

1 **A generalized statistical framework to assess mixing ability from incomplete mixing designs**
2 **using binary or higher order variety mixtures and application to wheat**

3

4 Authors and affiliations:

5 Emma Forst^{a,1}, Jérôme Enjalbert^a, Vincent Allard^b, Christophe Ambroise^{c,d}, Inès Krissaane^d,
6 Tristan Mary-Huard^{a,d}, Stéphane Robin^d, Isabelle Goldringer^a

7 ^a GQE- Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay,
8 91190, Gif-sur-Yvette, France

9 ^b GDEC, UMR 1095, INRA, UCA, 63000 Clermont-Ferrand, France

10 ^c LaMME, UMR 8071, CNRS, Université d'Evry Val-d'Essonne, Université Paris-Saclay,
11 91037, Evry Cedex, France

12 ^d MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France

13 Emails: emma.forst@inra.fr, jerome.enjalbert@inra.fr, vincent.allard@inra.fr,
14 christophe.ambroise@genopole.cnrs.fr, krissaane.ines@gmail.com, tristan.mary-
15 huard@agroparistech.fr, stephane.robin@agroparistech.fr, isabelle.goldringer@inra.fr

16

17 **HIGHLIGHTS**

- 18 • Statistical analyses for variety mixtures have made little progress in recent years
19 • Novel models are proposed to study mixing ability in incomplete designs
20 • The models account for inter and intra-genotypic interactions within mixtures
21 • The framework handles mixtures with any order and proportions of components
22 • This framework was shown to be relevant on wheat mixture trial analysis

¹ Corresponding author: Génétique Quantitative et Evolution - Le Moulon, INRA, Univ. Paris-Sud, CNRS, AgroParisTech, Université Paris-Saclay, Ferme du Moulon, 91190, Gif-sur-Yvette, France. E-mail: emma.forst@inra.fr

23

24 **ABSTRACT**

25 There has been recently a renewed interest for variety mixtures due to their potential capacity
26 to stabilize production through buffering abiotic and biotic stresses. Part of this results from
27 complementarity and/or compensation between varieties which can be assessed under
28 mixed stands only. Mixing ability of varieties can be partitioned into *General* and *Specific*
29 *Mixing Abilities* (GMA and SMA) that have been estimated so far through the evaluation of
30 binary mixtures in complete diallel designs. However, the number of mixtures increases
31 exponentially with the number of studied varieties, and the only feasible devices are
32 incomplete designs. Despite the long history of statistical analysis of variety mixtures, such
33 incomplete design analysis have rarely been addressed so far. To fill the gap, we proposed a
34 generalized statistical framework to assess mixing abilities based on mixed models and
35 BLUP method, with an original modeling of plant-plant interactions. The approach has been
36 applied to a panel of 25 winter wheat genotypes observed in two contrasted experimental
37 designs: (i) an incomplete diallel of 75 binary mixtures, and (ii) a trial including higher order
38 mixtures (four and eight components). The use of mixing ability models improved prediction
39 accuracy (of modeled values for observed traits) in comparison to predictions from the mean
40 of the pure stand components, especially in the first experiment. Genetic variability was
41 detected for the GMA of yield and its components, whereas variability for SMA was lower.
42 GMA predictions based on the diallel trial were highly correlated with the GMA of the second
43 trial providing accurate inter-trial predictions. A new model has been proposed to jointly
44 account for inter and intra-genotypic interactions for specific mixing ability, thus contributing
45 to a better understanding of mixture functioning. This framework constitutes a step forward to
46 the screening for mixing ability, and could be further integrated into breeding programs for the
47 development of intra- or inter-specific crop mixtures.

48

49 Key-words: intra-specific mixtures, plant-plant interactions, diallel, *Triticum aestivum*, BLUP

51 **1. INTRODUCTION**

52 Crop genetic diversity is a major lever that can be mobilized to face current challenges in
53 agriculture such as the increasing climatic stochasticity and the necessity to reduce the use
54 of synthetic inputs (Newton *et al.* 2009). Among diversifying strategies, cereal variety
55 mixtures have attracted increasing interest in recent decades (Kiær *et al.* 2009, 2012, Borg *et*
56 *al.* 2018, Reiss and Drinkwater 2018). In particular, mixtures of susceptible and resistant
57 cereal varieties have been shown to provide a larger reduction in the development of foliar
58 diseases than expected from the sum of their components when complementary resistances
59 are combined (Wolfe 1985; Finckh and Mundt 1992; Zhu *et al.* 2000; de Vallavieille-Pope
60 2004). Using variety mixtures may also allow buffering abiotic and other biotic stresses,
61 therefore leading to a stabilization of production (Østergård *et al.* 2005; Kaut *et al.* 2009;
62 Creissen *et al.* 2016), through complementarity and compensation mechanisms that might
63 occur between plants of different varieties, as shown in species mixtures (Tilman *et al.* 1997).

64 Despite these potential advantages of mixtures, very little literature provides practical
65 guidelines for designing mixtures, and consequently mixtures used in agriculture are often
66 composed of the varieties showing the best performances in pure stand (Borg *et al.* 2018).
67 However, it has been shown that mixtures performances are not necessarily correlated with
68 the means of their pure stand components: interaction between plants with different
69 genotypes (Finckh and Mundt 1992) can make it difficult to predict the behavior of mixtures.
70 Therefore, it is important to accurately estimate and predict the performances of varieties in
71 mixtures. Screening a set of lines for their *mixing ability* nevertheless presents many
72 methodological challenges, in particular in the case of a high number of genotypes and an
73 exponentially higher number of binary or higher order mixtures, and calls for adapted
74 statistical methods (Dawson and Goldringer 2012; Barot *et al.* 2017) that could be used to
75 detect the best varieties for mixing conditions, and to develop specific breeding schemes for
76 intra-specific mixture design.

77 Mixing ability relies on an analogy with the concept of *combining ability* developed for hybrid
78 breeding (Sprague and Tatum 1942). By assessing a set of p genotypes in all possible hybrid
79 combinations, the combining ability can be estimated and partitioned between the general
80 combining ability of the genotypes and the specific ability of each particular combination
81 (Griffing 1956). The analogy has first been proposed for the study of plant interactions
82 between rows of genotypes (Jensen and Federer 1965) and has then been broadly applied to
83 the study of binary mixtures for estimating mixing ability, both when the performance of each
84 component is accessible in the mixture (McGilchrist 1965; Chalbi 1967; Gallais 1970; Federer
85 1979; Federer *et al.* 1982), or when only the mixture performance is surveyed (Federer 1979;
86 Federer *et al.* 1982; Gizlice *et al.* 1989; Knott and Mundt 1990; Gallandt *et al.* 2001). The
87 work developed hereafter deals with this second case, where only the global performance of
88 each mixture is available. The *general mixing ability (GMA)* refers to the average
89 performance of a variety in mixture (additive term) and the *specific mixing ability (SMA)*
90 relates to the interaction between the two components of a particular combination of
91 varieties. According to Griffing's decomposition of combining ability in hybrids (1956), the
92 performance of binary mixtures can be modeled as:

$$93 \quad Y_{ijbr} = \mu + a_b + \frac{1}{2}(GMA_i + GMA_j) + SMA_{ij} + e_{ijbr},$$

94 where Y_{ijbr} is the performance of the mixture of varieties i and j for block b (and possibly
95 replicate r in block b), μ is the intercept, a_b is the effect of block b , GMA_i is the general
96 mixing ability of variety i , SMA_{ij} is the specific mixing ability of varieties i and j grown
97 together, and e_{ijbr} is the error term of the observation. The $\frac{1}{2}$ coefficient is applied on the
98 GMA term since each variety accounts for only half of the plants grown in the plot (in case of
99 equal proportions at sowing).

100 However, due to the need to grow all the possible binary mixtures (that is to use a complete
101 design) among a set of varieties to estimate GMA and SMA, the field evaluation can be

102 cumbersome regarding the cost and management of experimental trials. For this reason,
103 authors have usually applied this method to a limited number of varieties (e.g. six or eight in
104 Gallandt *et al.* 2001; Gizlice *et al.* 1989 respectively). A key constraint for experimental
105 efforts, this curse of dimensionality has also been encountered when studying hybrids'
106 combining ability, and solved by the use of incomplete designs (Jensen 1959). By reducing
107 the number of mixtures to grow for each variety, such designs allow a wider range of varieties
108 to be evaluated. Nonetheless, due to the reduced number of observed mixtures using the
109 Griffing's model (1956), the use of incomplete design raises the question of the definition of
110 GMA and SMA. If GMA and SMA are defined assuming a complete design setting, then
111 some GMA and SMA cannot be estimated in practice if the actual design is incomplete.
112 Otherwise, if GMA and SMA are defined conditionally to the set of observed mixtures only,
113 then the definition of GMA and SMA is specific to this set of mixtures.

114 Alternatively, hybrid's geneticists suggested the use of mixed linear models for the analysis of
115 crossing designs, in which the estimates of GCA and SCA are provided by Best Linear
116 Unbiased Predictors (BLUP, Möhring, Melchinger and Piepho 2011). In this framework, both
117 the parameters of the models (namely the intercept and the variance of the random effects)
118 and the (random) effects to be predicted have the same definition whatever the design (see
119 Appendix A for a theoretical comparison of the two approaches). Interestingly, despite the
120 known similarity between binary mixtures and hybrids, the methods developed for hybrids'
121 unbalanced designs have not been adapted so far to mixtures.

122 Moreover, there is also a need for suitable approaches to handle higher order mixtures (i.e.
123 mixtures including more than two components) as they have been found useful in different
124 studies (Lopez and Mundt 2000; Mille *et al.* 2006), and are closer to the three to five-way
125 mixtures commonly grown by farmers. Indeed, higher order mixtures present strong
126 agronomical interest, such as their ability to reduce disease development (Newton *et al.*
127 1997, Borg *et al.* 2018), and can provide multiple agroecosystem services (i.e. weed
128 suppression, yield and grain quality when considered altogether, Lazzaro, Costanzo, and
129 Bàrberi 2018). Here the mixtures reveal their fundamental divergence with hybrids: if hybrids

130 are constrained to binary and balanced combinations (1:1 of the two parental genomes,
131 except when considering polyploids), mixtures allow all degrees of freedom, considering any
132 number of components, in any possible proportion.

133
134 The aim of this article is to provide a methodological framework for the analysis of mixing
135 ability in incomplete designs for binary and higher order mixtures based on dedicated
136 statistical methods using mixed models. For a given panel of genotypes, this allows (i) the
137 assessment of the relative importance of the additive part of the mixing ability vs the part that
138 results from specific interactions due to the combinations of genotypes (using two modeling
139 of the inter and intra-genotypic interactions thus providing a better understanding of mixture
140 functioning), (ii) the identification of the best performers in mixtures, and (iii) the prediction of
141 mixture performances using mixing ability modeling. The approach was applied to two
142 contrasted cases: (i) a trial of wheat binary mixtures and their pure stands, and (ii) a trial
143 including higher order mixtures.

144

145

146 **2. MATERIAL AND METHODS**

147 **2.1 Modeling of mixing ability:**

148 *2.1.1 Griffing's model with fixed GMA and SMA effects*

149 Mixing ability can be modeled using the same formalism as proposed by Griffing for
150 combining ability (1956). In this setting, the GMA and SMA are defined as fixed effects and
151 the residual term as random. This model is associated to the particular context of complete
152 diallel experiment design i.e. (i) all the possible mixtures are assessed and (ii) only binary
153 mixtures are considered, in 1:1 proportions. The use of this model in a broader context raises
154 some statistical issues: in case of incomplete design the condition (i) is not fulfilled, so the
155 definition of the terms becomes unclear, the estimability of the GMA and SMA is not
156 warranted and the estimators initially proposed by Griffing are no longer valid; if requirement

157 (ii) is not met, a model for SMA in higher order mixture should be proposed. To address the
 158 problem of the definition of GMA and SMA resulting from assumption (i) we propose to treat
 159 the genetic effects as random effects in a mixed model (as proposed, but not developed, by
 160 Griffing 1956, and further developed by Möhring *et al.* 2011 for the prediction of combining
 161 ability in hybrids).

162

163 2.1.2 Models with random GMA and SMA effects

164 In order to analyze the diverse experimental designs allowed by mixtures, we have adapted
 165 the previous model to binary and higher order mixtures (but considering in a first step equal
 166 proportions for each component). First, the model can be defined to estimate a block effect
 167 and to predict GMA effects:

$$168 \quad Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \mathring{\mathbf{a}}_{k(n)=1}^{K(n)} GMA_{k(n)} + e_{nbr} \quad (\text{Model 1})$$

169 where Y_{nbr} stands for the performance of the r -th replicate of mixture (or genotype) n in block
 170 b , n is the mean of the experiment, a_b is the effect of block b , $K(n)$ is the number of
 171 components of mixture n and $k(n)$ stands for the k -th genotype within mixture n (or $K(n) = 1$
 172 and $k(n) = 1$ if n is a pure stand), $GMA_{k(n)}$ is the general mixing ability of the genotype, and
 173 e_{nbr} is the residual term.

