Selection for high spike fertility index in early generations increases grain number per unit area (and yield) of advanced breeding lines in bread wheat

MP Alonso[1,2]; MF Franco[1,2]; AC Pontaroli[1,2*].

[1] Unidad Integrada Balcarce [Facultad de Ciencias Agrarias, Universidad Nacional de Mar del Plata – Estación Experimental Agropecuaria Balcarce, Instituto Nacional de Tecnología Agropecuaria (INTA)]; CC 276 (7620) Balcarce, Argentina.

[2] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Argentina.

[*] Correspondence: Ana Clara Pontaroli, Email: pontaroli.ana@inta.gob.ar; phone: +54-2266-43-9013 (ext. 333/567).

ABSTRACT

Spike fertility index (a.k.a. fruiting efficiency) at maturity (FE_m) has been proposed as a promising selection criterion for improving grain yield (GY) in bread wheat. However, its usefulness as an early-generation selection criterion aiming at increasing GY in advanced generations remains to be established. On the other hand, there is no information about the feasibility of using spike harvest index (SHI) as an indirect selection criterion for increasing GY. In this study, a biparental population derived from a cross between 'Baguette 10' and 'Klein Chaja´', Argentinian spring bread wheat cultivars with contrasting FE_m, was evaluated from early generations (F_2 and F_3) to recombinant inbred lines (RILs, tested in seven experiments). Associations of FE_m and SHI measurements between early generations and RILs were established. Also, responses in grain number/m^2 (GN) and GY in the RILs after simulated selection for high FE_m or high SHI in early generations under four selection strategies comprising combinations of different selection intensities (50%, 25% and 10%) were assessed. Significant and positive correlation was found between FE_m measured in early generations and in the RILs ($r=0.32$, $P<0.05$) and SHI measured in the F_3 and in the RILs ($r=0.26$, $P<0.05$). After selection for high FE_m in early generations, responses in GN were positive in all the experiments under all four simulated selection strategies: GN increased by 5.1 to 22.1% as compared with the

population mean, depending on the selection strategy. Responses in GY were positive in three out of four selection strategies, averaging -1.3 to 2.1% depending on the selection strategy. Contrarily, responses to selection in GN and GY in the RILs using SHI as a selection criterion in early generations were negative or neutral. In this way, our results validate the use of FE_m as a selection criterion in early generations for increasing GN and GY in advanced generations of bread wheat breeding programs.

KEYWORDS

Fruiting efficiency at maturity; selection criterion; grain number; breeding; genetic progress

INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one of the most important crops that supply the world demand for food. Under a current and future scenario of increased global demand for grains and limited possibilities of expanding cropping areas, breeding efforts must concentrate on further improving grain yield (GY) (CIMMYT, 2019). One manner to achieve this faster is through the identification of yield-related traits which are amenable to high-throughput phenotyping, and their utilization as selection criteria in breeding programs.

Grain yield in bread wheat can be considered as the product between the number of grains per unit area (GN) and individual grain weight (GW); of these components, the former is the one which best explains GY variations (Fischer, 1985; Abbate et al., 1995; Calderini et al., 1999; González et al., 2014). Therefore, selection for high GN would seem a useful approach which breeders could apply for grain yield improvement. However, one of the difficulties with using variables quantified as per unit area is obtaining reliable data in early generations of breeding programs, in which not enough seed is available.

Fischer (1984) proposed that, under non-limiting growing conditions (*i.e.*, without water or nutrient limitations and in the absence of diseases and pests), GN can be considered as the product of (i) the duration of rapid spike growth period, (ii) crop growth rate during the spike growth period, (iii) dry

weight partitioning to spikes during the spike growth period, and (iv) the number of grains per unit of spike chaff dry weight, i.e., a "spike fertility" index (Abbate et al., 1998) or "fruiting efficiency" (Ferrante et al., 2012). This index was originally determined using the spike chaff dry weight at anthesis; later, Abbate et al. (2013) reported that it had a good correspondence with the index as determined with the spike chaff dry weight at maturity.

The spike fertility index at maturity (hereinafter referred to as "fruiting efficiency at maturity", or FE_m) has been widely proposed as a selection criterion in breeding programs (Slafer et al., 2015; Alonso et al., 2018b; Fischer and Rebetzke, 2018; Gerard et al., 2019). Firstly, FE_m is positively associated with GN (Abbate et al., 1998; Foulkes et al., 2015; Ferrante et al., 2017; Terrile et al., 2017; among others). Secondly, a high throughput, non-destructive method for assessing the trait in small samples at maturity is available (Abbate et al., 2013). In addition, FE_m has been shown to have moderate to high heritability (Martino et al., 2015; Alonso et al., 2018b; Pretini et al., 2020; Pontaroli et al., 2021) as well as substantial genetic variability and low genotype × environment interaction (Martino et al., 2015; Mirabella et al., 2016; Terrile et al., 2017; Alonso et al., 2018b; Pretini et al., 2020; Pontaroli et al., 2021). All these features turn FE_m into a very promising selection criterion in breeding programs aiming at increasing grain yield (Slafer et al., 2015; Elía et al., 2016; Terrile et al., 2017). Indeed, Alonso et al. (2018b) and Pontaroli et al. (2021) evaluated the genetic progress in GN and GY under different selection strategies (i.e., selection for high FE_m and/or GY *per se*) in recombinant inbred lines and advanced breeding lines. Results of these studies, in which both selection and testing were carried out in advanced lines, showed that using FE_m as selection criterion produced similarly positive and more stable responses in GY and GN than when selecting by high GY *per se*. Some trade-offs (namely with GW) were observed, but these could potentially be overcome by concurrent selection for both traits.

