jkaa017.

696   Huynh, B.-L., J. D. Ehlers, B. E. Huang, M. Muñoz-Amatriaín, S. Lonardi et al., 2018

697   A multi-parent advanced generation inter-cross (MAGIC) population for genetic

698   analysis and improvement of cowpea (Vigna unguiculata L. Walp.). Plant J. 93:

699   1129-1142.

700   Kover, P. X., W. Valdar, J. Trakalo, N. Scarcelli, I. M. Ehrenreich et al., 2009 A

701   multiparent advanced generation inter-cross to fine-map quantitative traits in

702   Arabidopsis thaliana. PLoS Genet. 5: e1000551.

703   Leisch, F., A. Weingessel and K. Hornik, 2021 bindata: generation of artificial binary

704   data. R package version 0.9-20. https://cran.r-project.org/package=bindata.

705   Li, X.-F., Z.-X. Liu, D.-B. Lu, Y.-Z. Liu, X.-X. Mao et al., 2013 Development and

706   evaluation of multi-genotype varieties of rice derived from MAGIC lines. Euphytica

707   192: 77-86.

708   Mackay, I, and W. Powell, 2007 Methods for linkage disequilibrium mapping in crops.

709   Trends Plant Sci. 12: 57-63.

710   Mackay, I. J., P. Bansept-Basler, T. Barber, A. R. Bentley, J. Cockram et al., 2014

711   An eight-parent multiparent advanced generation inter-cross population for winter-

712 sown wheat: creation, properties, and validation. G3: Genes, Genomes, Genetics 4:

713 1603-1610.

714 Ogawa, D., T. Sakamoto, H. Tsunematsu, N. Kanno, Y. Nonoue *et al.*, 2021

715 Haplotype analysis from unmanned aerial vehicle imagery of rice MAGIC population

716 for the trait dissection of biomass and plant architecture. J. Expt. Bot. 72: 2371-2382.

717 Pascual, L., N. Desplat, B. E. Huang, A. Desgroux, L. Bruguier *et al.*, 2015 Potential

718 of a tomato MAGIC population to decipher the genetic control of quantitative traits

719 and detect causal variants in the resequencing era. Plant Biotechnol. J. 13: 565-577.

720 R Core Team, 2021 *R: A language and environment for statistical computing.* R

721 Foundation for Statistical Computing, Vienna.

722 Raghavan, C., R. Mauleon, V. Lacorte, M. Jubay, H. Zaw *et al.*, 2017 Approaches in

723 characterizing genetic structure and mapping in a rice multiparental population. G3:

724 Genes, Genomics, Genetics 7: 1721-1730.

725 Sannemann, W., B. E. Huang, B. Mathew and J. Léon, 2015 Multi-parent advanced

726 generation inter-cross in barley: high-resolution quantitative trait locus mapping for

727 flowering time as a proof of concept. Mol. Breeding 35: 86.

728 Sannemann, W., A. Lisker, A. Maurer, J. Léon, E. Kazman *et al.*, 2018 Adaptive

729 selection of founder segments and epistatic control of plant height in the MAGIC

730 winter wheat population WM-800. BMC Genomics 19: 559.

731 Scott, M. F., O. Ladejobi, S. Amer, A. R. Bentley, J. Biernaside *et al.*, 2020 Multi-

732 parent populations in crops: a toolbox integrating genomics and genetic mapping

733 with breeding. Heredity 125: 396-416.

734 Scott, M. F., N. Fradgley, A. R. Bentley, T. Brabbs, F. Corke, *et al.*, 2021 Limited

735 haplotype diversity underlies polygenic trait architecture across 70 years of wheat

736 breeding. Genome Biol. in press.

737 https://www.biorxiv.org/content/10.1101/2020.09.15.296533v1.

738 Scutari, M., P. Howell, D. J. Balding and I. Mackay, 2014 Multiple quantitative trait

739 analysis using Bayesian networks. Genetics 198: 129-137.

740 Shah, R., B. E. Huang, A. Whan, M. Newberry, K. Verbyla *et al.*, 2019 The complex

741 genetic architecture of recombination and structural variation in wheat uncovered

742 using a large 8-founder MAGIC population. bioRxiv. doi: 10.1101/594317 (Preprint

743 posted March 31, 2019).

744 Stadlmeier, M., L. Hartl and V. Mohler, 2018 Usefulness of a multiparent advanced

745 generation intercross population with a greatly reduced mating design for genetic

746 studies in winter wheat. Front. Plant Sci. 9: 1825.

747 Stadlmeier, M., L. N. Jorgensen, B. Corsi, J. Cockram, L. Hartl *et al.*, 2019 Genetic

748 dissection of resistance to the three fungal plant pathogens *Blumeria graminis*,

749 *Zymoseptoria tritici*, and *Pyrenophora tritici-repentis* using a multiparental winter

750 wheat population. G3: Genes, Genomes, Genetics 9: 1745-1757.

751 Thépot, S., G. Restoux, I. Goldringer, F. Hospital, D. Gouache *et al.*, 2015 Efficient

752 tracking selection in a multiparental population: the case of earliness in wheat.

753 Genetics 199: 609-623.

754 Verbyla, A. P., C. R. Cavanagh and K. L. Verbyla, 2014 Whole-genome analysis of

755 multienvironment or multitrait QTL in MAGIC. G3: Genes, Genomes, Genetics 4:

756 1569-1584.

757 Wang, S., D. Wong, K. Forrest, A. Allen, S. Chao *et al.*, 2014 Characterization of

758 polyploid wheat genomic diversity using a high-density 90 000 single nucleotide

759 polymorphism array. Plant Biotechnol. J. 12: 787-796.

760 Yi, Q., R. A. Malvar, L. Alvarez-Iglesias, B. Ordas and P. Revilla, 2020 Dissecting the

761 genetics of cold tolerance in a multiparental maize population. Theor. Appl. Genet.

762 133: 503-516.

763 **Tables and Figures**

764 **Table 1. Summary of five analyzed MAGIC populations.**

765 The wheat-UK8 and wheat-DE8 datasets have been reduced to share the same
766 markers and maps for comparison. The proportion of recombinant haplotypes
767 recovered (PRHR) is calculated as number of recombinant haplotypes in actual
768 dataset divided by number of recombinant haplotypes in simulated dataset. PRHR is
769 shown as mean ± standard deviation.

dataset	n	genome (cM)	marker	distance (cM/marker)	PRHR	reference
wheat-UK8	8	5,262	5,138	1.024	0.296 ± 0.094	Mackay et al. (2014)
wheat-DE8	8	5,262	5,138	1.024	0.331 ± 0.058	Sannemann et al. (2018)
cowpea	8	979	32,114	0.030	0.674 ± 0.207	Huynh et al. (2018)
tomato	8	2,156	1,345	1.603	0.410 ± 0.106	Pascual et al. (2015)
wheat-UK16	16	5,262	1,065,178	0.005	0.799 ± 0.092	Scott et al. (2021)

770

Table 2. Number of informative recombinations in wheat-UK8 and wheat-DE8.

The number of informative recombinations (NR) is calculated for both simulated and actual wheat-UK8 and wheat-DE8 datasets. Note: recombinant inbred line (RIL), Morgan (M).

Chr	NR/RIL				NR/RIL/M			
	wheat-UK8		wheat-DE8		wheat-UK8		wheat-DE8	
	sim	actual	sim	actual	sim	actual	sim	actual
1A	7.21	1.91	7.21	2.35	3.13	0.83	3.13	1.02
1B	11.77	3.57	11.76	4.31	3.44	1.04	3.44	1.26
1D	3.74	0.68	3.73	0.99	2.92	0.53	2.92	0.78
2A	8.37	3.67	8.37	3.17	3.32	1.45	3.32	1.26
2B	12.43	2.63	12.43	3.61	3.34	0.71	3.34	0.97
2D	4.97	1.69	4.96	1.42	2.71	0.92	2.71	0.77
3A	10.21	4.50	10.19	4.16	3.37	1.48	3.36	1.37
3B	9.89	3.35	9.90	4.18	3.51	1.19	3.51	1.48
3D	4.45	0.61	4.44	0.45	2.29	0.32	2.28	0.23
4A	6.95	2.51	6.96	2.11	3.30	1.19	3.30	1.00
4B	7.55	2.21	7.58	2.23	3.40	0.99	3.41	1.01
4D	2.95	0.51	2.96	0.43	2.75	0.47	2.75	0.40
5A	10.23	4.15	10.23	4.47	3.26	1.32	3.26	1.43
5B	10.97	2.59	10.96	3.25	3.53	0.83	3.53	1.05
5D	4.76	1.33	4.75	1.52	2.39	0.67	2.39	0.76
6A	9.70	3.38	9.68	4.48	3.48	1.21	3.47	1.61
6B	8.85	1.97	8.84	2.46	3.40	0.76	3.40	0.95
6D	3.92	0.43	3.93	1.14	1.82	0.20	1.83	0.53
7A	13.53	4.21	13.50	5.07	3.53	1.10	3.52	1.32
7B	9.28	2.64	9.27	2.97	3.23	0.92	3.23	1.03
7D	4.40	0.74	4.40	1.05	2.36	0.40	2.35	0.56
All	166.10	49.24	166.05	55.84	3.16	0.94	3.16	1.06

36

Table 3. Five MAGIC population designs tested in magicdesign.

The numbers of replicates, selfing generations and crosses are listed separately for each generation. For example, in design 1, the eight-way individuals are replicated three times and then selfed for four generations. The total number of crosses is shown in parentheses.