174 In a second model, we included the SMA effects corresponding to first order interactions
 175 between genotypes (i.e. at the level of pairs of genotypes), supposing that higher order
 176 interactions are negligible (their estimation would require an even larger experimental effort).
 177 The model is defined as follows:

$$178 \quad Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \mathring{\mathbf{a}}_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{\binom{K(n)}{2}} \mathring{\mathbf{a}}_{k(n)=1}^{K(n)-1} \mathring{\mathbf{a}}_{k(n)=k(n)+1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \quad (\text{Model 2})$$

179 where a_b is the effect of block b , $\binom{K(n)}{2}$ is the number of all possible pairs of different
180 genotypes in mixture n of $K(n)$ components when $K(n) > 1$, and $SMA_{k(n)k'(n)}$ stands for
181 inter-genotypic interaction between components $k(n)$ and $k'(n)$ of the mixture.

182 It should be noticed that n can be either a mixture (m) or a pure stand (p). The pure stands
183 are thus included in the estimation of the GMA as an intra-genotypic mixture (in contrast with
184 Gallais 1970), with the originality to consider the SMA of pure stands (SMA_{ii}). This intra-
185 genotypic interaction SMA_{ii} indicates how a given genotype performs in pure stand compared
186 to the mean of mixtures comprising this genotype. In the case of pure stands (when
187 $K(n) = 1$ and only variety $l(n)$ is grown) the model writes:

$$188 \quad Y_{nbr} = \mu + a_b + GMA_{l(n)} + SMA_{l(n)l(n)} + e_{nbr},$$

189 where $SMA_{l(n)l(n)}$ stands for intra-genotypic interaction within the grown genotype in pure
190 stand. The importance of intra-genotypic competition has been raised before, but only in the
191 context of experiments where the performance of each component in a mixture is accessible
192 (Gallais 1970). The access to intra-genotypic SMA through the integration of pure stands
193 directly in the analysis therefore provides an original description of the effect of competition in
194 a pure stand.

195 In addition to the residual term, the GMA and SMA terms are also defined as random effects
196 which ensures estimability even for incomplete designs. The precise distribution of the
197 random effects will be detailed in the following section (2.2).

198 Higher order mixtures might be interesting to characterize mixing ability of genotypes, since
199 for a given number of plots, they allow to observe each genotype interacting with a higher
200 number of partners compared to a design of binary mixtures only, though with a lower
201 contribution of the genotype to each mixture performance.

202

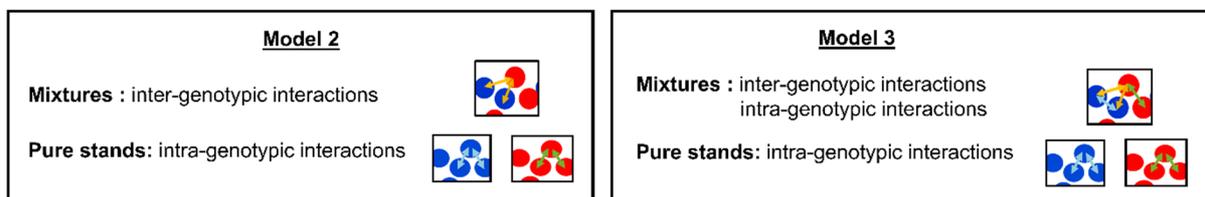
203 *2.1.3 Intra-genotypic interactions in mixture*

204 In a mixture, a plant is exposed to inter-genotypic interactions but also to intra-genotypic
 205 interactions, so we modified the model to better describe the biological reality for specific
 206 effects, according to the components proportions in mixture:

207
$$Y_{nbr} = \mu + a_b + \frac{1}{K(n)} \mathring{\mathbf{a}}_{k(n)=1}^{K(n)} GMA_{k(n)} + \frac{1}{(K(n))^2} \mathring{\mathbf{a}}_{k(n)=1}^{K(n)} \mathring{\mathbf{a}}_{k'(n)=1}^{K(n)} SMA_{k(n)k'(n)} + e_{nbr} \quad (\text{Model 3})$$

208 where a_b is the effect of block b . In this model, intra-genotypic interaction ($k(n)' = k(n)$) is
 209 therefore always part of the mixture performance, for any mixture order ($K(n)^3 - 1$). For
 210 instance, for a given binary mixture n of genotypes i and j , the interaction term between the
 211 components can be written as: $\frac{1}{4}SMA_{ii} + \frac{1}{2}SMA_{ij} + \frac{1}{4}SMA_{jj}$ since $SMA_{ji} = SMA_{ij}$.
 212 Observe that for $K(n) = 1$, Model 2 and Model 3 coincide. The coefficients $\frac{1}{(K(n))^2}$ represent
 213 the expected weight of each kind of neighborhood in the plant community, assuming a
 214 random distribution of genotypes in the plant community. It should be noticed that due to the
 215 differences in SMA effects weighting, the SMA variance and SMA BLUPs are expected to be
 216 higher with Model 3 than with Model 2. Nevertheless, the integration of neighboring
 217 probabilities in the modeling of mixing ability (especially through the introduction of intra-
 218 genotypic interactions within mixture) constitutes a novel advance in mixture analysis.

219



220

221 **Figure 1:** Schematic representation of plant interactions modeled by Model 2 and Model 3

222

223 The motivation for considering both models is that Model 2 only accounts inter-genotypic
 224 interactions within mixtures, while Model 3 also accounts for intra-genotypic interactions in
 225 mixtures (Figure 1). For convenience we adopted notations for all models similar to the
 226 Griffing's notations for combining ability. We emphasize that the interpretation of the different
 227 terms is quite different between Model 2 and Model 3. To illustrate these differences, we
 228 consider the expected performance over all possible equally weighted mixtures of order K
 229 that include a given genotype k . This amounts to work conditionally on all random effects that
 230 depend on k only, so we denote this quantity by $E(Y | k, K)$.

231 One has

232 $E(Y | k, K) = \mu + \frac{GMA_k}{K}$ if $K > 1$ in Model 2

233 $E(Y | k, K) = \mu + GMA_k + SMA_{kk}$ if $K = 1$ in Model 2

234 $E(Y | k, K) = \mu + \frac{GMA_k}{K} + \frac{SMA_{kk}}{K^2}$ in Model 3

235 As a consequence, SMA_{kk} can be interpreted as the expected difference in performance
 236 between pure stand and mixture in Model 2, but not in Model 3 where it also includes intra-
 237 genotypic interaction within mixture. Although the interpretation of SMA_{kk} is different from the
 238 one of $SMA_{kk'}$ in Model 2, we assumed a common distribution for these two terms in the
 239 following to reduce the number of variance parameters to be inferred.

240

241 **2.2 Statistical analysis:**

242 *Inference method*

243 The mixed model framework has already been widely applied to plant breeding (Bernardo
 244 1996; Falconer *et al.* 1996; Lynch and Walsh 1998; Piepho and Möhring 2007), but to our

245 knowledge not to crop mixing ability analysis. This framework allows studying mixtures of any
246 order in incomplete designs, ensuring model estimability (as discussed above).

247 Models 2 and 3 can be expressed in a matrix form as

$$248 \quad y = Xb + Z_1g + Z_2s + e ,$$

249 where y is the vector of performances, b is the vector of fixed effects, g is the vector of the
250 GMA random effects, and s is the vector of SMA random effects. X , Z_1 and Z_2 are the
251 corresponding design matrices.

252 The random effects are assumed to be normally distributed according to the following
253 distributions $g \sim N(0, s_{GMA}^2 A_1)$, $s \sim N(0, s_{SMA}^2 A_2)$ and $e \sim N(0, s_e^2 I)$. The variance of the
254 observed values can be decomposed as:

$$255 \quad Var(y) = Z_1 A_1 Z_1' s_{GMA}^2 + Z_2 A_2 Z_2' s_{SMA}^2 + I s_e^2 .$$

256 In the present work, we used identity matrices for A_1 and A_2 matrices. Note that one can
257 account for *a priori* similarities between genotypes by specifying more sophisticated variance
258 covariance structures, e.g. accounting for genetic similarities between genotypes through
259 relatedness matrices (Bernardo 1995).

260 As for the estimation of the variance parameters, we used the Restricted Maximum
261 Likelihood (REML). We used the Best Linear Unbiased Predictors (BLUP) to predict the GMA
262 and SMA effects.

263

264 **2.3 Experimental designs:**

265 The analysis has been carried on two data sets produced by two experiments assessing
266 mixing ability of a panel of 25 genotypes of winter wheat. This panel was designed to be
267 representative of the phenotypic diversity of a larger set of 58 genotypes, evaluated in pure

268 stand in a previous study for root and shoot architecture as well as other functional traits
269 (Cantarel *et al.* in prep., Dubs *et al.* 2018). This 25 genotypes panel is composed of nine elite
270 varieties cultivated for their high yielding performance in the Parisian Basin (Altigo, Apache,
271 Arezzo, Boregar, Grapeli, Renan, Sogood, Soissons, Trémie), four landraces (Blé autrichien,
272 Rouge de Bordeaux, Rouge du Roc, Saint Priest), six varieties bred for organic agriculture
273 (Alauda, Hermes, Maxi, Midas, Ritter, Skerzzo), and six lines from a MAGIC population (A22,
274 A208, A243, A398, F236, F426, Thépot *et al.* 2015) (Table B.1). The way mixtures of two,
275 four and eight genotypes of the panel were designed in the two following experiments has
276 been inspired from the ecology literature, and more specifically the Jena diversity experiment
277 (Weisser *et al.* 2017), exploring extensively the gradient of intraspecific functional diversity
278 (Dubs *et al.* 2018).

279 In the first experiment, the 25 genotypes were grown in a field trial both in pure stands and in
280 75 binary mixtures during the season 2014-2015 in the Le Moulon (LM) experimental station
281 (48°42'37.2"N, 2°09'37.6"E) in France, according to an incomplete diallel design (75 of the
282 300 binary combinations: Figure 2.A). Each entry was sown in a 7,5m² plot at a density of
283 160 grains.m⁻² (with 20 cm row width). All pure stand genotypes and binary mixtures were
284 replicated twice in two randomized complete blocks. The mixtures were prepared for sowing
285 using the thousand kernel weight in order to get equal proportion of seeds of each genotype.

302 In the second experiment (Dubs *et al.*, 2018), a subset of 16 genotypes of the panel (Altigo,
303 Arezzo, Boregar, Grapeli, Renan, Soissons, Trémie, Blé autrichien, Alauda, Hermes, Maxi,
304 Midas, Ritter, Skerzzo, A22, F426) has been assessed in the INRA experimental station of
305 Clermont-Ferrand (CF) (45°46'04.2"N, 3°08'52.2"E) in 2014-2015, in pure stand and in 24
306 mixtures of two genotypes (also observed in the Le Moulon trial), 28 mixtures of four
307 genotypes and 20 mixtures of eight genotypes (Figure 2.B), where the components of each
308 mixture were also in equal proportions of seeds. They were sown in 7,5m² plots, for a target
309 density of 250 plants.m⁻² and replicated in two randomized complete blocks.

310 In both experiments, three fungicide treatments were applied in order to study mixtures
311 without disease development since it has already been shown that mixtures can lead to
312 important disease reduction on wheat, and we aimed at focusing on compensation effects
313 (Stützel and Aufhammer 1990). Herbicides were applied to avoid weed competition to
314 interfere with wheat competition, and a nitrogen fertilization was applied (60 kg N/ha and then
315 40 kg N/ha of nitrogen fertilizer in the LM trial, and 60 kg N/ha in the CF trial) according to the
316 leftovers from the previous crops on each trial to reach the common objective of 150 kg N/ha.

317

318 *Phenotypic data*

319 Yield and its components (number of spikes.m⁻², number of grains/spike, thousand kernel
320 weight (TKW)) were recorded on each plot in the LM experiment, while only yield and protein
321 content were recorded in the CF experiment.

322 Yield was calculated as the weight of grain harvested on the plot surface, and after
323 measuring the humidity of a sample, it was standardized into q.ha⁻¹ at 15% humidity (1quintal
324 / hectare = 100kg / 10.000m² = 10⁻² kg.m⁻²). The number of spikes has been counted after
325 flowering date, on one meter length for two adjacent rows in each plot, and converted into a
326 number of spikes.m⁻² taking into account the distance between rows (20 cm). TKW was
327 measured after harvest and threshing, and the number of grains/spike was calculated based

328 on the yield components (= (yield/ TKW)/ number of spikes.m⁻²). Protein content was
329 estimated on grain samples using near-infrared spectroscopy (NIRS).