The question remains whether FE_m can also be effectively used as a selection criterion in early generations of a breeding program and still generate an increase in GN and, by extension, in GY, in the advanced lines derived from them (Fischer and Rebetzke, 2018). A first important prerequisite for this to be feasible is the existence of a good association between the FE_m measurements taken on

individual spaced plants or rows in early generations (F_2-F_3) and those taken on a random spike sample drawn from yield plots of advanced lines. If this were the case, one could expect an increase in GN of advanced lines after selection for high FE_m in early generations of a breeding program (Pretini et al., 2021).

Quail et al. (1989) found moderate association at one site (r=0.35) between FE_m (they simply called it grains per unit chaff weight) measured in individual F_3 plants and yield measured in F_7/F_8 plots. However, they considered the predictive value of FE_m in F_3 for mean F_7/F_8 yields across all the environments to be rather low (r=0.28). More recently, Martino et al. (2015) studying two segregating populations derived from parental cultivars differing for FE_m reported moderate positive correlations for this trait between F_2 spaced plants and F_3 rows. Moreover, selection by high FE_m in F_2 plants generated overall 28% and 12% increases in FE_m and grain number per spike, respectively, in F_3 rows. Despite these antecedents, the usefulness of FE_m as an early-generation selection criterion for increasing GN (and GY) in advanced generations remains to be established (Fischer and Rebetzke, 2018; Pretini et al., 2021).

In addition to FE_m, Fischer and Rebetzke (2018) proposed to explore the use of "a related and totally ignored partitioning trait, namely grain weight per unit spike weight at maturity, [hereinafter referred to as spike harvest index (SHI), that] could also readily be tested in FE_m studies and would avoid the issue of trade-offs with GW". Another advantage of SHI as compared with FE_m would be a reduction in sample processing time, as it would not be necessary to count the grains. However, there is virtually no information on SHI's mode of inheritance or the feasibility of using it as a selection criterion in breeding programs.

Therefore, the aims of this work were (1) to determine the association between FE_m measurements carried out in early-generation, spaced plants and/or rows and in plots of advanced lines, (2) to ascertain whether selection for high FE_m in early-generation, spaced plants and/or rows increases GN and, by extension, GY in the advanced lines derived from them, and (3) to evaluate the potential of SHI as a selection criterion for indirectly increasing GY.

1. MATERIAL & METHODS

1.1. Plant material

A recombinant inbred line (RIL) population was developed from the cross between two Argentinian cultivars with contrasting FE_m: Baguette 10 (B10; high FE_m) and Klein Chajá (KCJ; low FE_m). The development of the RIL population has been described in Martino et al. (2015), Mirabella et al. (2016) and Alonso et al. (2018b). Briefly, the F_1 resulting from the cross was selfed; approximately 2000 F_2 plants were obtained and 200 of them were randomly picked and harvested individually. The single seed descent method (without selection) was carried out from the F_3 generation onwards until reaching practical homozygosity. As a result, 146 RILs were generated.

1.2. Data generation

Most of the data used in this study have been previously published. The field experiments which data were used here are listed and briefly described in Table 1. All were carried out at the INTA Balcarce Experimental Station (Balcarce, Argentina, 37°45′S; 58°18′W; 130 m a.s.l.) except for one at the Chacra Experimental Miramar, Ministerio de Desarrollo Agrario de la Provincia de Buenos Aires (Miramar, Argentina, 38°10′S, 58°0′W, 45 m a.s.l.) and one at the INTA Marcos Juárez Experimental Station (Marcos Juárez, Argentina, 32°43′S; 62°06′W; 112 m a.s.l.). All experiments were conducted with no nutrient limitations and chemical control of fungal diseases, weeds, and pests. Supplemental irrigation was applied only in three experiments: BFE13, BFE14 and BFE15.

Table 1. Overview of field experiments which data were used in this study: genetic material, experiment name, location, year, experimental design (RCBD=randomized complete block design; ARCBD=augmented randomized complete block design), No. reps. (number of replications), plot characteristics and literature references where data have been previously published and/or where experiments and environmental conditions have been described in detail.

Genetic material	Experiment name	Location	Year	Experimental design	No. reps.	Plot characteristics		Reference
						Area (m^2)	Description	
F_2	BF209	Balcarce	2009	ARCBD	1	0.01	0.2 m x 0.05 m spaced plants	1; 2
F_3	MF310	Miramar	2010	RCBD	2	0.3	1 m-long single row; 0.3 m inter-row spacing	1
RILs	BFE13	Balcarce	2013	RCBD	2	3.5	3 rows (5 m long); 0.2 m inter-row and 0.3 m inter-plot spacing	3; 4
	BFE14	Balcarce	2014	RCBD	2	7	7 rows (5 m long); 0.2 m inter-row and 0.3 m inter-plot spacing	3; 4
	BFE15	Balcarce	2015	RCBD	2	7	7 rows (5 m long); 0.2 m inter-row and 0.3 m inter-plot spacing	3; 4
	BFF16-1	Balcarce	2016	RCBD	2	0.3	1 m-long single row; 0.3 m inter-row spacing	5
	BFF16-2	Balcarce	2016	RCBD	2	0.3	1 m-long single row; 0.3 m inter-row spacing	5
	BFF17	Balcarce	2017	RCBD	2	0.3	1 m-long single row; 0.3 m inter-row spacing	5
	MJFE17	M. Juárez	2017	ARCBD	1	0.9	3 rows (1 m long); 0.3 m inter-row and 0.3 m inter-plot spacing	4

References: 1: Martino et al. (2015); 2: Mirabella et al. (2016); 3: Alonso et al. (2018b); 4: Pontaroli et al. (2021);

5: Franco et al. (2021) (only description of crop husbandry is included in Franco et al. (2021), as FE_m and SHI data are shown in the present study for the first time).