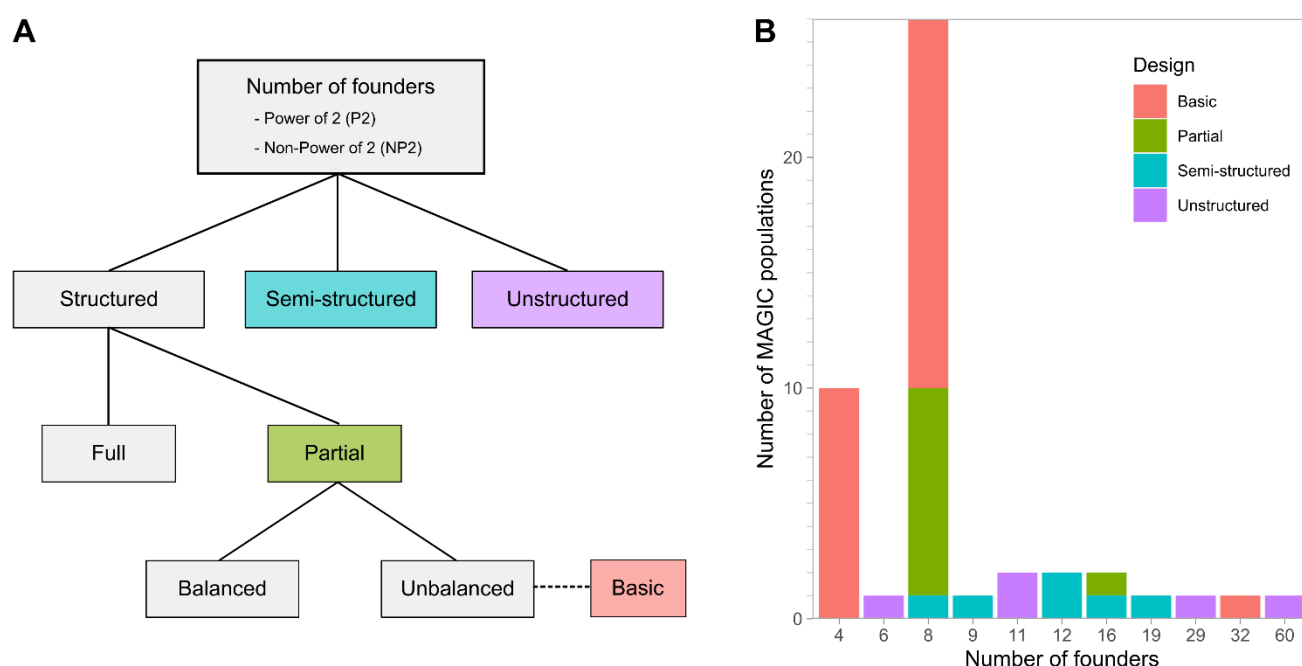|	Design 1	Design 2	Design 3	Design 4	Design 5
Founders	8	8	8	8	8
Type	Full	Partial balanced	Partial unbalanced	Partial balanced	Basic
Replicates	1, 1, 3	1, 9, 15	1, 9, 15	1, 9, 15	1, 4, 4, 15
Selfing	0, 0, 4	0, 0, 4	0, 0, 4	0, 1, 4	0, 0, 0, 4
Crosses	28, 210, 315 (553)	28, 14, 63 (105)	19, 13, 63 (95)	28, 14, 63 (105)	4, 2, 4, 64 (74)
Generations	7	7	7	8	8
RIL	945	945	945	945	960
Funnel	315	7	7	7	1

782

Figure 1. Classifications of MAGIC population designs.

[**A**] Flowchart of classifying MAGIC population designs based on their crossing schemes. [**B**] Distribution of MAGIC population designs in all 48 surveyed populations.
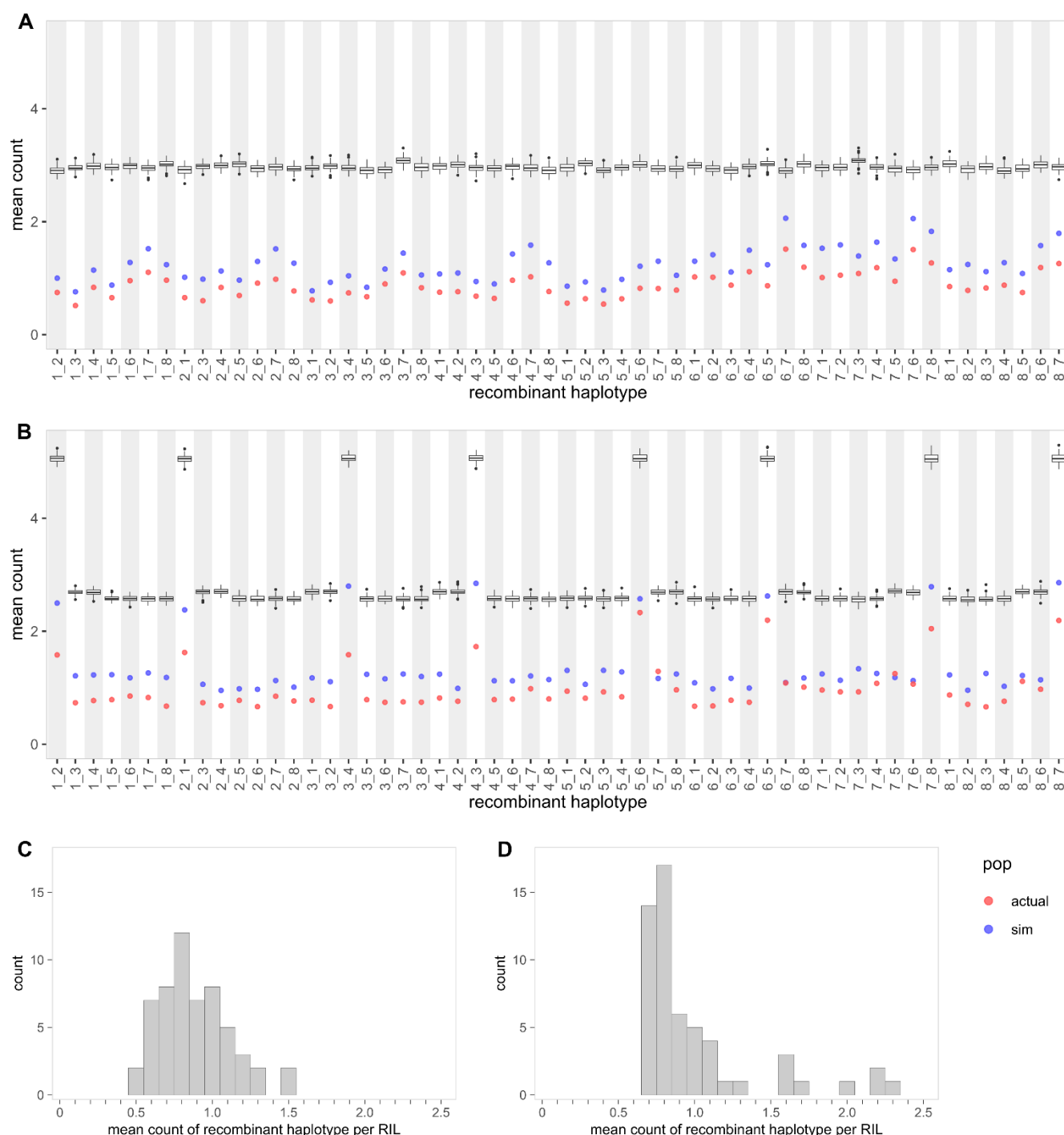
Figure 2. Distributions of recombinant haplotypes in two wheat MAGIC populations.

[**A**] Plot shows mean count of each recombinant haplotype in a single RIL in wheat-UK8. The boxplot shows mean count from true founder genotypes (100 simulated iterations). The red and blue points show mean count from inferred founder genotypes. [**B**] Plot shows mean count of each recombinant haplotype in a single RIL in wheat-DE8. [**C**] Histogram of the mean count in wheat-UK8. [**D**] Histogram of the mean count in wheat-DE8.
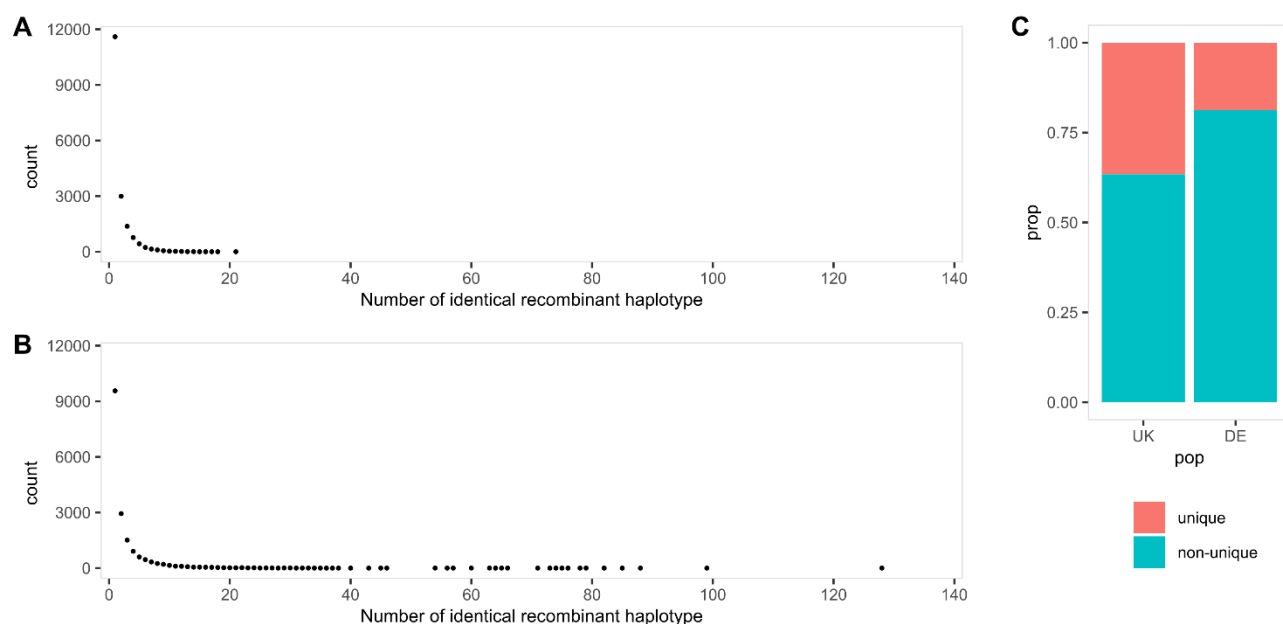
Figure 3. Distributions of unique and identical recombinant haplotypes in two wheat MAGIC populations.