330

331 **2.4 Data analysis:**

332 *2.4.1 Application of the models to the datasets of each location*

333 For each location, the correlation between the observed mixtures and the mean of their pure
334 stand components was calculated. The main analyses (model comparison, analysis of the
335 variance components and the BLUPs) were then performed with the three models on all the
336 observations (mixtures and pure stands) of each location separately, and the predicted
337 values are the modeled values fitted, on the already observed data. The correlation between
338 all the observed values and the predictions from (i) the means of the genotypic effects of the
339 corresponding pure stand components (according to Federer *et al.* 1982 and Gizlice *et al.*
340 1989), or (ii) a subset of observations of the same location (mixtures only using Model 3) has
341 been performed for each site. Finally, the CF observed values for different mixture orders
342 were compared with predictions obtained from CF or LM trials analyzed with Model 1. The
343 analysis was performed using an adaptation of the *lme4* R package (Bates *et al.* 2015; see
344 File S.1).

345

346 *2.4.2 Model comparison*

347 In order to compare the relative importance of GMA, SMA and intra-genotypic vs inter-
348 genotypic interaction, we compared the following models:

- 349 - Model 1 which includes only the block effect and the GMA,
- 350 - Model 2 which includes the GMA and the SMA, with inter-genotypic interactions in
351 mixture, and intra-genotypic interactions for pure stands only,

352 - Model 3 which includes the GMA and the SMA, with inter-genotypic interactions and
353 intra-genotypic interactions in mixture in addition to intra-genotypic interactions in
354 pure stand.

355 The models were compared based on the AIC, BIC and on Likelihood Ratio Tests (LRT,
356 using Maximum Likelihood procedure (ML)) which were further performed for nested models
357 (Model 2 vs Model 1, and Model 3 vs Model 1). We also computed the second-order AIC and
358 the conditional AIC using MuMIn and cAIC4 R packages respectively, but no difference with
359 the AIC values was observed. In addition, the R^2 and the RMSE were reported for the
360 different models. While open to criticism (Alday 2016), the calculation of a R^2 for mixed
361 models is a convenient tool for assessing the model quality using a common dimensionless
362 metric for the different response variables. The R^2 were calculated following Xu (2003) to
363 assess the proportion of the total variance which is explained by the model (conditional
364 variance, for both fixed and random effects), as one minus the ratio of the residual variance
365 of the full mixed model, over the residual variance of a null model (the fixed intercept-only
366 model).

367

368 *2.4.3 Within-location comparison of observed values with predictions from BLUP based on*
369 *subsets of observations*

370 We investigated the possible impact of including pure stands in the BLUP of mixture in the
371 same location. For that, the observed values (mixtures and pure stands) were correlated to (i)
372 the BLUP values predicted from mixture data only (i.e. without pure stands) and (ii) the BLUP
373 from the complete data set. In both cases, BLUP were performed using Model 1 and
374 alternatively using the best model for each response variable.

375

376 *2.4.4 Inter-trial predictions*

377 Finally, the observed values of the CF trial for different mixture orders were compared with
378 predictions based on the GMA predicted using Model 1 on all the observations of the LM
379 experiment.

380 The observed values were also compared to the predictions based on (i) the GMA in LM
381 predicted with Model 1 fitted on the mixture observations only, or on (ii) the mean of the
382 genotypic effects assessed on the LM pure stands only instead of GMA. Model 1 was
383 preferred for predictions because (i) regarding breeding strategy, it is preferable to being able
384 to predict mixture performances based on GMA than based on GMA and SMA, since using
385 SMA requires having observed the corresponding pairs of genotypes (for using Models 2 or
386 3), and (ii) for practical considerations, in this study no higher order mixture had all its SMA
387 predicted in the LM trial in order to predict their performances, so SMA are diluted in higher
388 order mixtures. Nevertheless, as all the binary mixtures observed in the CF experiment were
389 also observed in the LM trial (so they were all predictable from LM BLUPs), we assessed how
390 including SMA (using Models 2 and 3) affects prediction of the yields of binary mixtures in the
391 CF trial.

392

393 **2.5 Data and program availability:**

394 The data and the program are available on an online public repository
395 (<https://github.com/cambroise/lme4-adapt-for-variety-mixture>). The program is described in
396 File S.1.

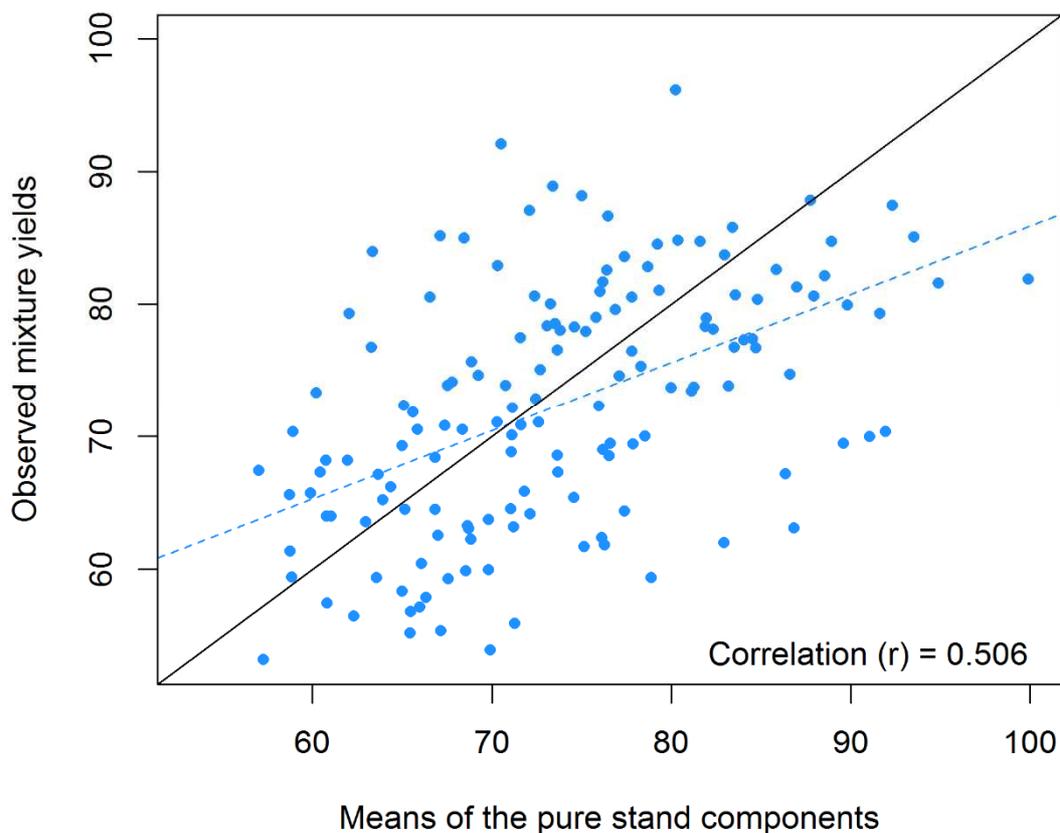
397

398

399 **3. RESULTS**

400 *3.1 Correlation between mixtures and pure stands*

401 The Pearson's correlation between the observed binary mixtures performances and the
 402 means of their pure stand components was high for TKW (0.91) but moderate for yield (0.51,
 403 Figure 3), for the number of spikes.m⁻² (0.51), and for the number of grains per spike (0.42) in
 404 the Le Moulon (LM) experiment, showing the need for modeling of mixing ability. The
 405 correlation in Clermont-Ferrand (CF) reached 0.75 for yield and 0.63 for the protein content
 406 (respectively 0.81 and 0.73 when considering only binary mixtures).



407
 408 **Figure 3:** Observed yield for binary mixtures and the means of their pure stand components for the Le
 409 Moulon experiment

410 The dotted line is the regression line ($y= 34.395 + 0.515 x$, with standard errors of 5.395 and 0.073
 411 respectively, and DF= 146). The black solid line is the $y=x$ line.

412
 413 *3.2 Model comparison*

414 Based on LRT (Likelihood Ratio Test) using ML procedure for model comparison, both
 415 models including SMA have been found significantly better than Model 1 for TKW (p-values
 416 of 0.026 and 0.030 respectively, Table 1). Models 2 and 3 had low but non-significant p-
 417 values (0.060 and 0.182) for spike density, while for yield and other response variables in LM
 418 and CF trials the p-values were non-significant.

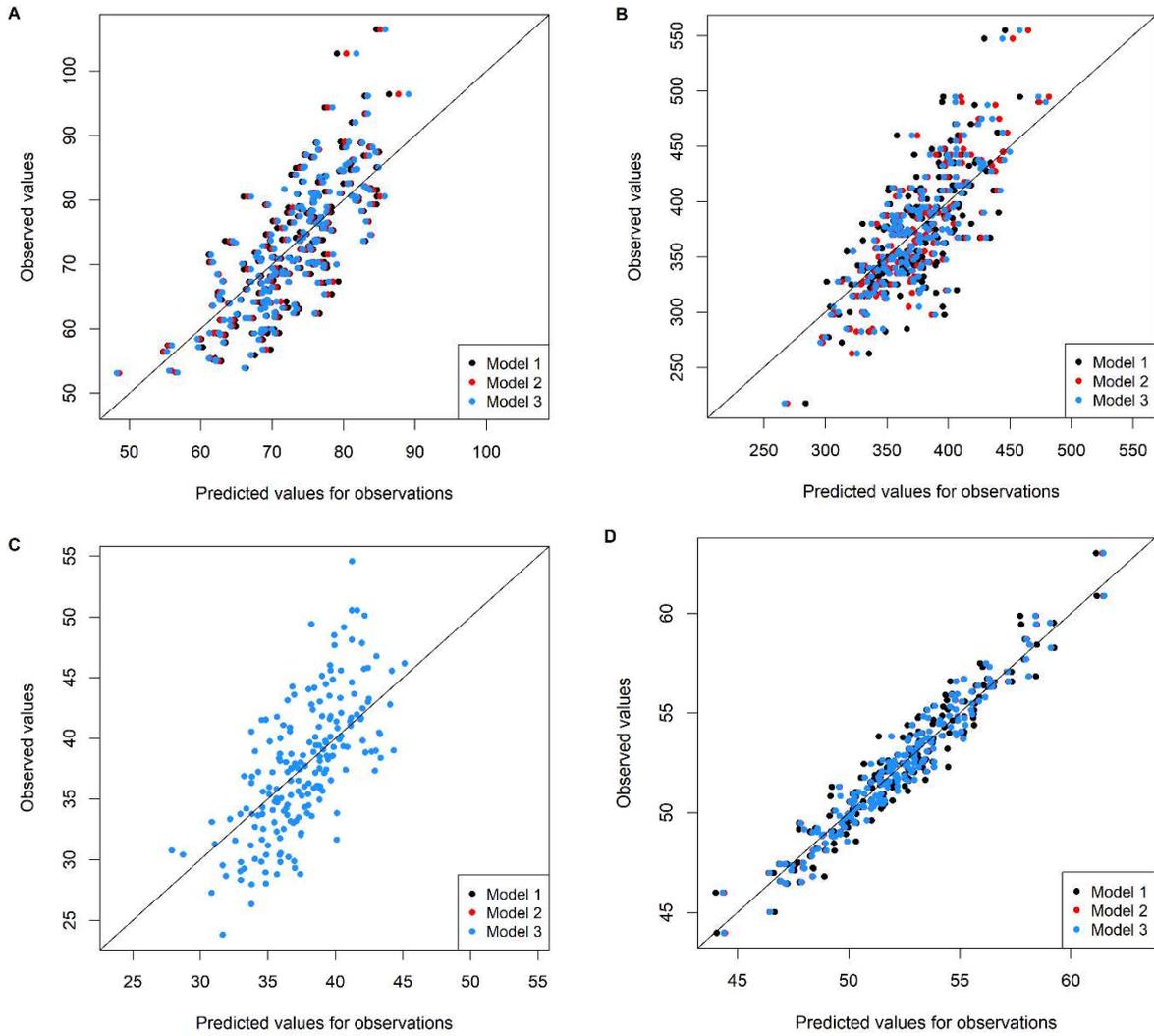
Response variable	Model	AIC	BIC	p-value
Yield	Model 1	1393.3	1406.5	NA
	Model 2	1395.1	1411.6	0.664
	Model 3	1394.6	1411.0	0.385
Spike density	Model 1	2075.5	2088.7	NA
	Model 2	2074.0	2090.4	0.060 (.)
	Model 3	2075.7	2092.2	0.182
Grain number per spike	Model 1	1188.4	1201.6	NA
	Model 2	1190.4	1206.9	1
	Model 3	1190.4	1206.9	1
Thousand kernel weight	Model 1	672.5	685.7	NA
	Model 2	669.6	686.0	0.026 (*)
	Model 3	669.8	686.3	0.030 (*)

419 **TABLE 1:** Comparison of Models 2 and 3 to Model 1 in LM trial using the maximum likelihood
 420 procedure

421 The p-values are based on LRT (likelihood ratio test) comparisons. “.” indicates significance < 0.1 and
 422 “*” significance < 0.05.