1.3. Measurements and calculations

1.3.1. Early generations

Two hundred F_2 plants (experiment BF209) and 200 F_3 rows (experiment MF310) were harvested at physiological maturity. All spikes per F_2 plant or F_3 row were cut at the base of the rachis, counted, weighed, and threshed. Grains were weighed and a grain sample was counted with an electronic counter. For each sample, grain weight per spike was calculated as the quotient between total grain weight and the number of spikes; similarly, grain number per spike was calculated as the quotient between grain number and the number of spikes. Spike chaff weight was calculated as the difference between total weight of the sample (before threshing) and grain weight, divided by the number of spikes. Then, FE_m was calculated as the quotient between grain number per spike and spike chaff weight (Abbate et al., 2013). Spike harvest index (SHI) was calculated as the quotient between grain weight per spike and spike chaff weight.

1.3.2. *Recombinant inbred lines*

One hundred and forty-six RILs were evaluated in seven experiments carried out between 2013 and 2017 (Table 1). At physiological maturity, a random 20-spike sample was taken from the harvestable area of each plot. Spikes were cut at the base of the rachis, weighted and threshed. Grains per sample were counted in an electronic counter and weighed. Spike chaff weight was calculated as the difference between total weight and grain weight. Fruiting efficiency at maturity was calculated as the quotient between total grain number and spike chaff weight (Alonso et al., 2018b; Pontaroli et al., 2021). Spike harvest index was calculated as the quotient between total grain weight and spike chaff weight. Grain yield was determined in experiments BFE13, BFE14 and BFE15 by mechanical harvest of the whole plot (when three-row plots were used) or its five central rows (when seven-row plots were used; Table 1). Grain moisture was obtained from a grain subsample of each plot to calculate dry grain yield (GY). Grain weight (GW) was determined by counting a 1000-grain sample with an electronic counter, weighing it and dividing total weight by 1000. Grain number per m^2 (GN) was then calculated as the quotient between GY and GW × 1000.

1.4. Evaluation of FE_m as a selection criterion in early generations

1.4.1. Correlation analysis

The association of FE_m measurements taken in early generation, spaced plants and/or rows *versus* in RILs was established by Pearson correlation analysis, with a 0.05 significance level, using R software (R Core Team, 2013). Only families represented in both early and advanced generations (N=146) were included in the analysis. Fruiting efficiency at maturity values used in this analysis were from individual plants in the F_2 generation (experiment BF209), average of two replications (rows) in the F_3 generation (experiment MF310) and best linear unbiased predictors (BLUP) in the RILs, as obtained from a mixed linear model which included experiment (*i.e.*, environment) as a fixed effect and genotype and genotype by environment interaction as random effects (Alonso et al., 2018b). Data from seven experiments carried out between 2013 and 2017 (Table 1) were used in this model. In

addition, correlation analysis was carried out between F_2 and F_3 FE_m values and the average FE_m value of each RIL at each of these seven experiments individually.

1.4.2. *Responses after simulated selection*

Only families represented in both early and advanced generations (N=146) were included in the simulated selection analysis. Firstly, simulated selection for high FE_m was carried out in the F_2 generation under three selection intensities: 50%, 25% and 10%. Response to selection in FE_m was assessed at the F_3 generation and in the RILs by calculating the mean FE_m of the selected group and comparing it with the population mean of each experiment. Response to selection was then expressed as percentual increase in the selected group as compared with the population mean.

Responses in GN and GY of RILs after simulated selection for high FE_m in early generations were calculated under four different selection strategies, depending on the selection intensity (*i*) used:

1) Selection of the top 50% F_2 followed by selection of the top 50% F_3 (i_{F2}=50% and i_{F3}=50%);
2) Selection of the top 25% F_2 followed by selection of the top 25% F_3 (i_{F2}=25% and i_{F3}=25%);
3) Selection of the top 25% F_2 followed by selection of the top 10% F_3 (i_{F2}=25% and i_{F3}=10%);
4) Selection of the top 10% F_2 followed by selection of the top 10% F_3 (i_{F2}=10% and i_{F3}=10%).

Responses after simulated selection were calculated for GN and GY as the difference between the mean values of the 'selected group' and the general mean of the population at each response experiment (BFE13, BFE14 and BFE15). They were then expressed as percentual increase in the selected group as compared with the population mean.

1.5. Evaluation of SHI as a selection criterion in early generations

1.5.1. Mode of inheritance

Frequency histograms were used for evaluating SHI distribution in the F_2, F_3 and RILs. Also, as described earlier for FE_m, BLUPs of SHI were obtained in the RILs from a mixed linear model which included environment as a fixed effect and genotype and genotype by environment interaction as

random effects. Variance components estimated by this model were used to calculate SHI's broad-sense heritability (H^2) according to Hallauer et al. (2010), as follows:

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + (\sigma_{ge}^2/e) + (\sigma_{res}^2/re)}$$

where σ_g^2 is the genotypic variance; (σ_{ge}^2 is the genotype x environment interaction variance; σ_{res}^2 is the error variance; *e* is the number of environments and *r* is the number of replications per experiment.

1.5.2. Correlation analysis

The association of SHI measurements taken in early generations *versus* in RILs was established by Pearson correlation analysis, following the same methodology as described earlier for FE_m. Pearson correlation analysis was also performed to determine the degree of association of SHI with GN and GY, with a 0.05 significance level.

1.5.3. Responses after simulated selection

The same simulated selection strategies and response to selection assessments carried out for FE_m, described above, were used for SHI.