Recombinant haplotypes are considered identical if they are of the same founder pairs and present in the same 1 cM interval, otherwise unique. [**A**] Counts of the number of identical recombinant haplotypes in wheat-UK8. The left most point is the count of unique recombinant haplotypes. [**B**] Counts of the number of identical recombinant haplotypes in wheat-DE8. [**C**] Proportions of unique and non-unique (identical) recombinant haplotypes in wheat-UK8 and wheat-DE8.
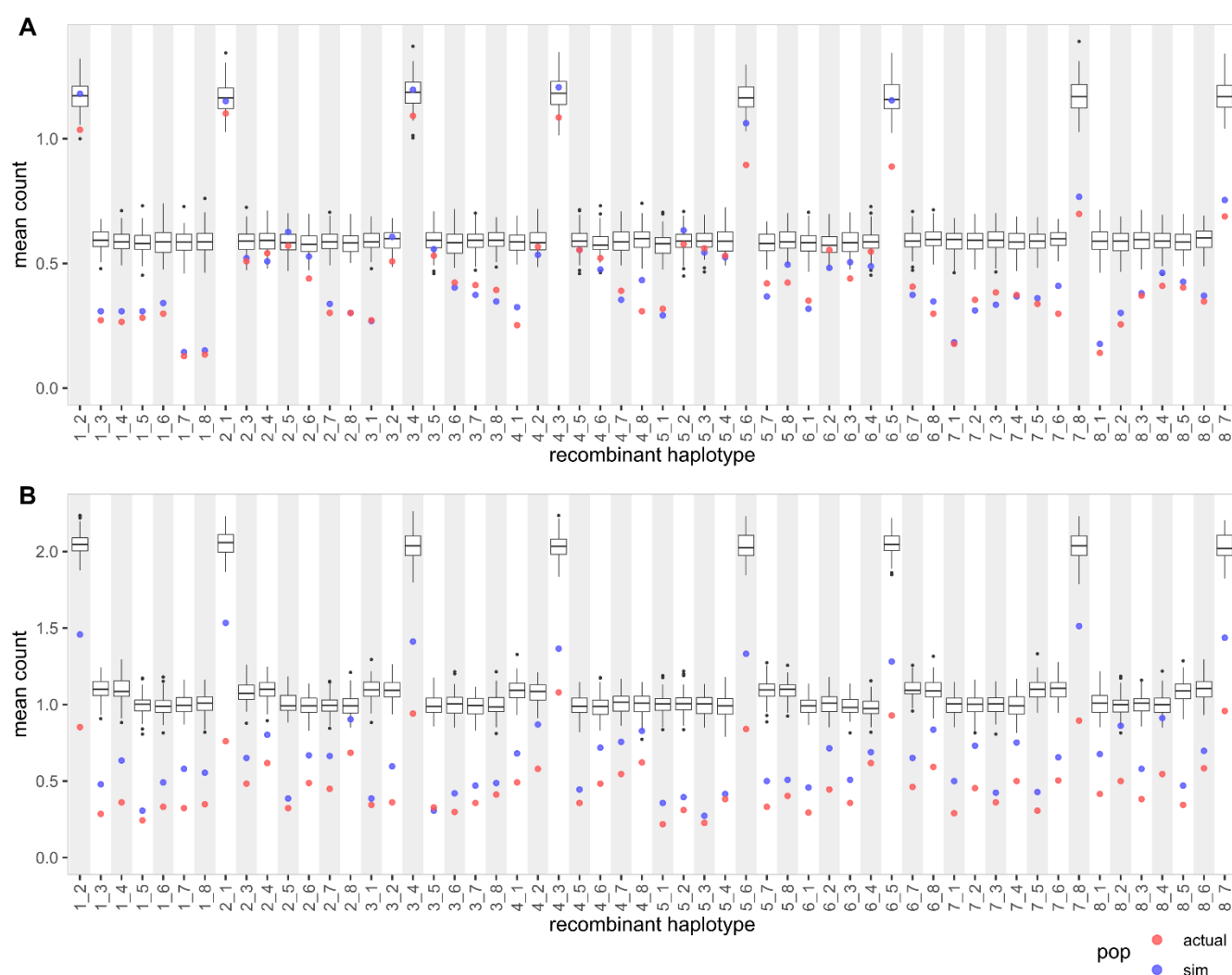
Figure 4. Distributions of recombinant haplotypes in cowpea and tomato MAGIC populations.

[**A**] Plot shows mean count of each recombinant haplotype in a single RIL in cowpea. The boxplot shows mean count from true founder genotypes (100 simulated iterations). The red and blue points show mean count from inferred founder genotypes. [**B**] Plot shows mean count of each recombinant haplotype in a single RIL in tomato.
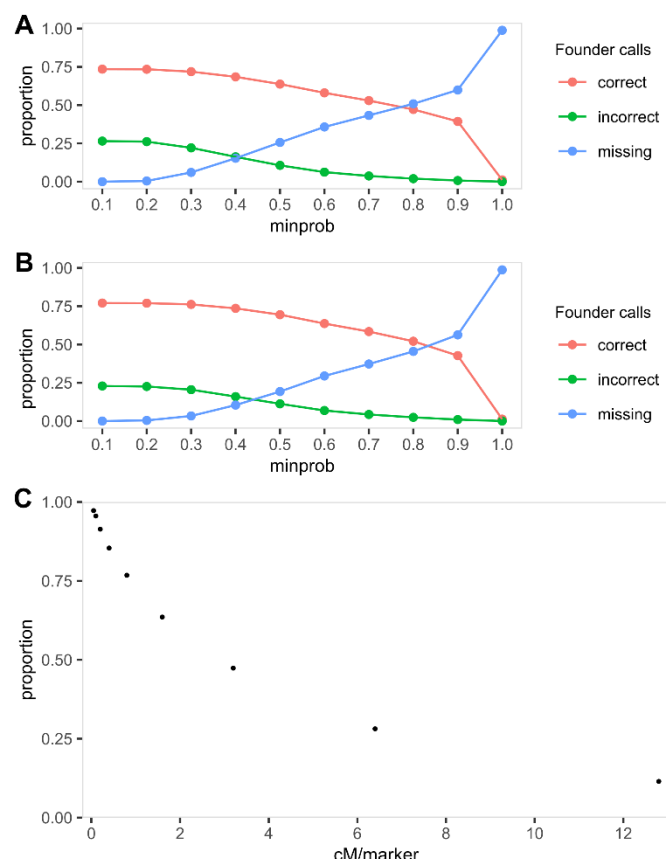
Figure 5. Ideal threshold for inferring founder genotypes and marker density in MAGIC population.

[**A**] Proportions of correct, incorrect and missing founder genotypes inferred at different minimum probability (minprob) in simulated wheat-UK8 population. [**B**] Proportions of correct, incorrect and missing founder genotypes inferred at different minprob in simulated wheat-DE8 population. [**C**] Proportions of recombinant haplotypes recovered (PRHR) at different marker density along a simulated chromosome of 200 cM. The marker density is adjusted by having markers equally spaced at 0.05, 0.10, 0.20, 0.40, 0.80, 1.60, 3.20, 6.40 and 12.80 cM apart.
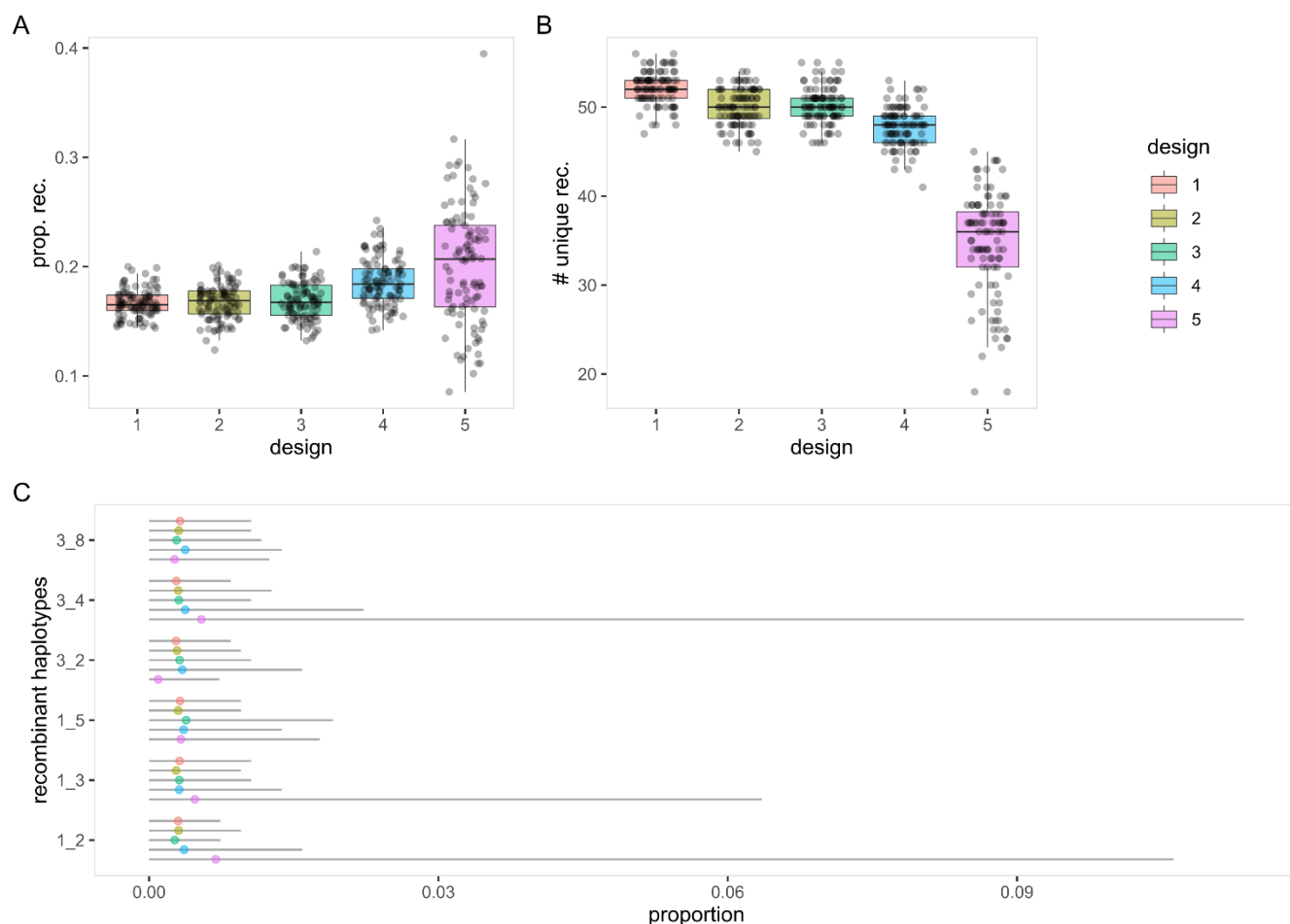
823

Figure 6. Distributions of recombinant haplotypes in five MAGIC population designs.