423

424 Model 1, the most parsimonious as only based on GMA effects, was as efficient as Models 2
 425 and 3 for the number of grains per spike in LM (Figure 4.C) or for the yield and the protein
 426 content in CF (Figures 5).



427

428

429

Figure 4: Observed and predicted values for each response variable in LM trial

430

A. Yield, B. Spike density, C. Grain number per spike, D. Thousand kernel weight. Model 1 is the

431

model comprising fixed effects and GMA, Model 2 additionally includes SMA effects (inter-genotypic

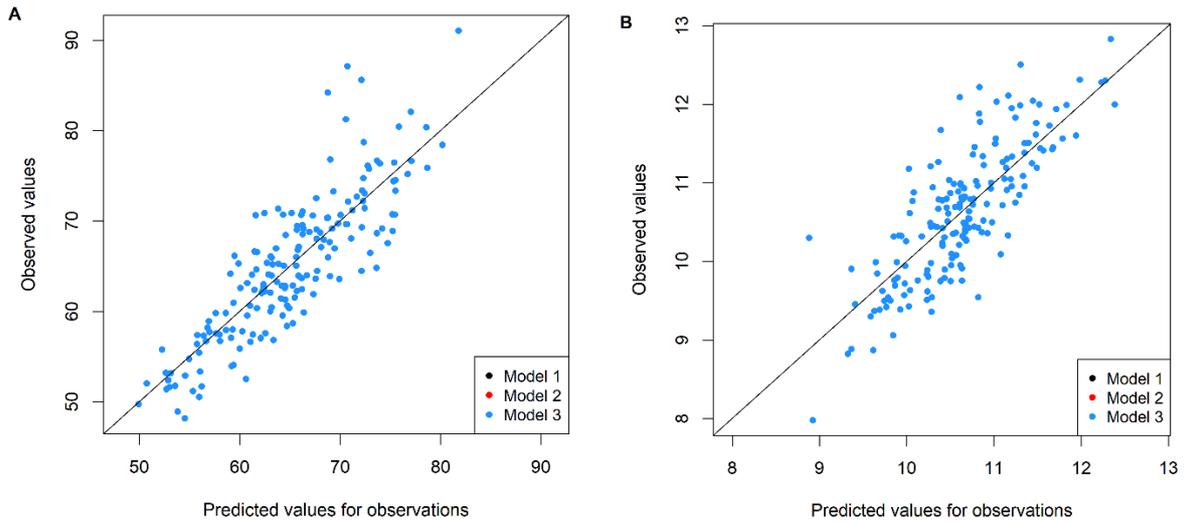
432

effects within pairs of genotypes within mixtures, and intra-genotypic effects for pure stands), Model 3

433

further includes intra-genotypic effects within mixtures. The black diagonal is the $y=x$ line.

434



435

436 **Figure 5:** Observed and predicted values for each response variables in CF trial

437 A. Yield, B. Protein content. The black and red dots are hidden by the blue dots due to equal values for
 438 the three models. The black diagonal is the $y=x$ line.

439

440 However, considering the yield in the LM experiment (Figure 4.A), Model 3 including SMA
 441 effects showed a higher R^2 and a lower RMSE (Table 2), although the likelihood of Model 3
 442 was not significantly different from the one of Model 1 with ML procedure. For the spike
 443 density and TKW (Figures 4.B and D), Models 2 and 3 had higher R^2 and lower RMSE than
 444 Model 1, Model 2 appearing slightly better than Model 3 for spike density. These models
 445 including the SMA effects seemed to partly reduce the shrinkage effect observed in Figure 4.

446

Response variable	Model	R^2	RMSE
Yield (LM)	Model 1	0.563	6.690
	Model 2	0.593	6.454
	Model 3	0.613	6.294
Spike density	Model 1	0.508	36.908
	Model 2	0.650	31.125
	Model 3	0.613	32.702
Grain number per spike	Model 1	0.456	4.023
	Model 2	0.456	4.023

	Model 3	0.456	4.023
TKW	Model 1	0.911	0.923
	Model 2	0.941	0.754
	Model 3	0.940	0.758
Yield (CF)	Model 1	0.728	4.174
	Model 2	0.728	4.174
	Model 3	0.728	4.174
Protein content	Model 1	0.642	0.502
	Model 2	0.642	0.502
	Model 3	0.642	0.502

447 **TABLE 2:** Criteria for characterization of the model fits in Le Moulon and Clermont-Ferrand trials

448

449 *3.3 Comparison with predictions based on the pure stands*

450 When using mixing ability models fitted on all the observations instead of the pure stand
451 genetic effects, the R^2 , the RMSE, the Pearson's correlation coefficient (on the values of the
452 mixtures and the pure stands) and the Kendall's correlation coefficient (on their ranking)
453 between observed and predicted values were improved for most of response variables in LM
454 experiment, but this was marginal on the CF trial, as shown in Table B.4. For instance, for the
455 yield in LM trial, the Pearson's correlation coefficient increased from 0.687 with predictions
456 based on the means of the pure stands (p), to 0.790 with predictions based on BLUP
457 predicted with Model 3 ($m+p$) showing a higher ability to predict mixture performances. In the
458 CF experiment, the correlation between observed and predicted values was as high with
459 predictions based on the mean of the genotypic effects of the pure stand components (0.840,
460 Table B.4) as with BLUP predicted from mixtures and pure stand observations (0.854, with
461 Model 1).

462

463 *3.4 Variances of GMA and SMA*

464 The variance components estimated with each model for yield and for the number of
465 spikes.m⁻² are detailed below for the LM and CF experiments (Table 3). The other responses

466 are given in supplementary material (Table B.2). It should be noticed that Models 2 and 3
 467 differ in the weighting of the SMA effects, and it is expected for estimated SMA variance to be
 468 higher with Model 3 than with Model 2 (as well as SMA BLUPs).

Response variable	Model	σ_{GMA}^2	σ_{SMA}^2	σ_e^2	$\sigma_{\text{SMA}}^2 / \sigma_{\text{GMA}}^2$
Yield Le Moulon trial	Model 1	63.57	NA	50.48	NA
	Model 2	63.10	2.11	48.66	0.03
	Model 3	54.69	14.67	47.40	0.27
Yield Clermont-Ferrand trial	Model 1	95.18	NA	19.20	NA
	Model 2	95.18	0	19.20	0
	Model 3	95.18	0	19.20	0
Number of spikes.m ⁻² Le Moulon trial	Model 1	1621.82	NA	1531.06	NA
	Model 2	1536.02	303.75	1267.18	0.19
	Model 3	1357.87	822.18	1338.40	0.61

469 **TABLE 3:** Variance components and ratio of variances in Le Moulon and Clermont-Ferrand trials

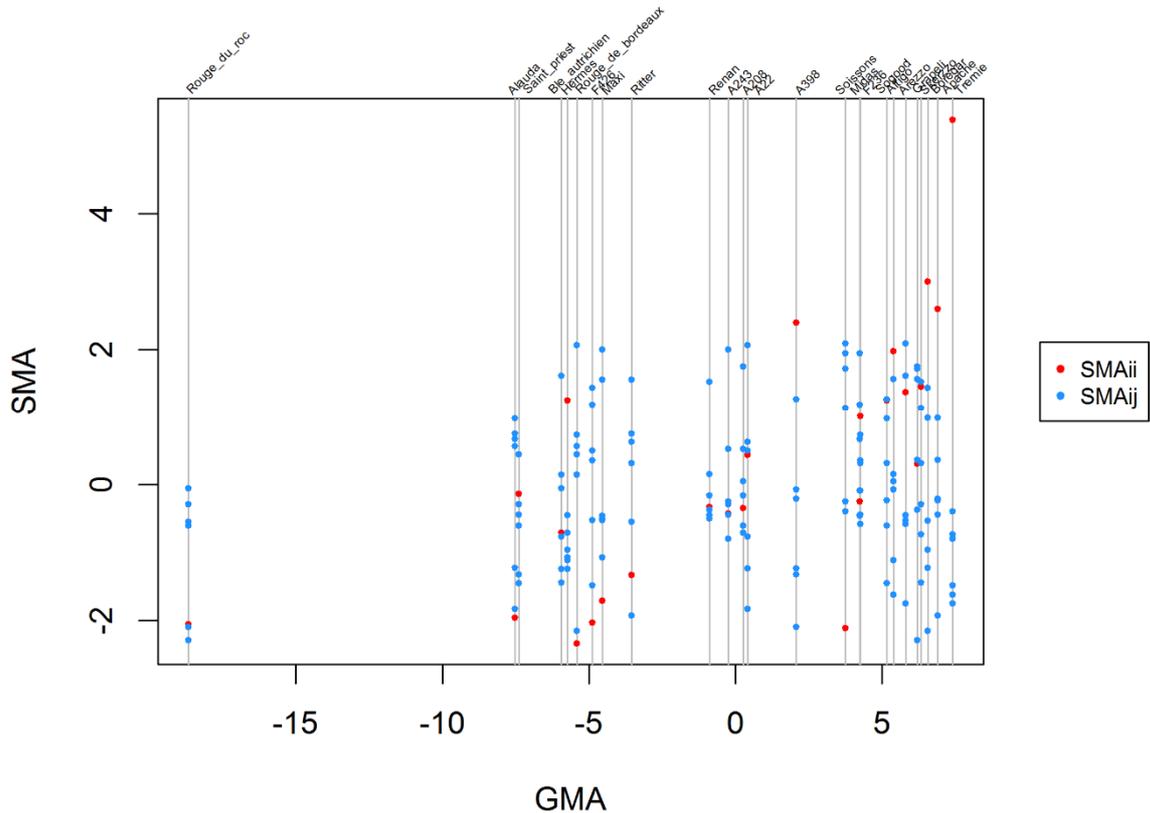
470

471 In all experiments and with all considered mixed models, the GMA variance was always
 472 higher than the residual variance and was more than 10 times higher in the case of the TKW
 473 in LM experiment (Table B.2). The amount of the SMA variance compared to the GMA
 474 variance ranged from 0 to 0.19 with Model 2 and from 0 to 0.61 with Model 3. The highest
 475 rates were obtained for spike density, while the number of grains per spike and the CF
 476 response variables did not show any specific effects. Consequently, the three models were
 477 equivalent for variance partitioning in the CF experiment.

478

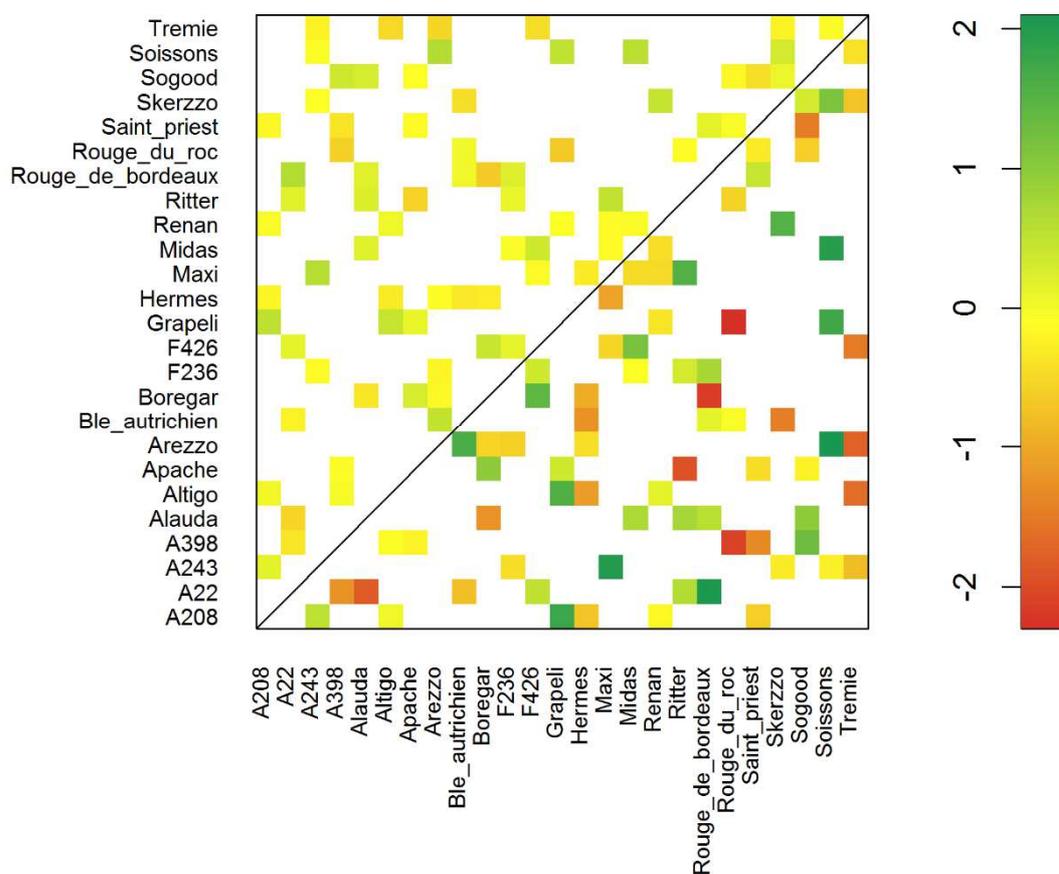
479 *3.5 GMA and SMA predictions*

480 The predicted values of GMA and SMA with Model 3 for yield in LM trial are shown in Figure
 481 6 (in Figure B.1 for spike density and TKW with Model 2 in LM trial) and the GMA values for
 482 each response variable with the best model are given in Table B.3.