2. RESULTS

2.1. Evaluation of FE_m as a selection criterion in early generations

2.1.1. Correlation analysis

Correlation coefficients of FE_m between early generations (F_2 and F_3) and RILs are shown in Table 2. A significant, positive association (r=0.46, P<0.05) between FE_m in F_2 (individual plant) and FE_m in F_3 (row) was detected. Moreover, significant, positive associations were found between FE_m in early generations and FE_m BLUPs in the RILs (r=0.32, P<0.05).

Additionally, correlation coefficients were calculated for FE_m between F_2 (experiment BF209), F_3 (experiment MF310) and RILs as evaluated at each of seven experiments individually (BFE13, BFE14, BFE15, BFF16-1, BFF16-2, BFF17 and MJFE17). As a result, correlation coefficients of FE_m between F_2 and RILs were positive and significant in all cases (r=0.23 to 0.32; P<0.05; Table S1) except in BFE14. Similarly, correlation coefficients of FE_m between F_3 and RILs were positive and significant for all experiments (r=0.19 to 0.33; P<0.05; Table S1).

Table 2. Pearson's correlation coefficients of fruiting efficiency at maturity between F_2 (individual plant values), F_3 [average of two replications (rows)] and RILs (BLUPs of seven experiments, Table 1) of the Baguette 10 x Klein Chajá population (N=146). All coefficients are significant (P<0.05).

	F_2	F_3
F_3	0.46	
RILs	0.32	0.32

2.1.2. Responses after simulated selection

Simulated selection for high FE_m in the F_2 generation yielded a consistent increase in FE_m both in the F_3 generation and in the RILs, as tested in seven experiments which spanned two very contrasting locations and different years (Table S2). Responses in GN and GY in the RILs after simulated selection for high FE_m carried out in the F_2 and F_3 generations under four selection strategies are presented in Table 3. Responses in GN were positive under all four simulated selection strategies and in all three experiments (BFE13, BFE14 and BFE15); that is, selection for high FE_m in early generations yielded GN means of the selected groups that were always higher than the population mean. The mean response for each strategy was positive and increased as selection pressure increased, from 5.1% (Selection strategy 1) to 22.1% (Selection strategy 4). Responses in GY of RILs to simulated selection for high FE_m in early generations were more variable than those in GN. In two out of three experiments in which RILs were evaluated (BFE14 and BFE15) mean responses to simulated selection for high FE_m in early generations were positive (i.e., under all four selection strategies, with

overall means of 7.9 and 4% respectively). Conversely, responses to selection were negative in BFE13 (Table 3). Considering the mean response values per selection strategy across all experiments, RILs' GY response to simulated selection for FE_m in early generations were 1.3, 1.3, -1.3 and 2.1% for Selection strategies 1 to 4, respectively.

Table 3. Simulated selection for high FE_m (grains/g spike chaff) carried out in the F_2 and F_3 generations under four selection strategies [(1) i_{F2}=50% and i_{F3}=50%; (2) i_{F2}=25% and i_{F3}=25%; (3) i_{F2}=25% and i_{F3}=10%, and (4) i_{F2}=10% and i_{F3}=10%]: response to selection in grain number (GN; grains/m^2) and grain yield (GY; g/m^2) in the RILs of the Baguette 10 x Klein Chajá population (N=146) as evaluated in experiments BFE13, BFE14 and BFE15. Response to selection expressed as percentual increase in the mean of the selected group as compared with the population mean.

Trait	GN				GY			
Experiment	BFE13	BFE14	BFE15	Mean	BFE13	BFE14	BFE15	Mean
Population mean	22092	9978	17763	16611	719.5	367.4	749.1	612.0
Selection strategy 1								
Mean of the selected group (N=36)	22986	10494	18910	17463	716.9	369.6	773.1	619.9
Response to selection (%)	**4.1**	**5.2**	**6.5**	**5.1**	**-0.4**	**0.6**	**3.2**	**1.3**
Selection strategy 2								
Mean of the selected group (N=9)	23167	11113	19525	17935	708.6	376.2	774.5	619.8
Response to selection (%)	**4.9**	**11.4**	**9.9**	**8.0**	**-1.5**	**2.4**	**3.4**	**1.3**
Selection strategy 3								
Mean of the selected group (N=4)	22852	11930	20761	18514	657.0	373.3	782.5	604.3
Response to selection (%)	**3.4**	**19.6**	**16.9**	**11.5**	**-8.7**	**1.6**	**4.5**	**-1.3**
Selection strategy 4								
Mean of the selected group (N=2)	22675	15702	22485	20287	624.7	465.8	784.1	624.9
Response to selection (%)	**2.6**	**57.4**	**26.6**	**22.1**	**-13.2**	**26.8**	**4.7**	**2.1**
Overall response to selection (%)	**3.8**	**23.4**	**15.0**	**11.7**	**-6.0**	**7.9**	**4.0**	**1.8**

2.2. Evaluation of SHI as a selection criterion in early generations

2.2.1. *Mode of inheritance*

Frequency histograms of SHI distribution in the F_2 and F_3 generations and SHI BLUPs in the RILs are shown in Figure 1. In all three cases, a bell-shaped, symmetrical distribution was observed. Spike harvest index values of the parents of the population in 2009 (as evaluated along with the F_2 plants in experiment BF209) were 3.2 g/g for B10 and 2.8 g/g for KCJ. In turn, SHI values in the F_2 ranged between 0.6 and 4.5 g/g, evidencing the existence of transgressive segregation for the trait in this population. Broad-sense heritability of SHI, as evaluated in the RILs, was 0.76.