Recombinant haplotypes are evaluated within a 5 cM interval over 100 iterations of simulation. [**A**] Proportions of total recombinant haplotypes. [**B**] Number of unique recombinant haplotypes. [**C**] Proportions of six chosen recombinant haplotypes. Complete results are available in Table S5.
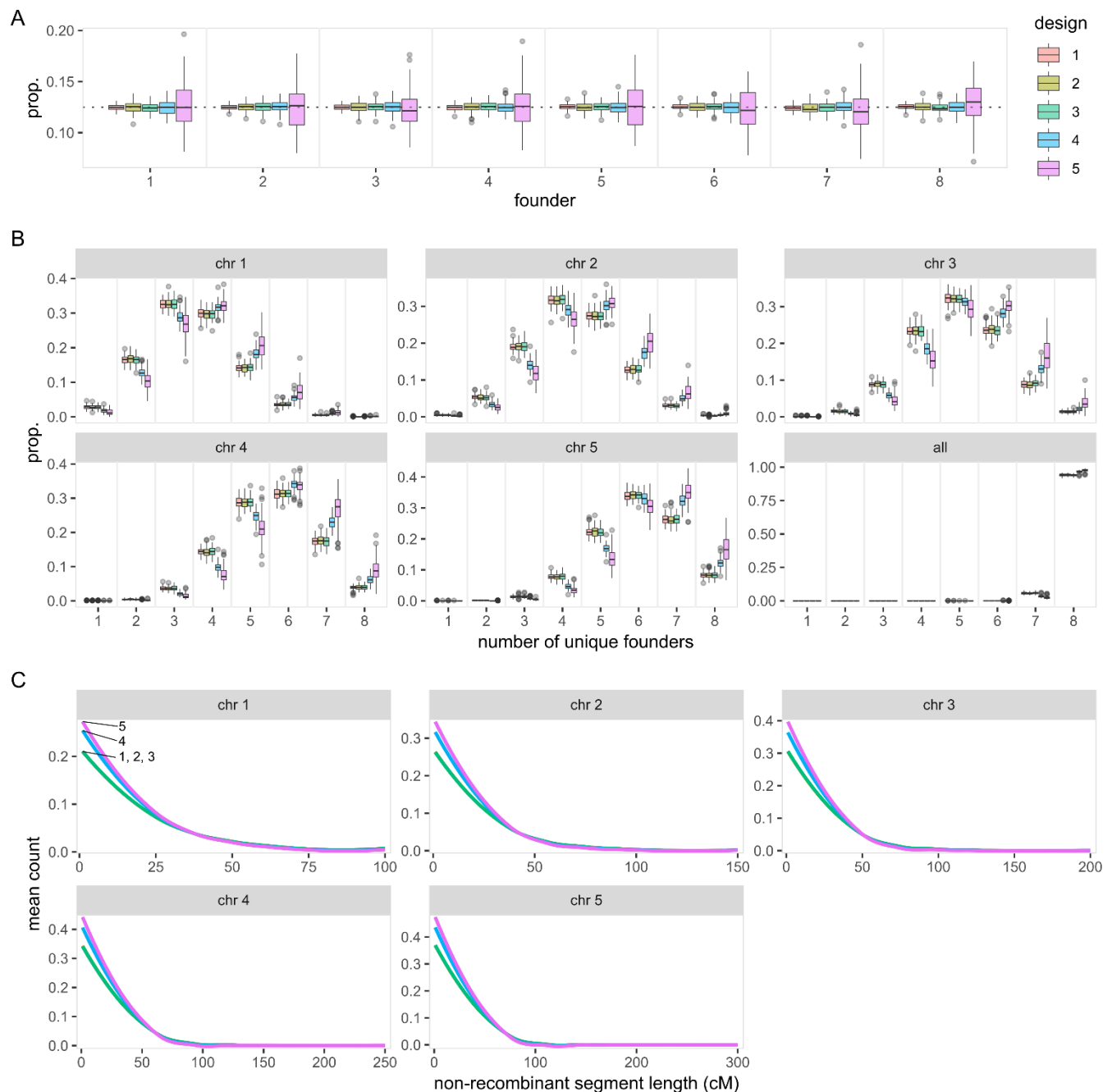
43

Figure 7. Distributions of founder genomes in five MAGIC population designs.

Founder genomes are evaluated from 100 iterations of simulation. [**A**] Proportions of each founder genome in the MAGIC RILs. [**B**] Proportions of the MAGIC RILs carrying tracts of 1 to 8 unique founder genomes in each chromosome. [**C**] Mean count of non-recombinant segment length in each RIL's chromosome.

1 **Supplementary Materials**

2 for

3 **Crafting for a better MAGIC: systematic design and test for multiparental advanced**
4 **generation inter-cross population.**

5 Chin Jian Yang

6 Rodney N. Edmondson

7 Hans-Peter Piepho

8 Wayne Powell

9 Ian Mackay

10

11

Supplementary Text

Source datasets

Links to source datasets without DOI have been archived at https://web.archive.org on April 13, 2021.

- wheat-UK8: https://www.niab.com/research/agricultural-crop-research/resources/niab-magic-population-resources
- wheat-DE8: https://doi.org/10.1186/s12864-018-4915-3 (Table S1)
- cowpea: https://doi.org/10.1111/tpj.13827 (Data S1)
- tomato: https://doi.org/10.1111/pbi.12282 (Table S1 and S2)
- wheat-UK16: http://mtweb.cs.ucl.ac.uk/mus/www/MAGICdiverse/

Supplementary Tables

Table S1. Number of funnels in a partial balanced set or full design.

In P2 design, a balanced design is defined where all founders are present equally in the funnels and all founder pairings are present equally at each level of crosses. The number of funnels required for a partial balanced set is $n-1$ and for full design is $n!/2^{n-1}$. In NP2 design, a balanced design is less strictly defined where only all founders are present equally in the funnels, without the latter requirement in P2. The number of funnels required for a partial balanced set is x where x must be the smallest positive integer that satisfies $(x \cdot n_0)/n \ mod \ 1 = 0$ and n_0 is the next highest P2 n. Number of funnels in a full design is provided up to $n = 8$ as the number of funnels in a full design for higher n is impractical.

n	set	full	n	set	full	n	set	full	n	set	full
3	3	3	35	35	NA	67	67	NA	99	99	NA
4	3	3	36	9	NA	68	17	NA	100	25	NA
5	5	240	37	37	NA	69	69	NA	101	101	NA
6	3	855	38	19	NA	70	35	NA	102	51	NA
7	7	945	39	39	NA	71	71	NA	103	103	NA
8	7	315	40	5	NA	72	9	NA	104	13	NA
9	9	NA	41	41	NA	73	73	NA	105	105	NA
10	5	NA	42	21	NA	74	37	NA	106	53	NA
11	11	NA	43	43	NA	75	75	NA	107	107	NA
12	3	NA	44	11	NA	76	19	NA	108	27	NA
13	13	NA	45	45	NA	77	77	NA	109	109	NA
14	7	NA	46	23	NA	78	39	NA	110	55	NA
15	15	NA	47	47	NA	79	79	NA	111	111	NA
16	15	NA	48	3	NA	80	5	NA	112	7	NA
17	17	NA	49	49	NA	81	81	NA	113	113	NA
18	9	NA	50	25	NA	82	41	NA	114	57	NA
19	19	NA	51	51	NA	83	83	NA	115	115	NA
20	5	NA	52	13	NA	84	21	NA	116	29	NA
21	21	NA	53	53	NA	85	85	NA	117	117	NA
22	11	NA	54	27	NA	86	43	NA	118	59	NA
23	23	NA	55	55	NA	87	87	NA	119	119	NA
24	3	NA	56	7	NA	88	11	NA	120	15	NA
25	25	NA	57	57	NA	89	89	NA	121	121	NA
26	13	NA	58	29	NA	90	45	NA	122	61	NA
27	27	NA	59	59	NA	91	91	NA	123	123	NA
28	7	NA	60	15	NA	92	23	NA	124	31	NA
29	29	NA	61	61	NA	93	93	NA	125	125	NA
30	15	NA	62	31	NA	94	47	NA	126	63	NA
31	31	NA	63	63	NA	95	95	NA	127	127	NA
32	31	NA	64	63	NA	96	3	NA	128	127	NA
33	33	NA	65	65	NA	97	97	NA			
34	17	NA	66	33	NA	98	49	NA			

33 **Table S2. Published MAGIC populations.**

34 Most, if not all, of the MAGIC populations that have been described in either published literature or pre-
35 prints are listed as of March 17, 2021.