483 **Figure 6:** Predicted values for GMA and SMA per genotype for yield in LM trial with Model 3
 484
 485 Values in quintal per hectare ($10^{-2} \text{ kg.m}^{-2}$). The red dots represent the intra-genotypic SMA, and the
 486 blue dots the inter-genotypic SMA.
 487 The GMA for yield at LM ranged from -18.67 q.ha^{-1} to 7.41 q.ha^{-1} (with Model 3), while the
 488 SMA showed lower ranges (from -2.34 q.ha^{-1} to 5.38 q.ha^{-1}). The range of SMA values per
 489 genotype with Model 3 varied among genotypes, with varieties such as Renan and Midas
 490 obtaining similar SMA with most genotypes (Figure 6). Overall, intra-genotypic SMA effects
 491 displayed both high and low values, while inter-genotypic SMA were somehow more
 492 intermediate (Figure 6). The intra-genotypic SMA obtained with Model 2 were less extreme
 493 for spike density and TKW than for yield (Figures B.1 A and B). The intra-genotypic SMA with
 494 Model 3 was overall positively correlated with the GMA for yield (0.679). This was also the
 495 case for spike density (0.659) but the correlation was lower for TKW (0.235).
 496 The elite variety Trémie had the highest GMA and displayed the highest SMA when grown in
 497 pure stand (Figure 6). Interestingly, two genotypes with relatively low GMA (A22 and Rouge

498 de Bordeaux, Figure 6) showed a rather high SMA value when grown together in a mixture
 499 (Figure 7). More extreme values were obtained with Model 3, due to higher SMA variance (as
 500 expected). This was also the case for spike density and TKW in LM trial although to a lesser
 501 extent (Figure B.2). The inter-genotypic SMA for Model 2 and Model 3 were very highly
 502 correlated for the yield, spike density and TKW (0.9993, 0.9996 and 0.9998 respectively).
 503



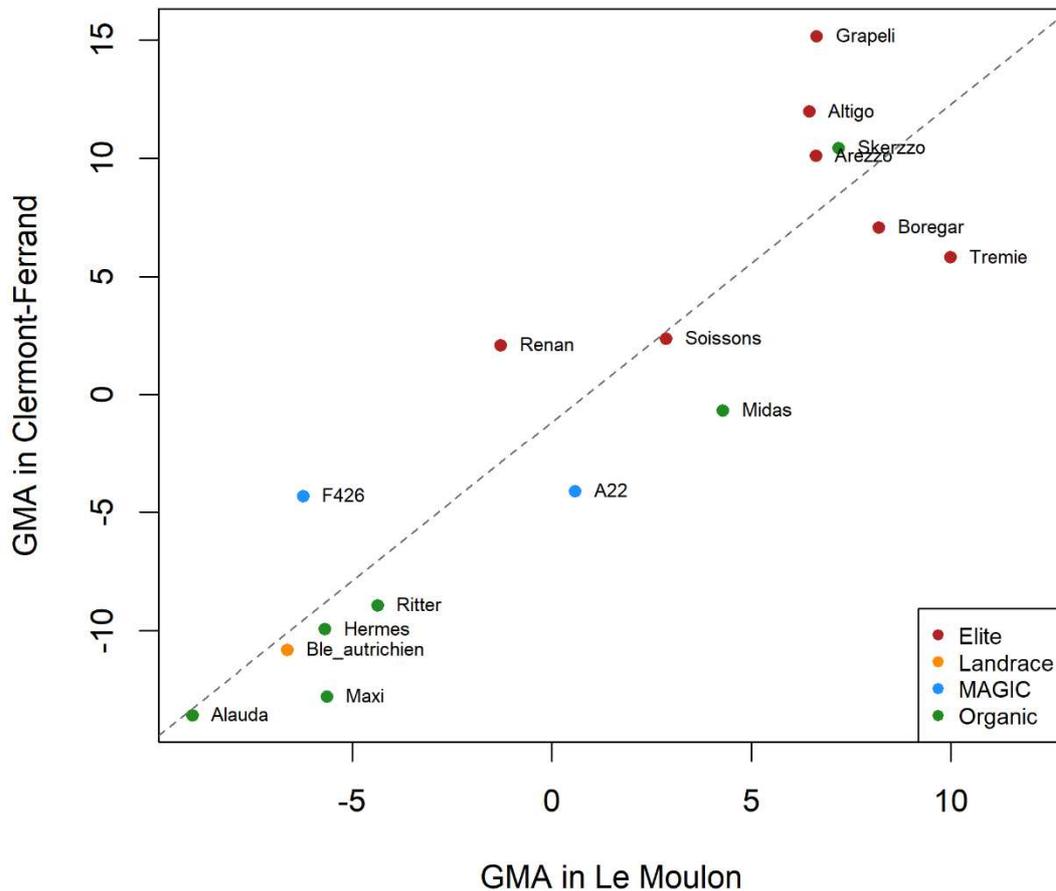
504 **Figure 7:** SMA predicted values for yield with Models 2 and 3 in LM trial
 505

506 The SMA obtained with Model 2 and Model 3 are plotted respectively above and below the diagonal
 507 (black line). The intra-genotypic SMA are not represented on the plot.

508
 509 The GMA was highly correlated with the pure stand genotypic effect in the CF trial (0.98 for
 510 the yield with Model 1, Table B.5), and to a lesser extent in the LM experiment (0.88). The
 511 correlation between the GMA obtained in both experiments for yield using Model 1 was also

512 quite high (0.903). The CF trial displayed higher variance for the GMA values than did the LM
513 trial (Figure 8).

514



515

516 **Figure 8:** GMA values predicted with Model 1 for yield in the two experiments (LM and CF)

517 The dotted line is the regression line between the GMA of the two experiments ($y = -1.174 + 1.347x$,
518 with standard errors of 1.066 and 0.171 respectively, and $DF = 14$).

519 The elite varieties obtained the highest GMA in both experiments, while most organic
520 varieties and the landrace had the lowest, although some organic varieties also obtained high
521 GMA. Organic varieties performed better in LM trial than in CF trial. The best varieties were
522 more spread in CF trial than in LM trial.

523

524 *3.6 Predictions based on subsets of observations*

525 In the LM experiment, when the analysis was performed on mixture observations only (m)
 526 instead of all observations ($m+p$), the variance of SMA for yield (estimated with both Models
 527 2 and 3) became null. The correlation between the observed and predicted mixtures with
 528 Model 3 was 0.786 for yield when using GMA_m , not different from the one obtained with $m+p$
 529 observations (0.790), and the RMSE was reduced from 6.3q.ha⁻¹ to 5.8q.ha⁻¹ (Table B.4).
 530 This result suggests that mixing ability analysis might be performed without requiring pure
 531 stands in the dataset. The GMA for yield were modified (Figure B.3) although the correlation
 532 between GMA_m and GMA_{m+p} with Model 3 was 0.91 (Table B.5). In the CF experiment, the
 533 correlation between GMA_m and GMA_{m+p} (with Model 1) was even higher (0.96, Table B.5),
 534 and the correlation between observed and predicted values was equal when excluding the
 535 pure stands for GMA predictions (0.834) to the one when pure stands were included (0.854,
 536 Table B.4).

537

538 *3.7 Predictions of CF high order mixtures based on CF BLUP*

539 Both Pearson's and Kendall's correlations between the observed values for pure stands and
 540 mixtures in the CF trial and the predicted values based on the CF mean of the genotypic
 541 effects in pure stands decreased with the order of the mixtures (Table 4), indicating that the
 542 four-way and eight-way mixtures had the largest level of deviation from additivity in the CF
 543 experiment.

544

Mixture order	Pearson's correlation						Kendall's correlation					
	CF GE _p	CF GMA _{m+p}	CF GMA _m	LM GE _p	LM GMA _{m+p}	LM GMA _m	CF GE _p	CF GMA _{m+p}	CF GMA _m	LM GE _p	LM GMA _{m+p}	LM GMA _m
All orders	0.840	0.854	0.834	0.653	0.785	0.741	0.651	0.669	0.664	0.498	0.598	0.573
1	0.973	0.955	0.876	0.718	0.876	0.770	0.871	0.831	0.734	0.548	0.706	0.621

2	0.836	0.862	0.786	0.662	0.805	0.751	0.649	0.677	0.700	0.505	0.667	0.613
4	0.742	0.777	0.789	0.599	0.695	0.727	0.516	0.554	0.582	0.405	0.481	0.541
8	0.700	0.713	0.719	0.574	0.649	0.680	0.544	0.569	0.569	0.438	0.500	0.538

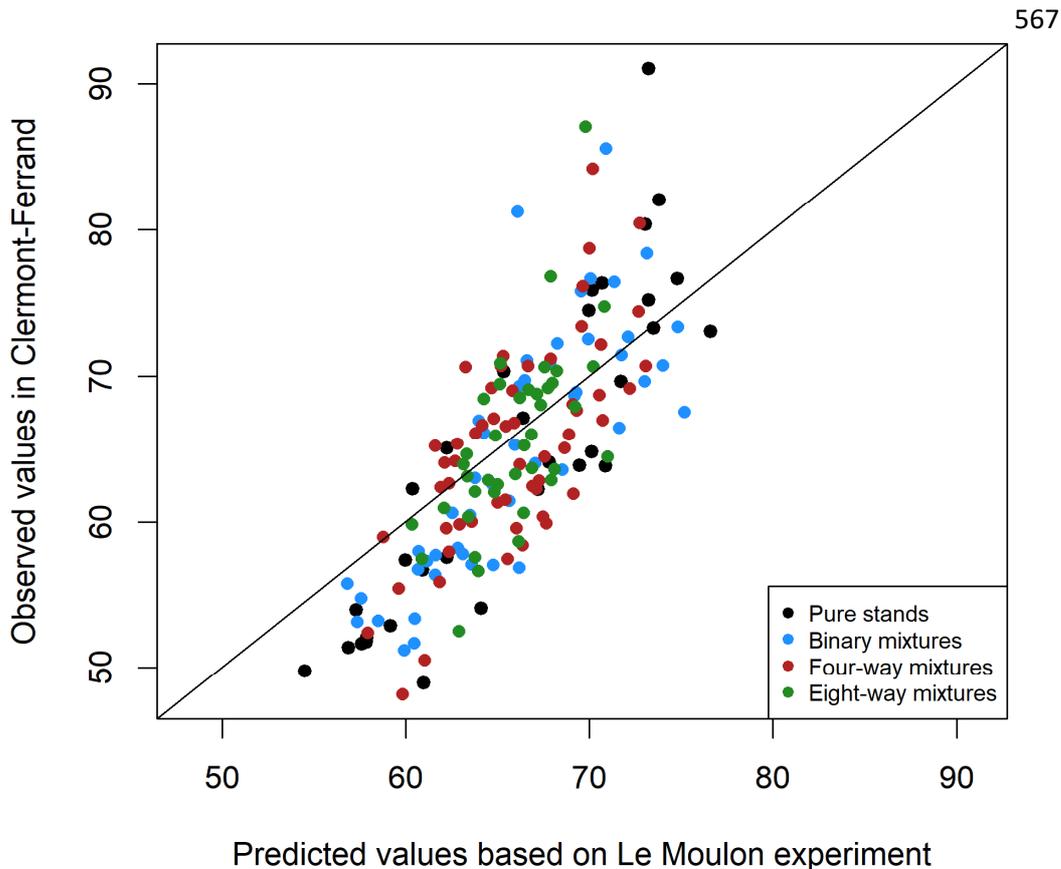
545 **TABLE 4:** Correlation between observed and predicted values or ranking for the yield in the Clermont-
546 Ferrand (CF) trial based on the CF or the Le Moulon (LM) mean of the genotypic effects in pure stands
547 (GE) or GMA predicted with Model 1 on different sets of observations in the CF and LM trials

548 The order of the mixture indicates the number of components: one for pure stand, two for binary
549 mixture, four for four-way mixture and eight for eight-way mixture. The capital letters indicate the trial
550 for observations for genotypic effects in pure stand or GMA predictions using Model 1. The lower case
551 letter specifies if pure stands observations were used (p) or mixture observations (m) or both (m+p) in
552 each trial for GMA predictions.