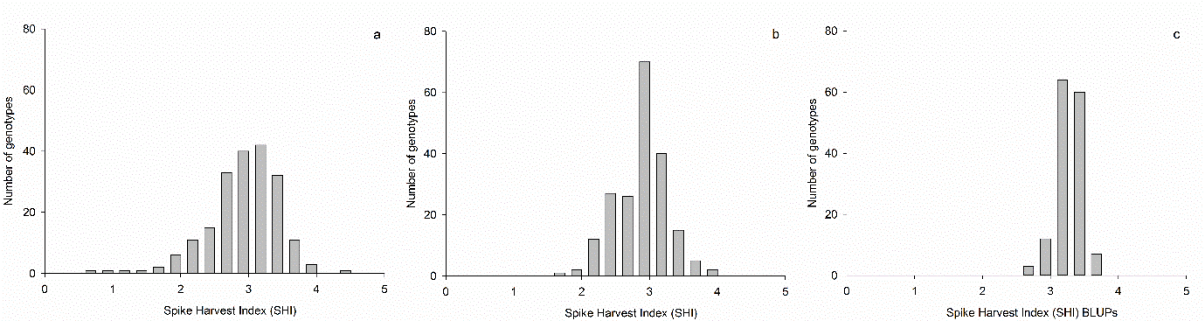

Figure 1. Frequency histograms of spike harvest index (SHI) distribution in the F_2 (a) and F_3 (b) generations, and SHI BLUPs in the RILs (c) of the Baguette 10 x Klein Chajá population.

2.2.2. Correlation analysis

A positive, significant association was found for SHI between the F_2 and F_3 generations, and between the F_3 generation and the RILs (Table 4). Conversely, no association was detected between SHI measured in the F_2 generation and SHI measured in the RILs.

Table 4. Pearson's correlation coefficients of spike harvest index between F_2 (individual plant values), F_3 [average of two replications (rows)] and RILs (BLUPs of seven experiments, Table 1) of the Baguette 10 x Klein Chajá population (N=146). Significant coefficients (P<0.05) are shown in bold.

	F_2	F_3
F_3	**0.29**	
RILs	0.15	**0.26**

Correlation analysis results for SHI with GN and GY are shown in Table 5. There was no association (P>0.05) between SHI and GN in this population, whereas SHI was positively associated with GY in only one out of three experiments (BFE13).

Table 5. Pearson's correlation coefficients between BLUPs of spike harvest index (SHI) and grain number (GN) and between SHI BLUPs and grain yield (GY) as determined in the RILs of the Baguette 10 x Klein Chajá population (N=146) in three experiments (BFE13, BFE14 and BFE15). Significant coefficients (P<0.05) are shown in bold.

		BLUPs SHI
GN	BFE13	-0.03
	BFE14	-0.06
	BFE15	-0.11
GY	BFE13	**0.17**
	BFE14	0.12
	BFE15	0.07

2.2.3. Responses after simulated selection

Recombinant inbred lines' GN and GY responses to selection using SHI as selection criterion in early generations are summarized in Table S3. Responses in GN were negative or neutral, except in experiment BFE15 under Selection strategy 3; on average, responses in GN ranged between -4.7 and -20.5% depending on the Selection strategy. Regarding responses in GY, they were also mostly

negative except for two cases in experiment BFE15. On average, responses ranged between -3.2 and -7.1% for Selection strategies 1 to 4, respectively.

3. DISCUSSION

For the first time in bread wheat breeding history, we show the feasibility of selecting for high FE_m in early generations and achieving a consistent increase in GN in advanced generations. This has the potential of impacting grain yield genetic progress in a similar way as we have observed in the past with early selection for dwarfing or photoperiod insensitivity genes.

We know that FE_m is tightly linked to GN and, in turn, GN is tightly linked to GY (Fischer, 1985; Abbate et al., 1995; Calderini et al., 1999; González et al., 2014). We also have extensive evidence that FE_m has all the features of a good selection criterion (see Pretini et al. 2021 for a review). Our previous work of simulated selection for high FE_m in advanced generations and assessment of the response in GN and GY in the same advanced generations showed increases in both GN and GY (Alonso et al., 2018b; Pontaroli et al., 2021). Even though we had FE_m data of early generations for one of the populations we used in the latter study, we were not quite convinced that early selection for high FE_m would be practical because it is time-consuming and tedious to have to cut each spike of each plant at the base of the rachis and do this in thousands of F_2 and / or F_3 plants. However, recent developments in automated phenotyping systems have given us hope that one such system can be designed to expedite this process. In the meantime, considering the positive results we observed when performing selection in advanced lines and, in an attempt to provide support to Fischer and Rebetzke (2018)'s proposal of using FE_m as a selection criterion in early-generation, spaced plants, we considered it worthwhile to go back and reappraise the early generations' data. The first thing we did was to determine the correlation between FE_m values as taken in F_2, F_3 and RILs (advanced lines without selection). As a result, positive, significant correlations were observed ($r_{F2\text{-}F3}$=0.46; $r_{F2\text{-}RILs}$=0.32; $r_{F3\text{-}RILs}$=0.32; P<0.05); this was not surprising, considering the high heritability of the trait (González et al., 2011; Martino et al., 2015; Mirabella et al., 2016; Alonso et al., 2018b; Ramirez et al., 2018; Pretini et al., 2020; Pontaroli et al., 2021). Nevertheless, it was encouraging to find these

correlations considering that F_2 and F_3 plants are highly heterozygous, and measurements were taken on individual plants or rows as opposed to the sampling done in advanced generations, which entailed drawing a random spike sample from yield plots in homozygous lines. Also, these correlations were calculated with the FE_m BLUPs in the RILs, which removed the environmental and genotype by environment effects. Still, a good repeatability was observed when the F_2-RILs and F_3-RILs correlation analyses was carried out with data of each experiment individually (Table S1). Similarly, good repeatability (r values between 0.4 and 0.66; $P<0.05$) was observed between measurements taken on the RILs across different experiments, which spanned two contrasting locations (Marcos Juárez and Balcarce) and five different crop seasons at Balcarce (Table S1).