Species	n	Design	Pop.	Data Avail.	References
Arabidopsis thaliana	19	Semi-structured	703	Yes	Gnan *et al.* (2014)
	8	Semi-structured	532	No	Huang *et al.* (2011)
Brassica napus (Rapeseed)	8	Basic	680	No	Zhao *et al.* (2017)
Brassica juncea (Chinese mustard)	8	Basic	408	No	Yan *et al.* (2020)
Gossypium hirsutum (Cotton)	11	Unstructured	550	No	Islam *et al.* (2016)
	12	Semi-structured	258	No	Li *et al.* (2016)
	8	Basic	960	No	Huang *et al.* (2018)
Fragaria x ananassa (Strawberry)	6	Unstructured	338	No	Wada *et al.* (2017)
Glycine max (Soybean)	8	Partial, 3 funnels	764	No	Shivakumar *et al.* (2018)
Hordeum vulgare (Barley)	8	Basic	533	No	Sannemann *et al.* (2015)
	32	Basic	324	No	Bülow *et al.* (2019)
	8	Basic	122	No	Novakazi *et al.* (2020)
	8	Basic	29	No	Novakazi *et al.* (2020)
	8	Basic	81	No	Novakazi *et al.* (2020)
	8	Basic	303	No	Novakazi *et al.* (2020)
Oryza sativa ssp. *indica* (Rice)	8	Basic	245	No	Li et al. (2013)
	4	Basic	532	No	Meng *et al.* (2016)
	4	Basic	271	No	Meng *et al.* (2016)
	8	Basic	268	No	Meng *et al.* (2016)
	8	Partial, 35 funnels	1316	Yes	Raghavan *et al.* (2017)
	8	Partial, 35 funnels, X1	144	No	Descalsota *et al.* (2018)
Oryza sativa ssp. *japonica* (Rice)	8	Partial, 35 funnels	500	No	Bandillo *et al.* (2013)
Oryza sativa ssp. *indica x japonica* (Rice)	12	Semi-structured	206	No	Li et al. (2014)
	8	Partial, 2 funnels	981	No	Ogawa *et al.* (2018)
	16	Semi-structured	1027	No	Zaw *et al.* (2019)
	4	Basic	247	No	Han *et al.* (2020)

36 *Table continued on the next page.*

4

37

Table continued from the previous page.

Species	n	Design	Pop.	Data Avail.	References
Solanum lycopersicum (Tomato)	8	Basic	397	Yes	Pascual *et al.* (2015)
	8	Basic	400	No	Campanelli *et al.* (2019)
Sorghum bicolor (Sorghum)	29	Unstructured	200	Yes	Ongom and Ejeta (2018)
Triticum aestivum (Bread wheat)	4	Basic	1579	No	Huang *et al.* (2012)
	8	Partial, 210 funnels	643	Yes	Mackay *et al.* (2014)
	60	Unstructured	1000	No	Thépot *et al.* (2015)
	8	Basic, X1	394	Yes	Stadlmeier *et al.* (2018)
	8	Basic	910	Yes	Sannemann *et al.* (2018)
	8	Partial, 311 funnels	2381	Yes	Shah *et al.* (2019)
	8	Partial, 313 funnels, X2	286	Yes	Shah *et al.* (2019)
	8	Partial, 313 funnels, X3	745	Yes	Shah *et al.* (2019)
	16	Partial, 15 funnels	504	Yes	Scott *et al.* (2021)
Triticum durum (Durum wheat)	4	Basic	338	No	Milner *et al.* (2016)
Vicia faba (Faba bean)	11	Unstructured	188	No	Sallam and Martsch (2015)
	4	Basic	1200	No	Khazaei *et al.* (2018)
Vigna unguiculata (Cowpea)	8	Basic	305	Yes	Huynh *et al.* (2018)
Zea mays (Maize)	9	Semi-structured	529	Yes	Dell'Acqua *et al.* (2015)
	4	Basic	120	No	Mahan *et al.* (2018)
	4	Basic, X1	234	No	Mahan *et al.* (2018)
	4	Basic, X2	106	No	Mahan *et al.* (2018)
	4	Basic, X3	545	No	Mahan *et al.* (2018)
	8	Basic, X6	672	No	Jiménez-Galindo *et al.* (2019)

38   Note: The founder marker data in the bread wheat MAGIC population from Stadlmeier et al. (2018) is not publicly
39   available. Since the rice founders in Li et al. (2014) are unspecified, this MAGIC population is placed under *indica* x
40   *japonica* group.

41

5

Table S3. Number of informative recombinations in wheat-UK8 (full dataset).

The number of informative recombinations (NR) is calculated for both simulated and actual wheat-UK8 dataset. Note: recombinant inbred line (RIL), Morgan (M).

Chr	NR/RIL		NR/RIL/M	
	sim	actual	sim	actual
1A	8.42	3.37	3.53	1.41
1B	12.94	4.64	3.69	1.32
1D	4.44	1.78	3.38	1.36
2A	9.36	5.18	3.61	2.00
2B	13.64	3.58	3.59	0.94
2D	6.50	2.97	3.27	1.49
3A	11.13	5.40	3.60	1.75
3B	10.45	4.29	3.69	1.51
3D	5.59	2.03	2.87	1.05
4A	7.94	3.32	3.64	1.52
4B	8.43	3.03	3.63	1.30
4D	3.95	1.36	3.14	1.08
5A	11.07	5.42	3.53	1.73
5B	11.38	3.23	3.65	1.04
5D	6.42	2.58	3.17	1.28
6A	10.42	4.23	3.66	1.49
6B	9.57	2.67	3.64	1.02
6D	5.44	1.03	2.53	0.48
7A	14.24	5.19	3.69	1.35
7B	10.19	3.78	3.55	1.32
7D	6.35	2.28	2.91	1.05
All	187.88	71.39	3.48	1.32

47

Table S4. Number of informative recombinations in wheat-DE8 (full dataset).

The number of informative recombinations (NR) is calculated for both simulated and actual wheat-DE8a dataset. Note: recombinant inbred line (RIL), Morgan (M).

Chr	NR/RIL		NR/RIL/M	
	sim	actual	sim	actual
1A	5.03	3.26	3.60	2.34
1B	5.96	5.11	3.54	3.04
1D	4.80	2.45	2.73	1.39
2A	6.07	5.46	3.36	3.03
2B	5.96	4.84	3.53	2.87
2D	4.22	4.14	3.12	3.07
3A	6.00	5.72	3.46	3.30
3B	4.97	5.35	3.58	3.85
3D	4.07	2.54	2.61	1.63
4A	5.36	3.2	3.44	2.06
4B	3.97	2.35	3.46	2.05
4D	3.53	0.81	2.36	0.54
5A	5.06	6.06	3.57	4.28
5B	7.21	4.27	3.33	1.97
5D	4.71	1.88	2.57	1.03
6A	4.93	4.4	3.58	3.20
6B	4.44	3.6	3.63	2.94
6D	4.03	2.73	2.67	1.81
7A	7.31	5.85	3.51	2.81
7B	6.08	3.95	3.43	2.23
7D	5.57	3.31	2.64	1.56
All	109.29	81.27	3.21	2.39

51

52

53

7

54 **Table S5. Proportions of individual recombinant haplotypes in five MAGIC population**
55 **designs.**

56 The mean proportions (standard deviations in parentheses) are shown for each recombinant haplotype
57 and design. Two-ways recombinant haplotypes in design 5 are highlighted.