553

554 *3.8 Inter-trial predictions*

555 In order to assess the ability of the BLUP values obtained in a design based on binary
556 mixtures to predict the observed values for higher order mixtures, we computed the predicted
557 values for the CF observations based on the LM GMA using Model 1 (Figure 9). The
558 Pearson's correlation between the observed values in CF and predictions based on LM was
559 0.785 (and 0.598 for Kendall's correlation, Table 4). The correlation was slightly lower when
560 excluding the pure stand observations in LM of the analysis (0.741). However, the correlation
561 was always much higher than when using the mean of the corresponding genotypic effects of
562 pure stand components (correlation of 0.653). When including SMA in addition to GMA
563 predicted on all observations in the LM trial (using Models 2 and 3) for predicting yields of
564 binary mixtures in the CF trial, the correlation between observed and predicted yields was not
565 clearly improved (0.814 with Model 2 and 0.817 with Model 3, vs 0.805 with Model 1).



568 **Figure 9:** Predicted yields for the Clermont-Ferrand experiment based on GMA predicted from all
 569 observations in the Le Moulon trial with Model 1 and the corresponding observed yields

570 The black diagonal is the $y=x$ line.

571

572 As for predictions based on CF BLUP, when using LM BLUP, the Kendall's correlation
 573 coefficient between the predicted values and the observed values decreased with the
 574 mixtures complexity (Table 4). Interestingly, when removing the pure stands in LM
 575 observations for fitting Model 1, the observed vs. predicted correlations were lower for pure
 576 stands and binary mixtures than when including pure stands in LM, while they were rather
 577 similar for four-way and eight-way mixtures.

578

579

580 4. DISCUSSION

581 Mixed models are pivotal statistical tools for the prediction of breeding values in modern
582 breeding programs, and are commonly used to analyze diallel designs of hybrid crosses. This
583 formalism has surprisingly never been applied to cultivar mixtures analysis, despite the
584 similarity between Combining Ability and Mixing Ability. Herein we adapted mixed models for
585 the study of mixing ability in diallel-like designs, unlocking the analysis of incomplete designs.

586 We combined the Griffing formalism for combining ability (Griffing 1956) and mixing ability
587 modeling with the recent developments in mixed models (Bates *et al.* 2015) adapted to fit
588 specialized linear mixed models. The approach is highly flexible as it can be applied to binary
589 and any order mixtures, including pure stands.

590

591 *4.1 General Mixing Ability predominates in two experiments on wheat mixtures*

592 This approach has been applied to two experimental designs of bread wheat mixtures: one
593 based on pure stands and binary mixtures, and the other including higher order mixtures too.
594 In the first experiment (Le Moulon, LM), the correlation between the observed mixture
595 performances and their predicted performances based on their pure stand means was
596 moderate (0.51, 0.51 and 0.42 respectively for yield, for the number of spike.m⁻² and for the
597 number of grain/spike), except for TKW (0.91), underpinning the importance of estimating
598 mixing ability of the genotypes. Using mixing ability modeling allowed to improve the
599 correlation between observed values and predictions to 0.79 for the yield in Le Moulon (LM)
600 experiment for instance (with Model 3).

601 The part of variance explained by GMA and SMA effects was relatively high, given that the
602 proportion of observed mixtures on the total number of possible mixtures was rather low
603 ($75/300=0.25$). This suggests that using a random sample of 25% of the total number of all
604 possible mixtures, in addition to the pure stands, seems efficient to assess GMA and SMA of

605 the genotypes used as components, and this is in accordance with previous findings for
606 combining ability in wheat (Zhao *et al.* 2015). This result calls for the possibility to screen a
607 large number of genotypes from a diversified panel for their GMA using incomplete designs
608 instead of growing all pairwise mixtures within a limited set of genotypes. In the second
609 experiment, the correlation between the mixture performances and the means of the pure
610 stands components was rather high for yield (0.75) and for protein content (0.63) so the
611 predictions of all the observations based on the pure stand performances were efficient and
612 were only marginally improved when using mixing ability modeling. This is explained by a
613 very high correlation between the GMA and the genotypic effects of the pure stands for yield
614 in this trial (0.98), for reasons discussed hereafter (section 4.3).

615

616 *4.2 An original modeling of specific mixing ability*

617 Specific Mixing Ability was modeled using two different approaches that could better reflect
618 the underlying plant-plant interactions:

619 - in addition to inter-genotypic interactions within mixtures, we introduced intra-genotypic
620 interactions within pure stands (Model 2). Note that in the literature, pure stands are not
621 always considered as particular cases of mixtures and therefore are usually not included as
622 such in the analysis;

623 - we further refined the SMA estimates, by accounting for intra-genotypic interactions not only
624 within pure stands but also within mixtures. For that, SMA effects were weighted by the
625 probabilities of neighborhood of varieties in a mixture (Model 3).

626 These two models resulted in differences in variance components and in ratio of variance
627 between SMA and GMA, with a higher SMA variance observed with Model 3, as expected.
628 The intra-genotypic SMA of yield had noticeably more extreme values in the SMA distribution
629 when using Model 3 compared to Model 2, suggesting that the intra-genotypic interactions in
630 mixtures significantly accounted for SMA variance in Model 3 for this trait. The interpretation

631 of the intra-genotypic SMA should be different depending on the model used. Model 3 might
632 be more realistic regarding the biological mechanisms occurring within mixtures. For instance
633 it is expected that the intra-genotypic interactions should be lower in higher order mixtures.
634 Intra-genotypic interactions SMA_{ii} are by nature quite different from the one of inter-genotypic
635 interactions SMA_{ij} . Consequently one should estimate separately one variance component for
636 SMA_{ii} and one for SMA_{ij} (with $i \neq j$) for both Models 2 and 3. However this would require the
637 inference of more variance parameters - each with less available information - which in turn
638 would require the evaluation of more mixtures per genotype than available in this study.

639 If the independence between the performance of pure lines and their hybrids (e.g. design 4 in
640 Griffing 1956) can be explained by genetic interactions (i.e. dominance), it is biologically
641 difficult to neglect intra-genotypic interactions in mixtures. However, if the focus is to detect
642 genotypes that are particularly good partners for complex mixtures, it might be more relevant
643 to remove pure stands. Running the analysis without the pure stands gave lower GMA
644 predictions for the high yielding elite varieties (e.g. Trémie, Boregar ... Figure B.3) while
645 different varieties such as Soissons showed higher GMA, therefore appearing as particular
646 good mixing partners. In that case, no SMA effect was detected for yield, while it was still
647 present for spike density ($p=0.06$) and TKW ($p=0.009$) (data not shown). Another approach
648 has been proposed in the literature to account for the inter- and intra-genotypic interactions
649 within mixtures. It is based on the partitioning of the GMA into a GPA (General Performing
650 Ability) reflecting the genotypic effect in pure stand, and a TGMA (True General Mixing
651 Ability) being the part of the GMA truly due to the mixing conditions obtained when analyzing
652 the over-yielding (OY, i.e. the difference between the yield of the mixture and the mean of its
653 components in pure stands) as a response variable (Federer *et al.* 1982; Gizlice *et al.* 1989;
654 Knott and Mundt 1990; Lopez and Mundt 2000). In these studies, both GPA and TGMA have
655 been found important for mixing ability. Here, we chose to include the pure stands in the
656 observations for modeling mixing ability, since SMA of the pure stands (SMA_{ii}) in Model 2
657 provides sufficient information on how each genotype performs in pure stands in comparison
658 with its GMA in the experiment. Both approaches are strongly divergent in their philosophy: in

659 the GPA/TGMA approach, pure stands and mixtures performances are used independently
660 to estimate GPA and GMA, and the pure stand performances are considered as a reference
661 to compute the TGMA. In the present model, pure stands and mixtures are jointly used to
662 estimate GMA and SMA, and the intra-genotypic interaction (SMA_{ij}) has the same status as
663 the inter-genotypic interaction (SMA_{ij}), which seems biologically sound.

664

665 *4.3 Variance components vary between the two studied locations*

666 For all response variables in the binary mixtures trial (LM), the variance of GMA was higher
667 than the one of SMA, in accordance with Knott and Mundt (1990) in similar trials, while no
668 SMA effect at all was found in the experiment involving higher order mixtures, as in Lopez
669 and Mundt (2000). However, only GMA/SMA estimates for yield can be compared among the
670 two experiments since the other response variables differed among LM and CF. The absence
671 of SMA in the CF experiment might be due to three possible factors which are confounded in
672 this experimental design, and might contribute to limit the inter-genotypic interactions in CF.
673 Firstly, the CF trial was conducted under more favorable conditions (higher yield objective
674 and nitrogen inputs) that might have led to less stresses for the plants and therefore less
675 opportunity for complementarity or synergy mechanisms to express among mixtures'
676 components. Secondly, the panel used in the CF experiment was a subset of the panel used
677 for the Le Moulon experiment, the genotypes were less contrasted than in the LM trial (only
678 one landrace included in the panel, lower diversity in functional traits) resulting in lower
679 competition/synergies. Thirdly, the CF experiment involved higher order mixtures in which
680 SMA within each pair of genotypes might have been too low to be detected (increasing the
681 number of binary interactions results in a dilution of their effects), or the possible occurrence
682 of higher order interactions might have masked the binary ones. Thus, this could lead to
683 reducing the possibility to observe SMA effects in the CF trial compared to the LM trial.

684 For yield in the LM trial, the share of the genetic variance due to SMA effect was larger with
685 Model 3 than with Model 2, indicating that the SMA variance might also be due to intra-

686 genotypic interactions within mixtures. However, both models including SMA did not provide
687 a significantly better fit than the model with GMA only (and the differences in AIC values for
688 the three models were very small for all response variables (Table 1)), as already found by
689 Gizlice *et al.* (1989). In contrast, Gallandt *et al.* (2001) in wheat or Federer *et al.* (1982) in
690 bean found significant SMA. Spike density displayed a SMA variance ratio of 0.61 with Model
691 3, indicating strong interactions between plants for this response variable. This is consistent
692 with the fact that the number of tillers and therefore the number of spikes are known to be the
693 most plastic traits when changing plant density (Darwinkel 1978) and/or the phenotypes of
694 neighboring plants in mixtures (Jackson and Wennig 1997; Cowger and Weisz 2008,
695 Lecarpentier 2017, Lecarpentier *et al.* in press). This is also in line with the clear-cut
696 difference observed in the range of spike density in the LM trial when comparing pure stands
697 and binary mixtures (respectively 217.5 to 490 for pure stands and 262.5 to 555 for mixtures).
698 These specific effects observed on mixtures occurred in the experiment characterized by a
699 low density leading to plasticity for tillering, and selection effects between genotypes due to
700 differences in potential productivity. As this leads to differences in genotype frequencies in
701 the harvested grain, it might be interesting to measure the proportion of grains (or spikes) of
702 each genotype in mixtures to further investigate the specific effects and better predict the
703 mixture performances based on individual contributions of the components. The estimation of
704 selection effects may help to better understand the changes in proportions of the genotypes
705 over time (Finckh and Mundt 1992; Loreau and Hector 2001; Barot *et al.* 2017) but this
706 requires particular designs or systems for separating the mixture components.

707

708 *4.4 BLUP*

709 In the LM experiment, the Trémie genotype showed the highest BLUP value for both GMA
710 and SMA_{ii} due to its high productivity in particular in pure stand. However, the binary mixtures
711 involving this genotype had among the lowest SMAs and all had lower yields than expected
712 based on the pure stands components (i.e. negative OY), therefore indicating that Trémie, as

713 an elite variety selected for performing well in pure stand conditions, might not be the best
714 “mixing partner” to combine. On the contrary, the genotype Soissons had intermediate GMA
715 (in both trials) and a low SMA in pure stand, while its SMAs in mixtures were high and the
716 mixtures displayed high yields (Soissons also had the highest GMA when considering
717 mixtures only). This could be due to a favorable mixture design for this genotype, but the OY
718 was always positive for this genotype (except when mixed with Trémie) with a mean of
719 $+7.7\text{q}\cdot\text{ha}^{-1}$ showing its potential for mixing conditions. The important mixing ability for yield of
720 this genotype might be explained by a high mixing ability for spike density. These results
721 show the importance of taking both GMA and SMA into account for selecting candidates for
722 mixtures.