The following logical thing to ascertain was whether selection for high FE_m in early generations could increase FE_m in advanced lines. Given the high heritability of the trait and the positive correlations mentioned above, it was expected to find a favourable response to selection. Effectively, this was the case (Table S2): performing selection for high FE_m in the F_2 generation yielded 8.8 and 11.9% increases in FE_m in the F_3 generation using 25% and 10% selection intensities, respectively. Response to the same selection schemes but in the RILs resulted in average 3.6 and 8.4% increases in FE_m for 25% and 10% selection intensities in the F_2, respectively (Table S2). These results suggest that selection for high FE_m at each generation of a breeding program could potentially produce a steady increase in FE_m across generations.

Next, we wanted to ascertain whether the use of FE_m as a selection criterion in early generations indirectly increased GN and GY of advanced lines, as we had observed when performing both selection for high FE_m and assessment of GN and GY response in the advanced lines (Alonso et al., 2018b; Pontaroli et al., 2021). We show here that selection for high FE_m in both the F_2 and F_3 generations always increases GN of advanced lines derived from them, regardless of the environmental conditions tested in this study and proportionally to the selection intensity (Table 3). Response in GN in the RILs was +5.1% as compared with the population mean under a 50% FE_m selection intensity in both F_2 and F_3 generations, whereas it jumped to 22.1% when a more stringent selection strategy was used (10% FE_m selection intensity in both F_2 and F_3 generations). Response in

GY, although as not as compelling as response in GN, was also positive on average (1.8%; Table 3). As a result of selection for high FE_m in the F_2 and F_3 generations, GY was increased in experiments BFE14 and BFE15, whereas in experiment BFE13 it decreased 0.4 (Selection strategy 1) to 13.2% (Selection strategy 4) as compared with the population mean. As described in Alonso et al. (2018a), environmental conditions during experiment BFE13 greatly favoured GN determination (Table 3, mean GN=22092 grains/m^2; ~2.2 and ~1.2-fold greater than GN in experiments BFE14 and BFE15, respectively); the grain filling period, on the other hand, was shorter and warmer than in experiment BFE15 (the one with the highest mean GY; Table 3). This suggests that early generation selection for high FE_m alone may lead to an increase in GN together with a reduced grain weight due to grain filling limitations under certain conditions. We also observed this in some cases when evaluating simulated selection for high FE_m in advanced lines and assessment of GN and GY response to selection in the same advanced lines (Pontaroli et al., 2021). Then, concurrent selection for both high FE_m and grain filling quality in early generations appears to be a pertinent strategy for securing a more stable GY increase in advanced lines, particularly for target environments of high yield potential.

We considered it worthwhile to explore the use of the spike harvest index as an alternative selection criterion in early generations, under the hypothesis that it could potentially impact both GN and GW simultaneously in advanced lines. Although (i) substantial genetic variation for SHI was detected in the population under study, (ii) broad sense heritability was quite high (0.76), and (iii) the trait determined in the F_3 was significantly and positively associated with the trait determined in the RILs, virtually no association was detected between SHI and GN or GY (Table 5). Hence, it was not surprising to find no increase in GN or GY after simulated selection for high SHI in early generations, when following the same selection strategies as used with FE_m. In fact, most responses were negative: on average, GN and GY respectively decreased 9.3 and 4.4% as compared with the population means (Table S3). Therefore, at least in the population and the environmental conditions under study, SHI would not be a useful selection criterion for increasing GN or GY.

Although further validation in different breeding populations and in a broader spectrum of environments is warranted, the results shown here strongly support the feasibility of increasing GN

and GY of advanced lines after performing early generation selection for high FE_m. Provided that a high-throughput, automated sample processing system becomes available and / or that more validated QTL and candidate genes associated with FE_m are detected and can be used in marked assisted selection (Pretini et al., 2021), this process can certainly boost the much-needed yield genetic progress in bread wheat.

AUTHOR CONTRIBUTION

MPA, MFF and ACP conceived the study. Field experiments were designed by ACP and carried out by ACP, MPA and MFF. Statistical analyses were performed by MPA with the contribution of ACP and MFF; production of figures and tables was performed by MFF with the contribution of MPA and ACP. The manuscript was written by all the authors. All authors read and approved the final manuscript.

ACKNOLEDGMENTS

We thank members of the Grupo Trigo Balcarce (EEA Balcarce INTA) and Mejoramiento de Trigo y Biotecnología (EEA Marcos Juárez INTA) for help with the experiments and technical assistance. Scholarships granted to M.F. Franco and M.P. Alonso by CONICET are acknowledged. This research was partially funded by INTA (PNCyO 1127044).

REFERENCES

Abbate, P., Andrade, F., Lazaro, L., Bariffi, J., Berardocco, H., Inza, V., Marturano, F., 1998. Grain yield increase in recent Argentine wheat cultivars. Crop Sci. 38, 1203-1209. https://doi.org/10.2135/cropsci1998.0011183X003800050015x.

Abbate, P.E., Andrade, F.H., Culot, J.P., 1995. The effects of radiation and nitrogen on number of grains in wheat. The Journal of Agricultural Science 124, 351-360. https://doi.org/10.1017/S0021859600073317.

Abbate, P.E., Pontaroli, A.C., Lázaro, L., Gutheim, F., 2013. A method of screening for spike fertility in wheat. The Journal of Agricultural Science 151, 322-330. https://doi.org/10.1017/S0021859612000068.

Alonso, M., Abbate, P., Mirabella, N., Merlos, F., Panelo, J., Pontaroli, A., 2018a. Analysis of sink/source relations in bread wheat recombinant inbred lines and commercial cultivars under a high yield potential environment. European Journal of Agronomy 93, 82-87. https://doi.org/10.1016/j.eja.2017.11.007.