Rec. hap.	Design 1	Design 2	Design 3	Design 4	Design 5
1_2	0.0030 (0.0019)	0.0031 (0.0020)	0.0027 (0.0018)	0.0036 (0.0036)	0.0069 (0.0206)
1_3	0.0032 (0.0021)	0.0028 (0.0020)	0.0031 (0.0022)	0.0031 (0.0030)	0.0048 (0.0108)
1_4	0.0030 (0.0020)	0.0029 (0.0025)	0.0025 (0.0021)	0.0033 (0.0028)	0.0058 (0.0104)
1_5	0.0032 (0.0019)	0.0030 (0.0021)	0.0038 (0.0038)	0.0036 (0.0034)	0.0033 (0.0036)
1_6	0.0035 (0.0021)	0.0029 (0.0024)	0.0031 (0.0022)	0.0027 (0.0027)	0.0029 (0.0034)
1_7	0.0031 (0.0018)	0.0033 (0.0027)	0.0023 (0.0016)	0.0030 (0.0027)	0.0033 (0.0033)
1_8	0.0029 (0.0019)	0.0034 (0.0028)	0.0030 (0.0024)	0.0032 (0.0033)	0.0033 (0.0036)
2_1	0.0029 (0.0018)	0.0031 (0.0023)	0.0028 (0.0016)	0.0030 (0.0036)	0.0092 (0.0212)
2_3	0.0027 (0.0018)	0.0027 (0.0021)	0.0032 (0.0026)	0.0036 (0.0035)	0.0030 (0.0083)
2_4	0.0030 (0.0019)	0.0029 (0.0025)	0.0026 (0.0018)	0.0030 (0.0028)	0.0022 (0.0057)
2_5	0.0035 (0.0020)	0.0027 (0.0020)	0.0027 (0.0017)	0.0037 (0.0032)	0.0021 (0.0029)
2_6	0.0031 (0.0020)	0.0030 (0.0025)	0.0024 (0.0019)	0.0032 (0.0030)	0.0023 (0.0032)
2_7	0.0030 (0.0019)	0.0029 (0.0025)	0.0040 (0.0036)	0.0031 (0.0031)	0.0018 (0.0023)
2_8	0.0029 (0.0017)	0.0036 (0.0028)	0.0034 (0.0026)	0.0038 (0.0035)	0.0022 (0.0024)
3_1	0.0027 (0.0018)	0.0027 (0.0022)	0.0030 (0.0021)	0.0028 (0.0026)	0.0022 (0.0056)
3_2	0.0028 (0.0018)	0.0029 (0.0021)	0.0032 (0.0023)	0.0034 (0.0034)	0.0009 (0.0015)
3_4	0.0028 (0.0019)	0.0030 (0.0024)	0.0031 (0.0024)	0.0037 (0.0040)	0.0054 (0.0170)
3_5	0.0031 (0.0020)	0.0032 (0.0024)	0.0033 (0.0024)	0.0033 (0.0028)	0.0023 (0.0028)
3_6	0.0032 (0.0019)	0.0028 (0.0025)	0.0033 (0.0030)	0.0036 (0.0031)	0.0025 (0.0037)
3_7	0.0031 (0.0020)	0.0031 (0.0024)	0.0030 (0.0018)	0.0036 (0.0035)	0.0023 (0.0035)
3_8	0.0032 (0.0021)	0.0031 (0.0024)	0.0029 (0.0026)	0.0037 (0.0032)	0.0026 (0.0032)
4_1	0.0029 (0.0017)	0.0027 (0.0022)	0.0032 (0.0028)	0.0032 (0.0032)	0.0037 (0.0103)
4_2	0.0029 (0.0021)	0.0034 (0.0027)	0.0024 (0.0014)	0.0029 (0.0027)	0.0049 (0.0109)
4_3	0.0027 (0.0019)	0.0032 (0.0027)	0.0032 (0.0024)	0.0032 (0.0030)	0.0076 (0.0190)
4_5	0.0030 (0.0016)	0.0029 (0.0023)	0.0028 (0.0024)	0.0034 (0.0033)	0.0028 (0.0030)
4_6	0.0029 (0.0017)	0.0027 (0.0021)	0.0034 (0.0033)	0.0033 (0.0034)	0.0028 (0.0031)
4_7	0.0028 (0.0017)	0.0032 (0.0025)	0.0029 (0.0024)	0.0032 (0.0033)	0.0023 (0.0028)
4_8	0.0031 (0.0019)	0.0028 (0.0026)	0.0035 (0.0026)	0.0034 (0.0036)	0.0032 (0.0035)
5_1	0.0030 (0.0018)	0.0030 (0.0023)	0.0034 (0.0030)	0.0028 (0.0026)	0.0033 (0.0036)
5_2	0.0029 (0.0017)	0.0028 (0.0023)	0.0026 (0.0017)	0.0036 (0.0029)	0.0023 (0.0029)
5_3	0.0030 (0.0018)	0.0029 (0.0024)	0.0029 (0.0025)	0.0036 (0.0037)	0.0027 (0.0030)
5_4	0.0029 (0.0015)	0.0026 (0.0020)	0.0029 (0.0025)	0.0030 (0.0033)	0.0028 (0.0032)
5_6	0.0027 (0.0018)	0.0030 (0.0023)	0.0027 (0.0016)	0.0033 (0.0031)	0.0074 (0.0211)
5_7	0.0028 (0.0018)	0.0034 (0.0029)	0.0027 (0.0016)	0.0034 (0.0030)	0.0045 (0.0112)
5_8	0.0028 (0.0017)	0.0026 (0.0021)	0.0032 (0.0028)	0.0029 (0.0024)	0.0057 (0.0110)
6_1	0.0032 (0.0019)	0.0031 (0.0022)	0.0030 (0.0023)	0.0035 (0.0030)	0.0029 (0.0031)
6_2	0.0031 (0.0020)	0.0031 (0.0025)	0.0024 (0.0018)	0.0028 (0.0029)	0.0021 (0.0028)
6_3	0.0026 (0.0018)	0.0028 (0.0024)	0.0038 (0.0029)	0.0031 (0.0028)	0.0026 (0.0032)
6_4	0.0030 (0.0019)	0.0029 (0.0024)	0.0034 (0.0032)	0.0030 (0.0024)	0.0029 (0.0031)
6_5	0.0033 (0.0016)	0.0032 (0.0022)	0.0026 (0.0019)	0.0032 (0.0029)	0.0088 (0.0238)
6_7	0.0029 (0.0018)	0.0031 (0.0025)	0.0028 (0.0023)	0.0037 (0.0033)	0.0025 (0.0063)
6_8	0.0031 (0.0020)	0.0028 (0.0027)	0.0028 (0.0022)	0.0030 (0.0028)	0.0043 (0.0102)
7_1	0.0026 (0.0017)	0.0034 (0.0026)	0.0024 (0.0016)	0.0030 (0.0028)	0.0029 (0.0028)
7_2	0.0029 (0.0018)	0.0027 (0.0025)	0.0048 (0.0043)	0.0035 (0.0031)	0.0018 (0.0024)
7_3	0.0034 (0.0023)	0.0029 (0.0024)	0.0030 (0.0019)	0.0034 (0.0028)	0.0023 (0.0028)
7_4	0.0029 (0.0017)	0.0031 (0.0024)	0.0028 (0.0021)	0.0030 (0.0031)	0.0023 (0.0030)
7_5	0.0031 (0.0018)	0.0033 (0.0024)	0.0029 (0.0017)	0.0030 (0.0029)	0.0039 (0.0094)
7_6	0.0030 (0.0020)	0.0029 (0.0026)	0.0027 (0.0022)	0.0037 (0.0035)	0.0030 (0.0068)
7_8	0.0028 (0.0020)	0.0025 (0.0021)	0.0025 (0.0018)	0.0033 (0.0029)	0.0059 (0.0182)
8_1	0.0029 (0.0019)	0.0030 (0.0026)	0.0034 (0.0026)	0.0037 (0.0035)	0.0032 (0.0037)
8_2	0.0027 (0.0018)	0.0027 (0.0021)	0.0031 (0.0025)	0.0036 (0.0029)	0.0023 (0.0030)
8_3	0.0030 (0.0018)	0.0035 (0.0027)	0.0032 (0.0024)	0.0034 (0.0033)	0.0025 (0.0027)
8_4	0.0028 (0.0019)	0.0027 (0.0023)	0.0031 (0.0025)	0.0041 (0.0042)	0.0031 (0.0035)
8_5	0.0031 (0.0022)	0.0032 (0.0027)	0.0033 (0.0028)	0.0036 (0.0035)	0.0047 (0.0111)
8_6	0.0030 (0.0021)	0.0032 (0.0027)	0.0028 (0.0024)	0.0035 (0.0032)	0.0043 (0.0117)
8_7	0.0030 (0.0020)	0.0030 (0.0024)	0.0024 (0.0015)	0.0033 (0.0038)	0.0063 (0.0180)

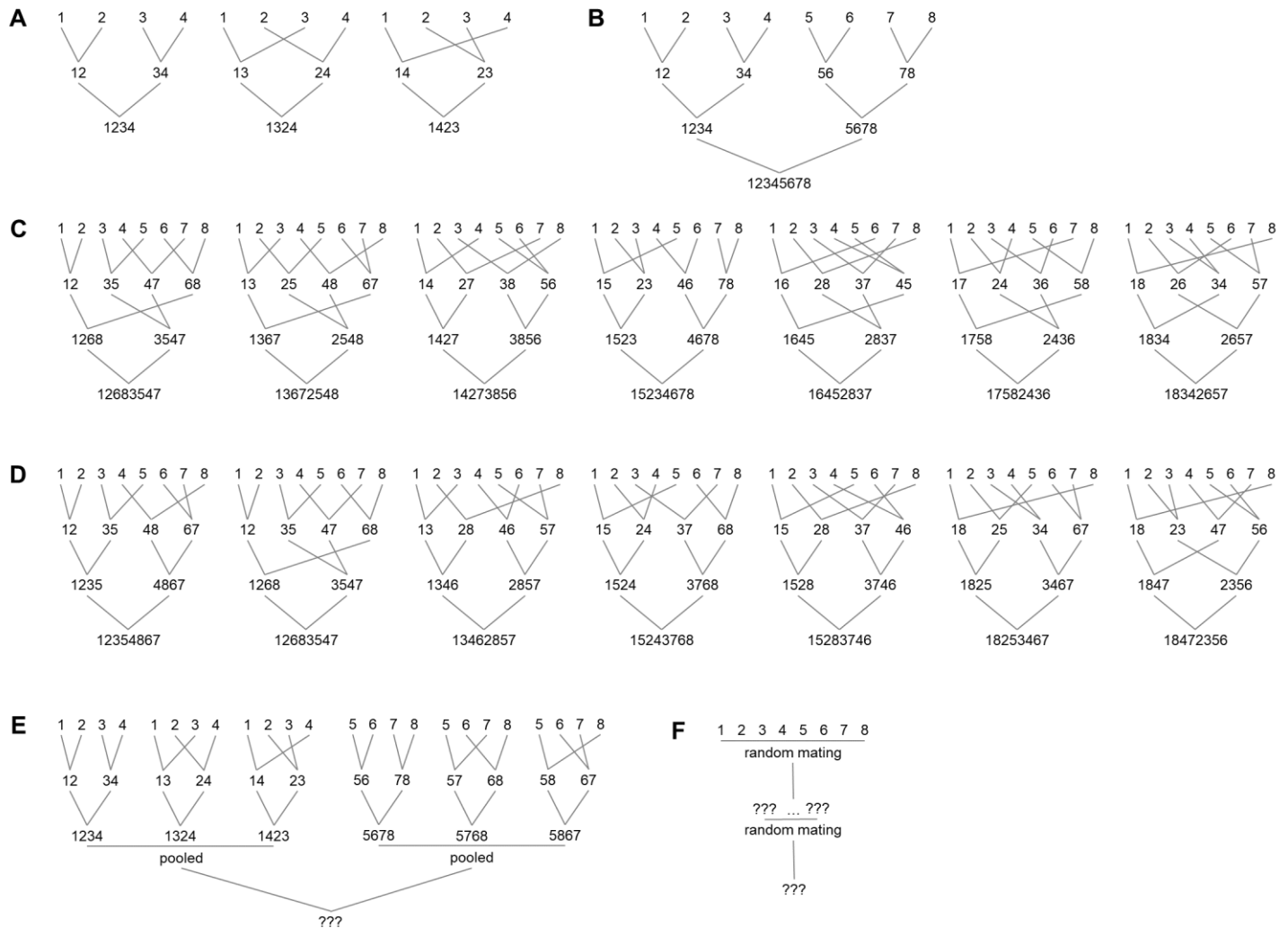## Supplementary Figures

Figure S1. Examples of MAGIC population designs.