723

724 *4.5 Prediction of mixtures and pure stands performances based on BLUP values*

725 Prediction of extreme observations was less accurate for response variables for which the
726 correlation between mixtures and corresponding means of the pure stand means was lower.
727 This might be because the GMA is an average effect leading to a shrinkage of the predicted
728 values towards the mean of the observed mixtures performances involving the given
729 genotype (which is partly corrected when including SMA in the models). It should also be
730 noted that the GMA of the genotypes are relative to the assessed panel. This is particularly
731 important since the GMA are computed using a common measure for all the components of a
732 given mixture. In this regard, Federer *et al.* (1982) proposed an alternate model to analyze
733 mixtures when individual component performances are available. Using this alternate model
734 should lead to a reduction of the shrinkage effect since the GMA of a genotype is not
735 dependent on the performances of its associated genotypes. Finally, the mixture
736 performances are predicted assuming that the proportions of the components are the same
737 as those used for sowing.

738

739 *4.6 Predictions of higher order mixtures from binary mixtures evaluated in another*
740 *experiment*

741 The GMA and SMA have been found quite variable over environments in other studies (Knott
742 and Mundt 1990; Gallandt *et al.* 2001) but interestingly the GMAs predicted for yield by the
743 experiments at Le Moulon and Clermont-Ferrand proved to be highly correlated (0.903)
744 although one site had higher order mixtures, and the Clermont-Ferrand panel was only a
745 subset of the Le Moulon panel. We assessed the ability of the pure stands and the binary
746 mixtures of the LM experiment to predict the CF experiment mixtures and especially the
747 higher order mixtures. The correlation between predicted and observed values was found
748 decreasing with increasing orders mixtures (but for binary mixtures, it should be borne in
749 mind that all binary mixtures observed in the CF trial were also observed in the LM trial).
750 However, predictions based on binary mixtures were better than based on pure stands, as
751 previously observed in literature (Lopez and Mundt 2000; Mille *et al.* 2006). The Kendall's
752 correlation coefficient was also calculated since the concordance in the ranking of mixture
753 performances is desirable for breeding perspectives. This correlation was at least equal when
754 the four-way and the eight-way mixtures were predicted from the LM observations without the
755 pure stands (m) than when they were included ($m+p$), suggesting that the exclusion of pure
756 stands from the GMA predictions does not degrade the predictions of high order mixtures. A
757 higher correlation between observed and predicted values was expected when using all
758 observations ($m+p$) than when using mixtures only (m), but the aim was to assess the impact
759 of the pure stand information on the correlation.

760

761 *4.7 Improving screening and breeding strategies*

762 The two mixture designs used in this study were performed empirically, as in all agronomical
763 and ecological experiments we have reviewed. This raises the question of how to optimize
764 the experimental designs used to better estimate GMA/SMA. As far as we know, this
765 question has been poorly addressed in the literature. For example Federer (2002) and

766 Raghavarao and Federer (2003) have proposed balanced incomplete designs to achieve
767 estimability of the GMA and SMA for a specific number of varieties and three-way mixtures,
768 without giving rules on the way varieties should be mixed. The experimental design
769 optimization remains to be studied. The BLUP framework proposed here will allow such
770 optimization, exploring various GMA/SMA ratios, according to panel/species biological
771 characteristics.

772 Different designs should be chosen for the estimation of mixing ability, depending on the
773 objective of the experiment (i.e. assessment of GMA-SMA for a panel, improvement of mixing
774 ability in breeding scheme) and the structuration of GMA and SMA effects. Once a design
775 has been chosen for estimating mixing ability, it can be optimized using power tests to
776 maximize the accuracy of the GMA and SMA predictions, whether or not using higher order
777 mixtures instead or in addition to binary mixtures, whether or not including the pure stands,
778 varying the proportion of possible mixtures to be observed, the number of observations per
779 genotype, the number of replicates per mixture, the mixture composition, ... and taking into
780 account the gain in accuracy regarding the costs and efforts invested in the experimental
781 design. The comparison of the power of different types of incomplete designs (nested design,
782 balanced incomplete factorial design, topcross design, random design as used by Zhao *et al.*
783 (2015) for combining ability analysis) may allow to identify the most adapted and powerful
784 type of design regarding a particular situation, and to tailor it into an optimized design for
785 mixing ability estimation. Since mixing ability might vary across environments and might be
786 subject to interactions with environment, the decision to replicate the experimental design in
787 different environments would depend on the objectives, resources and potential GxE
788 interactions.

789 Parameters to consider for assessing mixing ability are the correlation between the
790 performances of mixtures and predictions from their pure stand components, and the
791 importance of the GMA vs SMA effects. Therefore, in a breeding scheme, a preliminary
792 experiment must be carried out in order to characterize the panel first and thus guide the

793 breeding strategy and objectives. This study exemplifies the kind of preliminary study that
794 can be developed to assess variance components. Three cases are expected: i) high
795 correlation between the mixture performances and the mean of their pure stand components,
796 ii) important GMA versus SMA, iii) important SMA.

797 In the case of a strong correlation between mixtures and the mean of their pure stand
798 components, information on the performances of pure stands can be used since it requires
799 less observation. In contrast, if the GMA is moderately or weakly correlated to the genotypic
800 effects in pure stands or if SMA effects are detected, it would be necessary to include
801 mixtures in the design, and consider excluding the pure stands in order to limit the number of
802 plots especially for the development of high order mixtures. In addition to this advantage, in
803 the LM trial, the exclusion of pure stands for predicting GMA and SMA effects led to a slight
804 improvement in correlations between observed and predicted values for yield, spike density
805 and TKW, while the SMA was reduced to zero for the yield.

806 When the GMA is the major source of variation, compared to the SMA, it could be wiser to
807 focus on the components *per se* for mixing rather than on combinations of components. The
808 experimental design should be based on genotype screening under conditions that allow
809 interactions between different genotypes such as alternate rows (as proposed by Barot *et al.*
810 2017) or mixtures with a tester (that can be a genotype with high GMA or eventually a mixture
811 itself to save space and if the further purpose is the development of higher order mixture) for
812 instance. If the SMA is low but not negligible, it might be interesting to search for panel
813 structure to optimize the experimental design.

814 When SMA effects are important, the existence of complementarity groups for mixing ability
815 should be investigated. As carried out for hybrid development, the use of clustering methods
816 to search for heterotic-like pattern might allow to capture some of the specific effects in the
817 GMA (increasing the accuracy of mixture performance predictions), and to further improve
818 the complementarity between components within mixtures structuring the panel into groups
819 (Zhao *et al.* 2015). For the development of higher order mixtures, the presence of higher

820 order specific effects might be investigated (Federer 1999) to decide if the assessment of
821 binary mixture is sufficient or if higher order mixtures should be included in the design. If each
822 genotype is observed in several mixtures, the stability of the SMA effects involving the
823 genotype might be assessed, allowing to identify the most stable genotypes for specific
824 effects; and stability of SMA over environments might be investigated.

825 The mixing ability structuration between GMA and SMA has consequences on the conception
826 of the breeding strategy. When the mixture performances are highly predictable based on
827 pure stands (first case), any particular adaptation of the breeding scheme is required for
828 selecting the best performing genotypes, other than assuring that the final mixtures are
829 agronomically coherent. When the selection focuses on GMA (second case), genotypes
830 should be assessed and selected in interaction with each other or with testers. Interaction
831 traits (e.g. the number of spikes per plant) and the potential plasticity for these traits can be
832 integrated into the screening or used as selection criteria. When SMA effects are important
833 (third case), it is required to assess combinations of genotypes, but it is also desired to
834 reduce the number of mixtures to evaluate. Prediction of mixture performances may then rely
835 on both genotypic and phenotypic information (kinship, molecular markers, interaction traits).

836 The inclusion of genetic relatedness matrices as covariance matrices for the GMA and the
837 SMA random effects in the mixed models may improve the prediction of the mixing effects,
838 while allowing for predicting unobserved mixtures or mixtures involving unobserved
839 genotypes (as for combining ability analysis and predictions of performances of hybrids,
840 Bernardo 1995; Falconer *et al.* 1996; Lynch and Walsh 1998). Phenotypic similarities
841 assessed on pure stand performances and traits involved in plant-plant interactions could
842 also be used as covariance matrices. On a more general level, the identification of traits
843 associated to mixing ability could be relevant for integrating new selection criteria in breeding
844 for mixtures, but should also help to better understand the functioning of the mixtures (also
845 regarding traits complementarity between genotypes or species) and to improve the design of
846 the mixtures (Barot *et al.* 2017).

847 Note that this methodological framework can be easily further adapted to mixtures with
848 varying proportions of the different components (by adjusting the weighting coefficients of
849 GMA and SMA ($\frac{1}{K(n)}$, $\frac{1}{\binom{K(n)}{2}}$ and $\frac{1}{(K(n))^2}$, in Models 1, 2 and 3) based on their proportion
850 of sowed seeds and neighboring probabilities), and notably to inter-specific mixtures
851 analysis.

852

853

854 **ACKNOWLEDGMENTS**

855 E. Forst benefited from a thesis grant from a PICRI (Partenariat Institutions-Citoyens pour la
856 Recherche et l'Innovation) project (2015-2017) funded by the Région Ile-de-France, and her
857 post-doctoral salary was paid by the European Union's Horizon 2020 research and
858 innovation programme for the LIVESEED project (<https://www.liveseed.eu/>), under grant
859 agreement number 727230 and for the ReMIX project (<https://www.remix-intercrops.eu/>)
860 under grant agreement number 727217. This work was also supported by the ANR
861 WHEATAMIX project (www6.inra.fr/wheatamix/), grant ANR-13-AGRO-0008 of the French
862 National Research Agency, for field trials. The authors are grateful to the Wheatamix
863 consortium for early discussions on the approach and the design of the experiments, and to
864 Didier Tropée (and the technical team IE GQE-Le Moulon) and Alain Chassin (and all the
865 technical staff of UE 1375 PHACC) for conducting the field trials. We would like to thank
866 Fabien Laporte for his valuable contribution to the early developments of the BLUP
867 programing. We also thank the reviewers for their insightful comments that led us to improve
868 the paper and to add an appendix on mixed effect model development.

869

870

871 Declaration of conflicts of interest: none.

872

873

874 **REFERENCES**

875 Alday, P., 2016 A brief introduction to mixed effects models - analogues to the coefficient of
876 determination R^2

877 <https://www.biorxiv.org/content/biorxiv/suppl/2016/07/06/062299.DC1/062299-1.pdf>

878 Barot, S., V. Allard, A. Cantarel, J. Enjalbert, A. Gauffreteau, I. Goldringer, J-C. Lata, X.

879 Leroux, A. Niboyet, E. Porcher, 2017 Designing mixtures of varieties for

880 multifunctional agriculture with the help of ecology, a review. *Agron. Sustain. Dev.* 37

881 (2): 13. <https://doi.org/10.1007/s13593-017-0418-x>

882 Bates, D., M. Mächler, B. Bolker, S. Walker, 2015 Fitting linear mixed-effects models using

883 lme4. *Journal of Statistical Software*, 67 (1). <https://doi.org/10.18637/jss.v067.i01>.

884 Bernardo, R. 1995 Genetic Models for Predicting Maize Single-Cross Performance in

885 Unbalanced Yield Trial Data. *Crop Science* 35 (1): 141-47.

886 <https://doi.org/10.2135/cropsci1995.0011183X003500010026x>.

887 Bernardo, R., 1996 Best linear unbiased prediction of maize single-cross performance. *Crop*

888 *Sci*, 36 (1): 50-56. doi:10.2135/cropsci1996.0011183X003600010009x

889 Borg, J., L. P. Kiær, C. Lecarpentier, I. Goldringer, A. Gauffreteau, S. Saint-Jean, S. Barot, J.

890 Enjalbert, 2018 Unfolding the potential of wheat cultivar mixtures A meta-analysis

891 perspective and identification of knowledge gaps. *Field Crop Res.* 221, 298-313.

892 <https://doi.org/10.1016/j.fcr.2017.09.006>

893 Chalbi, N., 1967 La compétition entre géotypes et ses effets sur les caractères quantitatifs

894 de la Luzerne. *Annales d'Amélioration des Plantes* 17 (1): 67-82.

895 Cowger, C., R. Weisz., 2008 Winter wheat blends (mixtures) produce a yield advantage in

896 North Carolina. *Agron. J.* 100 (1): 169-77. <https://doi.org/10.2134/agronj2007.0128>.

897 Creissen, H. E., T. H. Jorgensen, J. K. M. Brown, 2016 Increased yield stability of field-grown
898 winter barley (*Hordeum Vulgare* L.) varietal mixtures through ecological processes.
899 Crop Prot. 85: 1-8. <https://doi.org/10.1016/j.cropro.2016.03.001>

900 Darwinkel, A., 1978 Patterns of tillering and grain production of winter wheat at a wide range
901 of plant densities. Neth. J. Agric. Sci., 26, 383-98.