Alonso, M., Mirabella, N., Panelo, J., Cendoya, M., Pontaroli, A., 2018b. Selection for high spike fertility index increases genetic progress in grain yield and stability in bread wheat. Euphytica 214, 112. https://doi.org/10.1007/s10681-018-2193-4.

Calderini, D., Reynolds, M., Slafer, G., 1999. Genetic gains in wheat yield and main physiological changes associated with them during the 20th century. In: Satorre, E.H., Slafer, G. (Eds.), Wheat: Ecology and Physiology of Yield Determination. Food Product Press, New York, p. 351—377.

CIMMYT, CENTRO INTERNACIONAL DE MEJORAMIENTO DE MAIZ Y TRIGO, 2019. Nutrition, health and food security. ‹https://www.cimmyt.org/news/nutrition-health-food-security/›.

Elía, M., Savin, R., Slafer, G.A., 2016. Fruiting efficiency in wheat: physiological aspects and genetic variation among modern cultivars. Field Crops Res. 191, 83-90. https://doi.org/10.1016/j.fcr.2016.02.019.

Ferrante, A., Cartelle, J., Savin, R., Slafer, G.A., 2017. Yield determination, interplay between major components and yield stability in a traditional and a contemporary wheat across a wide range of environments. Field Crops Res. 203, 114-127. https://doi.org/10.1016/j.fcr.2016.12.028.

Ferrante, A., Savin, R., Slafer, G.A., 2012. Differences in yield physiology between modern, well adapted durum wheat cultivars grown under contrasting conditions. Field Crops Res. 136, 52-64. https://doi.org/10.1016/j.fcr.2012.07.015.

Fischer, R., 1984. Wheat. In: Smith, W., Banks, S. (Eds.), Proceedings of symposium on potential productivity of field crops under different environments, IRRI, Philippines, pp. 129–154.

Fischer, R., 1985. Number of kernels in wheat crops and the influence of solar radiation and temperature. The Journal of Agricultural Science 105, 447-461. https://doi.org/10.1017/S0021859600056495.

Fischer, R., Rebetzke, G., 2018. Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: a review. Crop and Pasture Science 69, 439-459. https://doi.org/10.1071/CP17409.

Foulkes, J., Rivera, C., Trujillo, E., Sylvester-Bradley, R., Reynolds, M., 2015. Achieving a step-change in harvest index in high biomass wheat cultivars. In: Reynolds, M., Molero, G., Mollins, J., Braun, H. (Eds.), Proceedings of the International TRIGO (Wheat) Yield Potential Workshop 2015. CENEB, CIMMYT, Obregón, Sonora, Mexico, pp. 31-35.

Franco, M., Lori, G., Panelo, J., Alonso, M., NE, M., Malbrán, I., Cendoya, M., Pontaroli, A.C., 2021. Using anthesis date as a covariate to accurately assessing type II resistance to Fusarium head blight in field-grown bread wheat. Crop Protect. 142, 105504. https://doi.org/10.1016/j.cropro.2020.105504.

Gerard, G.S., Alqudah, A., Lohwasser, U., Börner, A., Simón, M.R., 2019. Uncovering the genetic architecture of fruiting efficiency in bread wheat: a viable alternative to increase yield potential. Crop Sci. 59, 1853-1869. https://doi.org/10.2135/cropsci2018.10.0639.

González, F., Terrile, I.I., Falcón, M., 2011. Spike fertility and duration of stem elongation as promising traits to improve potential grain number (and yield): variation in modern Argentinean wheats. Crop Sci. 51, 1693-1702. https://doi.org/10.2135/cropsci2010.08.0447.

González, F.G., Aldabe, M.L., Terrile, I.I., Rondanini, D.P., 2014. Grain weight response to different postflowering source: Sink ratios in modern high-yielding Argentinean wheats differing in spike fruiting efficiency. Crop Sci. 54, 297-309. https://doi.org/10.2135/cropsci2013.03.0157.

Hallauer, A.R., Carena, M.J., Miranda Filho, J.d., 2010. Quantitative genetics in maize breeding. Springer Science & Business Media, Iowa.

Martino, D.L., Abbate, P.E., Cendoya, M.G., Gutheim, F., Mirabella, N.E., Pontaroli, A.C., 2015. Wheat spike fertility: inheritance and relationship with spike yield components in early generations. Plant Breeding 134, 264-270. https://doi.org/10.1111/pbr.12262.

Mirabella, N., Abbate, P., Ramirez, I., Pontaroli, A., 2016. Genetic variation for wheat spike fertility in cultivars and early breeding materials. The Journal of Agricultural Science 154, 13-22. https://doi.org/10.1017/S0021859614001245.

Pontaroli, A.C., Alonso, M.P., Mirabella, N.E., Panelo, J.S., Franco, M.F., Vanzetti, L.S., Lorenzo, M., 2021. Increasing Grain Yield in Bread Wheat by Selection for High Spike Fertility Index. bioRxiv. https://doi.org/10.1111/pbr.12945.

Pretini, N., Alonso, M.P., Vanzetti, L.S., Pontaroli, A.C., González, F.G., 2021. The physiology and genetics behind fruiting efficiency: a promising spike trait to improve wheat yield potential. J. Exp. Bot. 72, 3987-4004. https://doi.org/10.1093/jxb/erab080.

Pretini, N., Terrile, I.I., Gazaba, L.N., Donaire, G.M., Arisnabarreta, S., Vanzetti, L.S., González, F.G., 2020. A comprehensive study of spike fruiting efficiency in wheat. Crop Sci. 60, 1541-1555. https://doi.org/10.1002/csc2.20143.