[**A**] Full design with 4 founders. [**B**] Basic design with 8 founders. [**C**] Partial balanced design with 7 funnels and 8 founders. [**D**] Partial unbalanced design with 7 funnels and 8 founders. [**E**] Semi-structured design with 8 founders. [**F**] Unstructured design with 8 founders.
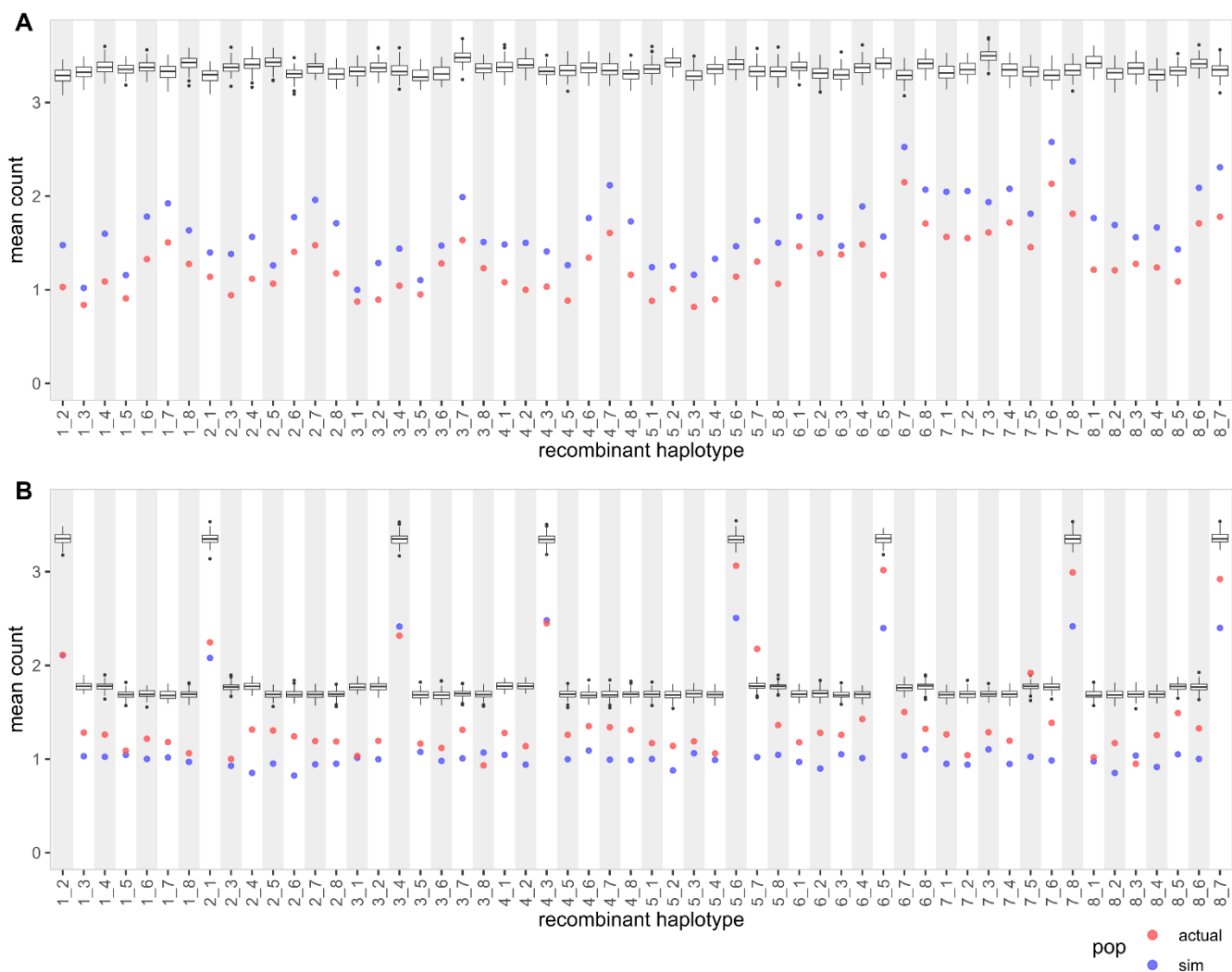
9

Figure S2. Distributions of recombinant haplotypes in two wheat MAGIC populations (full datasets).

[**A**] Plot shows mean count of each recombinant haplotype in a single RIL in wheat-UK8 (full dataset with 643 RILs and 18,599 markers). The boxplot shows mean count from true founder genotypes (100 simulated iterations). The red and blue points show mean count from inferred founder genotypes. [**B**] Plot shows mean count of each recombinant haplotype in a single RIL in wheat-DE8 (full dataset with 910 RILs and 7,579 markers).

91
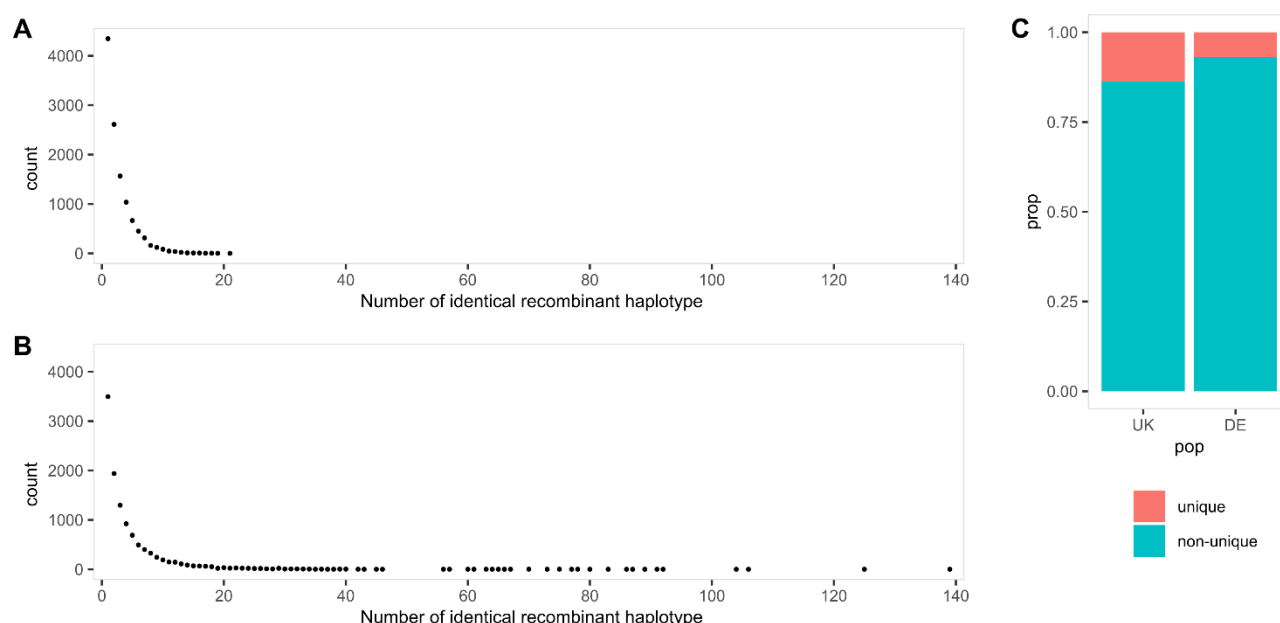
Figure S3. Distributions of unique and identical recombinant haplotypes in two wheat MAGIC populations.

Recombinant haplotypes are considered identical if they are of the same founder pairs and present in the same 10 cM interval, otherwise unique. [**A**] Counts of the number of identical recombinant haplotypes in wheat-UK8. The left most point is the count of unique recombinant haplotypes. [**B**] Counts of the number of identical recombinant haplotypes in wheat-DE8. [**C**] Proportions of unique and non-unique (identical) recombinant haplotypes in wheat-UK8 and wheat-DE8.

100
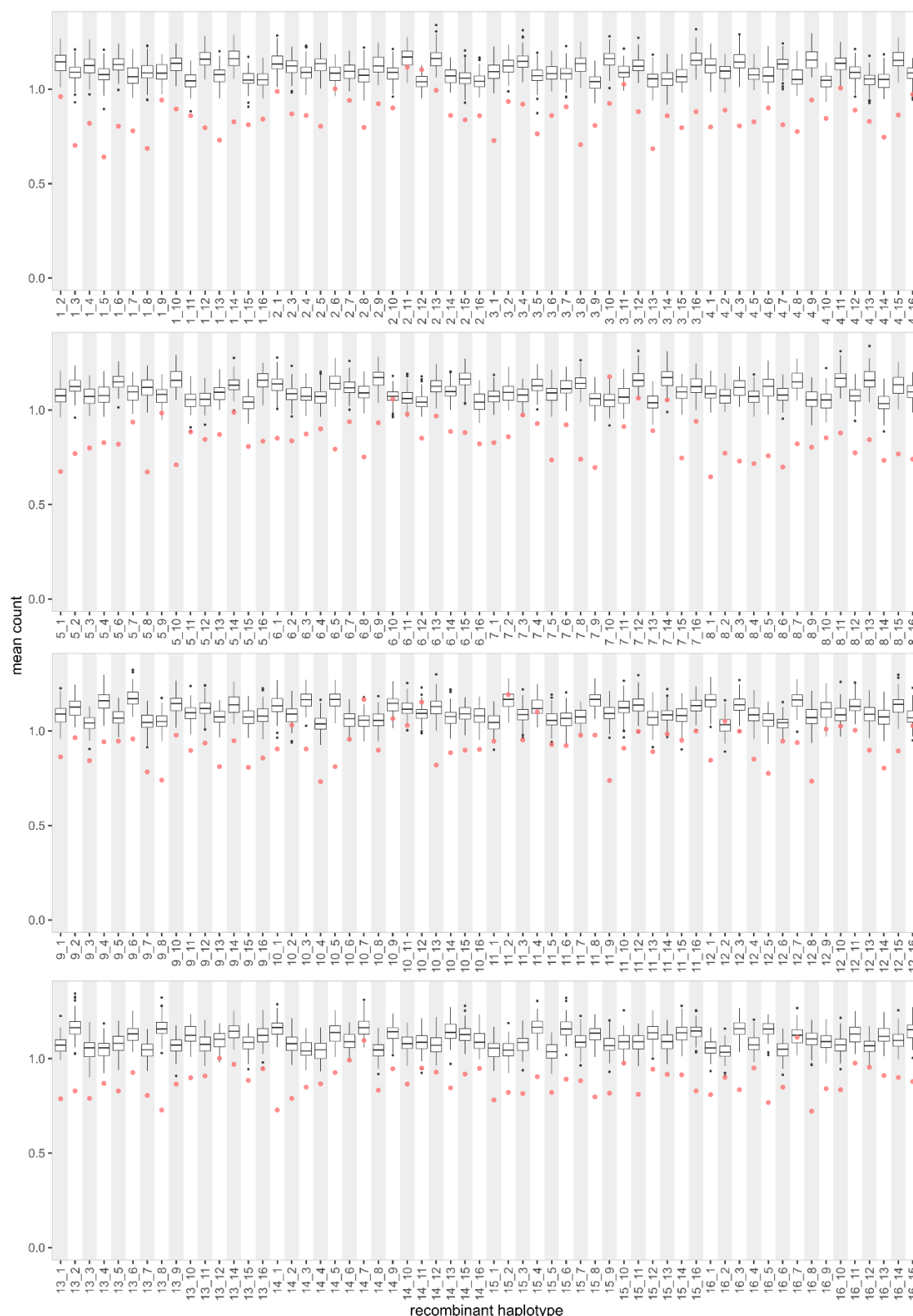
Figure S4. Distributions of recombinant haplotypes in wheat-UK16.