902 Dawson, J. C., I. Goldringer, 2012 Breeding for genetically diverse populations: variety
903 mixtures and evolutionary populations. Organic Crop Breeding, 77-98.
904 <https://doi.org/10.1002/9781119945932.ch5>

905 Dubs, F., X. Leroux, V. Allard, B. Andrieu, S. Barot, et al., 2018 An experimental design to
906 test the effect of wheat variety mixtures on biodiversity and ecosystem services, HAL.
907 . <https://hal.archives-ouvertes.fr/hal-01843564>

908 Falconer, D. S., T. F. C. Mackay, R. Frankham, 1996 Introduction to Quantitative Genetics.
909 4th ed. Benjamin Cummings. Trends in Genetics, 12 (7), 280.

910 Federer, W. T., 1979 Statistical designs and response models for mixtures of cultivars.
911 Agron. J. 71 (5): 701. <https://doi.org/10.2134/agronj1979.00021962007100050003x>.

912 Federer, W. T., J. C. Connigale, J. N. Rutger, A. Wijesinha, 1982 Statistical analyses of yields
913 from uniblends and biblends of eight dry bean cultivars. Crop Sci. 22 (1): 111.
914 <https://doi.org/10.2135/cropsci1982.0011183X002200010026x>.

915 Federer, W. T., 1999 Statistical Design and Analysis for Intercropping Experiments. Volume
916 II. Three or More Crops. Springer, New York.

917 Federer, W. T., 2002 Statistical issues in intercropping. Encyclopedia of environmetrics. 1st
918 edition, New York: Wiley, 1064-1069.
919 <https://doi.org/10.1002/9781118445112.stat07718>

920 Finckh, M. R., C. C. Mundt, 1992 Stripe rust, yield and plant competition in wheat cultivar
921 mixtures. Phytopathology 82 (9): 905-13. DOI: 10.1094/Phyto-82-905

922 Gallais, A., 1970 Modèle pour l'analyse des relations d'associations binaires. Biométrie-
923 Praximétrie XI (2-3): 51-80.

924 Gallandt, E. R., S. M. Dofing, P. E. Reisenauer, E. Donaldson, 2001 Diallel analysis of
925 cultivar mixtures in winter wheat. *Crop Sci.* 41 (3): 792-796.
926 <https://doi.org/10.2135/cropsci2001.413792x>.

927 Gizlice, Z., T. E. Carter, J. W. Burton, T. H. Emigh, 1989 Partitioning of blending ability using
928 two-way blends and component lines of soybean. *Crop Sci.* 29 (4): 885.
929 <https://doi.org/10.2135/cropsci1989.0011183X002900040008x>.

930 Griffing, B., 1956 Concept of General and Specific Combining Ability in relation to diallel
931 crossing systems. *Aust. J. Biol. Sci.* 9: 463-93. <https://doi.org/10.1071/BI9560463>

932 Jackson, L. F., R. W. Wennig, 1997 Use of wheat cultivar blends to improve grain yield and
933 quality and reduce disease and lodging. *Field Crops Res.* 52 (3): 261-269.
934 [https://doi.org/10.1016/S0378-4290\(97\)00007-5](https://doi.org/10.1016/S0378-4290(97)00007-5).

935 Jensen, S. D. 1959 Combining Ability of Unselected Inbred Lines of Corn from Incomplete
936 Diallel and Top-Cross Tests. PhD thesis, Iowa State University.
937 <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=3159&context=rtd>.

938 Jensen, N. I., W. T. Federer 1965 Competing ability in wheat. *Crop Sci* 5 (5): 449-52.

939 Kaut, A. H. E. E., H. E. Mason, A. Navabi, J. T. O'Donovan, D. Spaner, 2009 Performance
940 and stability of performance of spring wheat variety mixtures in organic and
941 conventional management systems in western Canada. *J. Agron. Crop Sci.* 147 (2):
942 141-53. <https://doi.org/10.1017/S0021859608008319>

943 Kiær, L. P., I. M. Skovgaard, H. Østergård, 2009 Grain yield increase in cereal variety
944 mixtures: a meta-analysis of field trials. *Field Crops Res.* 114 (3): 361-373.
945 <https://doi.org/10.1016/j.fcr.2009.09.006>.

946 Kiær, L. P., I. M. Skovgaard, H. Østergård, 2012 Effects of inter-varietal diversity, biotic
947 stresses and environmental productivity on grain yield of spring barley variety
948 mixtures. *Euphytica*, 185(1): 123-138. <https://doi.org/10.1007/s10681-012-0640-1>

949 Knott, E. A., C. C. Mundt, 1990 Mixing ability analysis of wheat cultivar mixtures under
950 diseased and nondiseased conditions. *Theor. Appl. Genet.* 80 (3): 313-20.
951 <https://doi.org/10.1007/BF00210065>.

952 Lazzaro, M., A. Costanzo, P. Bàrberi. 2018 Single vs multiple agroecosystem services
953 provided by common wheat cultivar mixtures: weed suppression, grain yield and
954 quality. *Field Crops Res.*, 221 (May): 277-97. <https://doi.org/10.1016/j.fcr.2017.10.006>

955 Lecarpentier, C., 2017 Plasticité de l'architecture aérienne du blé en réponse à la compétition
956 pour la lumière au sein de cultures pures ou d'associations variétales :
957 Caractérisation expérimentale et développement d'un modèle 3D. PhD thesis. Ecole
958 doctorale Sciences du végétal, Université Paris-Saclay.
959 <http://www.sudoc.fr/219507198>

960 Lecapentier, C., R. Barillot, E. Blanc, M. Abichou, I. Goldringer, P. Barbillon, J. Enjalbert, B.
961 Andrieu. WALTER: A three-dimensional wheat model to study competition for light
962 through the prediction of tillering dynamics. *Annals of Botany*. *in press*.
963 <https://doi.org/10.1093/aob/mcy226>

964 Lopez, C. G., C. C. Mundt, 2000 Using mixing ability analysis from two-way cultivar mixtures
965 to predict the performance of cultivars in complex mixtures. *Field Crops Res.* 68 (2):
966 121-32. [https://doi.org/10.1016/S0378-4290\(00\)00114-3](https://doi.org/10.1016/S0378-4290(00)00114-3).

967 Loreau, M., A. Hector, 2001 Partitioning selection and complementarity in biodiversity
968 experiments. *Nature* 412 (6842): 72-76. <https://doi.org/10.1038/35083573>.

969 Lynch, M., B. Walsh, 1998 *Genetics and Analysis of Quantitative Traits*. Vol. 1. M. A.
970 Sunderland, Sinauer.

971 McGilchrist, C. A., 1965 Analysis of competition experiments. *Biometrics*, 975-985. DOI:
972 10.2307/2528258

973 Mille, B., M. Belhaj Fraj, H. Monod, C. de Vallavieille-Pope, 2006 Assessing four-way
974 mixtures of winter wheat cultivars from the performances of their two-way and
975 individual components. *Eur J Plant Pathol.* 114 (2): 163-73.
976 <https://doi.org/10.1007/s10658-005-4036-0>.

977 Möhring, J., A.E. Melchinger, H.-P. Piepho, 2011 REML-based diallel analysis. *Crop Sci* 51,
978 470-478. doi:10.2135/cropsci2010.05.0272

979 Nakagawa, S. H. Schielzeth, 2013 A general and simple method for obtaining R^2 from
980 generalized linear mixed-effects models. *Methods Ecol Evol*, 4: 133-142.
981 doi:10.1111/j.2041-210x.2012.00261.x

982 Newton, A. C., R. P. Ellis, C. A. Hackett, D. C. Guy, 1997 The Effect of component number
983 on *Rhynchosporium secalis* infection and yield in mixtures of winter barley cultivars.
984 *Plant Pathology* 46 (6): 930-38. <https://doi.org/10.1046/j.1365-3059.1997.d01-83.x>.

985 Newton, A. C., G. S. Begg, J. S. Swanston, 2009 Deployment of diversity for enhanced crop
986 function. *Annals of Applied Biology*, 154(3): 309-322. <https://doi.org/10.1111/j.1744-7348.2008.00303.x>

987

988 Østergård, H., K., Kristensen, J. W., Jensen, 2005 Stability of variety mixtures of spring
989 barley, pp. 28-30 in *Proceedings of the COST SUSVAR/ECO-PB Workshop on*
990 *Organic Plant Breeding Strategies and the Use of Molecular Markers*, edited by E. T.
991 Lammerts van Bueren, I. Goldringer and H. Østergård. Driebergen, The Netherlands.
992 <http://orgprints.org/8152/>.

993 Piepho, H.-P., J. Möhring, 2007 Computing heritability and selection response from
994 unbalanced plant breeding trials. *Genetics* 177 (3): 1881-88.
995 <https://doi.org/10.1534/genetics.107.074229>

996 Raghavarao, D., W. T. Federer, 2003 Sufficient conditions for balanced incomplete block
997 designs to be minimal fractional combinatorial treatment designs, *Biometrika*, 90 (2):
998 465-470. <https://doi.org/10.1093/biomet/90.2.465>

999 Reiss, E. R., L. E. Drinkwater, 2018 Cultivar mixtures: A meta-analysis of the effect of
1000 intraspecific diversity on crop yield. *Ecol Appl*. 28: 62-77
1001 <https://doi.org/10.1002/eap.1629>

1002 Sprague, G. F., L. A. Tatum, 1942 General vs. Specific Combining Ability in single crosses of
1003 corn. *J. Amer. Soc. Agron.* 34 (10): 923-32.

1004 Stützel, H., W. Aufhammer, 1990 The physiological causes of mixing effects in cultivar
1005 mixtures - a general hypothesis. *Agric. Sys.* 32 (1): 41-53.
1006 [https://doi.org/10.1016/0308-521X\(90\)90029-P](https://doi.org/10.1016/0308-521X(90)90029-P).

1007 Thépot, S., G. Restoux, I. Goldringer, F. Hospital, D. Gouache, I. Mackay, J. Enjalbert, 2015
1008 Efficiently tracking selection in a multiparental population: the case of earliness in
1009 wheat. *Genetics* 199 (2): 609-23. <https://doi.org/10.1534/genetics.114.169995>.

1010 Tilman, D., J. Knops, D. Wedin, P. Reich, M. Ritchie, E. Siemann, 1997 The influence of
1011 functional diversity and composition on ecosystem processes. *Science* 277 (5330):
1012 1300-1302. <https://doi.org/10.1126/science.277.5330.1300>.

1013 Vallavieille-Pope, C. de., 2004 Management of disease resistance diversity of cultivars of a
1014 species in single fields: controlling epidemics. *Comptes Rendus Biologies* 327 (7):
1015 611-620. <https://doi.org/10.1016/j.crv.2003.11.014>.

1016 Weisser, W. W., C. Roscher, S. T. Meyer, A. Ebeling, G. Luo, et al. 2017 Biodiversity effects
1017 on ecosystem functioning in a 15-year grassland experiment: patterns, mechanisms,
1018 and open questions. *Basic Applied Ecology* 23: 1-73.
1019 <https://doi.org/10.1016/j.baae.2017.06.002>

1020 Wolfe, M. S., 1985 The current status and prospects of multiline cultivars and variety mixtures
1021 for disease resistance. *Annu. Rev. Phytopathol.* 23 (1): 251-273.
1022 <https://doi.org/10.1146/annurev.py.23.090185.001343>

1023 Wu, H. X., A. C. Matheson, 2000 Analysis of half-diallel mating design with missing crosses:
1024 theory and SAS program for testing and estimating GCA and SCA fixed effects. *Silvae*
1025 *genetica*, 49 (3), 130-136.

1026 Xu, R., 2003 Measuring explained variation in linear mixed effects models. *Statist. Med.*, 22:
1027 3527-3541. doi:10.1002/sim.1572

1028 Zhao, Y., Z. Li, G. Liu, Y. Jiang, H. P. Maurer, T. Würschum, H.-P. Mock, A. Matros, E.
1029 Ebmeyer, R. Schachschneider, E. Kazman, J. Schacht, M. Gowda, C. F. H. Longin, J.
1030 C. Reif., 2015 Genome-based establishment of a high-yielding heterotic pattern for
1031 hybrid wheat breeding. *Proc. Natl. Acad. Sci.* 112 (51): 15624-29.
1032 <https://doi.org/10.1073/pnas.1514547112>.

1033 Zhu, Y., H. Chen, J. Fan, Y. Wang, Y. Li, J. Chen, J.X. Fan, S. Yang, L. Hu, H. Leung, T. W.
1034 Mew, P. S. Teng, Z. Wang, C. C. Mundt, 2000 Genetic diversity and disease control in
1035 rice. *Nature* 406 (6797): 718-22. <https://doi.org/10.1038/35021046>.
1036
1037