Quail, K., Fischer, R., Wood, J., 1989. Early generation selection in wheat. I. Yield potential. Aust. J. Agric. Res. 40, 1117-1133. https://doi.org/10.1071/AR9891117.

R Core Team, 2013. R: A language and environment for statistical computing. R Core Team. R Foundation for Statistical Computing, Vienna, Austria.

Ramirez, I.A., Abbate, P.E., Redi, I.W., Pontaroli, A.C., 2018. Effects of photoperiod sensitivity genes Ppd-B1 and Ppd-D1 on spike fertility and related traits in bread wheat. Plant Breeding 137, 320-325. https://doi.org/10.1111/pbr.12585.

Slafer, G.A., Elia, M., Savin, R., García, G.A., Terrile, I.I., Ferrante, A., Miralles, D.J., González, F.G., 2015. Fruiting efficiency: an alternative trait to further rise wheat yield. Food and Energy Security 4, 92-109. https://doi.org/10.1002/fes3.59.

Terrile, I.I., Miralles, D.J., González, F.G., 2017. Fruiting efficiency in wheat (*Triticum aestivum* L): Trait response to different growing conditions and its relation to spike dry weight at anthesis and grain weight at harvest. Field Crops Res. 201, 86-96. https://doi.org/10.1016/j.fcr.2016.09.026.

Supplementary information

(Three tables)

Table S1. Pearson's correlation coefficients for FE_m in the Baguette 10 x Klein Chajá population (N=146) between F_2 (individual plants evaluated in exp. BF209), F_3 [average of two replications (rows) from exp. MF310] and RILs as evaluated in seven experiments (BFE13, BFE14, BFE15, BFF16-1, BFF16-2, BFF17 and MJFE17). All coefficients were significant (P<0.05) except where noted (ns).

				RILs					
		F_2	F_3	BFE13	BFE14	BFE15	BFF16-1	BFF16-2	BFF17
	F_3	0.46							
RILs	BFE13	0.28	0.28						
	BFE14	0.16^{ns}	0.24	0.53					
	BFE15	0.32	0.33	0.63	0.66				
	BFF16-1	0.25	0.23	0.49	0.60	0.66			
	BFF16-2	0.28	0.20	0.46	0.51	0.61	0.51		
	BFF17	0.30	0.23	0.57	0.56	0.59	0.46	0.60	
	MJFE17	0.23	0.19	0.41	0.40	0.57	0.62	0.59	0.43

Table S2. Simulated selection for high FE_m (grains/g spike chaff) in the F_2 generation under two selection intensities (i=25% and i=10%) and response to selection in FE_m in the F_3 generation and in the RILs of the Baguette 10 x Klein Chajá population (N=146). Response to selection expressed as percentual increase in the mean of the selected group as compared with the population mean.

			RILs						
	F_2	F_3	BFE13	BFE14	BFE15	BFF16-1	BFF16-2	BFF17	MJFE17
Population mean	79.4	76.2	98.5	92.0	89.4	92.6	89.4	101.7	90.5
Mean of top 25% F_2 plants	94.6	82.9	101.8	93.0	93.1	96.4	92.9	106.2	94.4
Mean of top 10% F_2 plants	101.7	85.3	104.3	97.1	97.2	102.3	96.1	111.7	100.0
Response to selection (i=25%), %	19.2	8.8	3.4	1.2	4.2	4.1	3.9	4.4	4.3
Response to selection (i=10%), %	28.1	11.9	5.9	5.6	8.8	10.6	7.4	9.9	10.5

Table S3. Simulated selection for high spike harvest index (SHI; g of grains/g spike chaff) carried out in the F_2 and F_3 generations under four selection strategies [(1) i_{F2}=50% and i_{F3}=50%; (2) i_{F2}=25% and i_{F3}=25%; (3) i_{F2}=25% and i_{F3}=10%, and (4) i_{F2}=10% and i_{F3}=10%]: response to selection in grain number (GN; grains/m^2) and grain yield (GY; g/m^2) in the RILs of the Baguette 10 x Klein Chajá population (N=146) as evaluated in experiments BFE13, BFE14 and BFE15. Response to selection expressed as percentual increase in the mean of the selected group as compared with the population mean.

Trait	GN				GY			
Experiment	BFE13	BFE14	BFE15	Mean	BFE13	BFE14	BFE15	Mean
Population mean	22092	9978	17763	16611	719.5	367.4	749.1	612.0
Selection strategy 1								
Mean of the selected group (N=36)	20692	9412	17253	15786	696.1	349.9	736.7	594.2
Response to selection (%)	**-6.3**	**-5.7**	**-2.9**	**-5.0**	**-3.3**	**-4.8**	**-1.7**	**-3.2**
Selection strategy 2								
Mean of the selected group (N=9)	19389	9784	17761	15644	668.4	351.5	759.2	593.0
Response to selection (%)	**-12.2**	**-1.9**	**0.0**	**-4.7**	**-7.1**	**-4.3**	**1.3**	**-3.4**
Selection strategy 3								
Mean of the selected group (N=4)	20810	8126	18342	15760	678.9	324.1	793.7	598.9
Response to selection (%)	**-5.8**	**-18.6**	**3.3**	**-7.0**	**-5.6**	**-11.8**	**6.0**	**-3.8**
Selection strategy 4								
Mean of the selected group (N=2)	13486	8294	16759	12846	627.1	336.7	747.2	570.3
Response to selection (%)	**-39.0**	**-16.9**	**-5.7**	**-20.5**	**-12.8**	**-8.3**	**-0.3**	**-7.1**
Overall response to selection (%)	**-15.8**	**-10.8**	**-1.3**	**-9.3**	**-7.2**	**-7.3**	**1.4**	**-4.4**