Plot shows mean count of each recombinant haplotype in a single RIL in wheat-UK16. The boxplot shows mean count from true founder genotypes (100 simulated iterations). The red points show mean count from inferred founder genotypes in actual dataset.

Literature Cited

Bandillo, N., C. Raghavan, P. A. Muyco, M. A. L. Sevilla, I. T. Lobina *et al.*, 2013 Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice 6: 11.

Bülow, L., M. Nightingale and L. Frese, 2019 A MAGIC population as an approach to the maintenance and development of breeding usable genetic diversity of winter barley through on-farm management. J. Kulturpflanzen 71: 286-298.

Campanelli, G., S. Sestili, N. Acciarri, F. Montemurro, D. Palma *et al.*, 2019 Multi-parental advanced generation inter-cross population, to develop organic tomato genotypes by participatory plant breeding. Agronomy 9: 119.

Dell'Acqua, M., D. M. Gatti, G. Pea, F. Cattonaro, F. Coppens *et al.*, 2015 Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in *Zea mays*. Genome Biol. 16: 167.

Descalsota, G. I. L., B. P. M. Swamy, H. Zaw, M. A. Inabangan-Asilo, A. Amparado *et al.*, 2018 Genome-wide association mapping in a rice MAGIC Plus population detects QTLs and genes useful for biofortification. Front. Plant Sci. 9: 1347.

Gnan, S., A. Priest and P. X. Kover, 2014 The genetic basis of natural variation in seed size and seed number and their trade-off using *Arabidopsis thaliana* MAGIC lines. Genetics 198: 1751-1758.

Han, Z., G. Hu, H. Liu, F. Liang, L. Yang *et al.*, 2020 Bin-based genome-wide association analyses improve power and resolution in QTL mapping and identify favorable alleles from multiple parents in a four-way MAGIC rice population. Theor. Appl. Genet. 133: 59-71.

Huang, B. E., A. W. George, K. L. Forrest, A. Kilian, M. J. Hayden *et al.*, 2012 A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotechnol. J. 10: 826-839.

Huang, C., C. Shen, T. Wen, B. Gao, D. Zhu *et al.*, 2018 SSR-based association mapping of fiber quality in upland cotton using an eight-way MAGIC population. Mol. Genet. Genom. 293: 793-805.

Huang, X., M.-J. Paulo, M. Boer, S. Effgen, P. Keizer *et al.*, 2011 Analysis of natural allelic variation in *Arabidopsis* using a multiparent recombinant inbred line population. Proc. Natl. Acad. Sci. USA 108: 4488-4493.

Huynh, B.-L., J. D. Ehlers, B. E. Huang, M. Muñoz-Amatriaín, S. Lonardi *et al.*, 2018 A multi-parent advanced generation inter-cross (MAGIC) population for genetic analysis and improvement of cowpea (*Vigna unguiculata* L. Walp.). Plant J. 93: 1129-1142.

Islam, M. S., G. N. Thyssen, J. N. Jenkins, L. Zeng, C. D. Delhom *et al.*, 2016 A MAGIC population-based genome-wide association study reveals functional association of *GhRBB1_A07* gene with superior fiber quality in cotton. BMC Genomics 17: 903.

143 Jiménez-Galindo, J. C., R. A. Malvar, A. Butrón, R. Santiago, L. F. Samayoa *et al.*, 2019
144 Mapping of resistance to corn borers in a MAGIC population of maize. BMC Plant Biol. 19: 431.

145 Khazaei, H., F. L. Stoddard, R. W. Purves and A. Vandenberg, 2018 A multi-parent faba bean
146 (*Vicia faba* L.) population for future genomic studies. Plant Genet. Res. 16: 419-423.

147 Li, D. G., Z. X. Li, J. S. Hu, Z. X. Lin and X. F. Li, 2016 Polymorphism analysis of multi-parent
148 advanced generation inter-cross (MAGIC) populations of upland cotton developed in China.
149 Genet. Mol. Res. 15: gmr15048759.

150 Li, X.-F., Z.-X. Liu, D.-B. Lu, Y.-Z. Liu, X.-X. Mao *et al.*, 2013 Development and evaluation of
151 multi-genotype varieties of rice derived from MAGIC lines. Euphytica 192: 77-86.

152 Li, Z., G. Ye, M. Yang, Z. Liu, D. Lu *et al.*, 2014 Genetic characterization of a multiparent
153 recombinant inbred line of rice population. Res. Crops 15: 28-37.

154 Mackay, I. J., P. Bansept-Basler, T. Barber, A. R. Bentley, J. Cockram *et al.*, 2014 An eight-
155 parent multiparent advanced generation inter-cross population for winter-sown wheat: creation,
156 properties, and validation. G3: Genes, Genomes, Genetics 4: 1603-1610.

157 Mahan, A. L., S. C. Murray and P. E. Klein, 2018 Four-parent maize (FPM) population:
158 development and phenotypic characterization. Crop Sci. 58: 1106-1117.

159 Meng, L., L. Guo, K. Ponce, X. Zhao and G. Ye, 2016 Characterization of three *indica* rice
160 multiparent advanced generation intercross (MAGIC) populations for quantitative trait loci
161 identification. Plant Genome 9: 1-14.

162 Milner, S. G., M. Maccaferri, B. E. Huang, P. Mantovani, A. Massi *et al.*, 2015 A multiparental
163 cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp.
164 *durum*). Plant Biotechnol. J. 14: 735-748.

165 Novakazi, F., L. Krusell, J. D. Jensen, J. Orabi, A. Jahoor *et al.*, 2020 You had me at "MAGIC"!:
166 four barley MAGIC populations reveal novel resistance QTL for powdery mildew. Genes 11:
167 1512.

168 Ogawa, D., E. Yamamoto, T. Ohtani, N. Kanno, H. Tsunematsu *et al.*, 2018 Haplotype-based
169 allele mining in the Japan-MAGIC rice population. Sci. Rep. 8: 4379.

170 Ongom, P. O. and G. Ejeta, 2018 Mating design and genetic structure of a multi-parent
171 advanced generation intercross (MAGIC) population of Sorghum (*Sorghum bicolor* (L.)
172 Moench). G3: Genes, Genomes, Genetics 8: 331-341.

173 Pascual, L., N. Desplat, B. E. Huang, A. Desgroux, L. Bruguier *et al.*, 2015 Potential of a tomato
174 MAGIC population to decipher the genetic control of quantitative traits and detect causal
175 variants in the resequencing era. Plant Biotechnol. J. 13: 565-577.

176 Raghavan, C., R. Mauleon, V. Lacorte, M. Jubay, H. Zaw *et al.*, 2017 Approaches in
177 characterizing genetic structure and mapping in a rice multiparental population. G3: Genes,
178 Genomes, Genetics 7: 1721-1730.

179 Sallam, A. and R. Martsch, 2015 Association mapping for frost tolerance using multi-parent
180 advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). Genetica
181 143: 501-514.

182 Sannemann, W., A. Lisker, A. Maurer, J. Léon, E. Kazman *et al.*, 2018 Adaptive selection of
183 founder segments and epistatic control of plant height in the MAGIC winter wheat population
184 WM-800. BMC Genomics 19: 559.

185 Sannemann, W., B. E. Huang, B. Mathew and J. Léon, 2015 Multi-parent advanced generation
186 inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a
187 proof of concept. Mol. Breed. 35: 86.

188 Scott, M. F., N. Fradgley, A. R. Bentley, T. Brabbs, F. Corke *et al.*, 2021 Limited haplotype
189 diversity underlies polygenic trait architecture across 70 years of wheat breeding. Genome Biol.
190 in press. https://www.biorxiv.org/content/10.1101/2020.09.15.296533v1.

191 Shah, R., B. E. Huang, A. Whan, M. Newberry, K. Verbyla *et al.*, 2019 The complex genetic
192 architecture of recombination and structural variation in wheat uncovered using a large 8-
193 founder MAGIC population. bioRxiv. doi: 10.1101/594317 (Preprint posted March 31, 2019).

194 Shivakumar, M., G. Kumawat, C. Gireesh, S. V. Ramesh and S. M. Husain, 2018 Soybean
195 MAGIC population: a novel resource for genetics and plant breeding. Curr. Sci. 114: 906-908.

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

199 Thépot, S., G. Restoux, I. Goldringer, F. Hospital, D. Gouache *et al.*, 2015 Efficient tracking
200 selection in a multiparental population: the case of earliness in wheat. Genetics 199: 609-623.

201 Wada, T., K. Oku, S. Nagano, S. Isobe, H. Suzuki *et al.*, 2017 Development and
202 characterization of a strawberry MAGIC population derived from crosses with six strawberry
203 cultivars. Breed. Sci. 67: 370-381.

204 Yan, W., H. Zhao, K. Yu, T. Wang, A. N. Khattak *et al.*, 2020 Development of a multiparent
205 advanced generation intercross (MAGIC) population for genetic exploitation of complex traits in
206 *Brassica juncea*: glucosinolate content as an example. Plant Breed. 139: 779-789.

207 Zaw, H., C. Raghavan, A. Pocsedio, B. P. M. Swamy, M. L. Jubay *et al.*, 2019 Exploring genetic
208 architecture of grain yield and quality traits in a 16-way *indica* by *japonica* rice MAGIC global
209 population. Sci. Rep. 9: 19605.

210 Zhao, F.-Y., H. Zhao, X.-L. Wang and X.-F. Li, 2017 Construction and application potential of
211 MAGIC population on genetic breeding of rapeseed (*Brassica napus* L.). Chinese J. Oil Crop
212 Sci. 39: 149-151.