Increasing Grain Yield in Bread Wheat (*Triticum aestivum* L.) by Selection for High Spike Fertility Index

Ana Clara Pontaroli[1,2*], María Pía Alonso[1,2], Nadia Estefania Mirabella[1], Juan Sebastián Panelo[1], María Fiorella Franco[1,2], Leonardo Sebastián Vanzetti[2,3], Máximo Lorenzo[1]

[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.

[3] Estación Experimental Agropecuaria Marcos Juárez, INTA; Ruta 12 km 3 (2580) Marcos Juárez, Argentina.

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

Running title: Spike fertility index selection in wheat

ACKNOWLEDGMENTS

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 and to N.E. Mirabella by INTA are acknowledged. This research was partially funded by INTA (PNCyO 1127044).

ABSTRACT

Spike fertility index (SF; a.k.a. fruiting efficiency at maturity) has been proposed as a promising selection criterion for increasing grain yield (GY) in bread wheat. In this study, changes in GY and related traits after simulated selection (10% intensity) for high SF or high GY were assessed in (i) a RIL population of 90 individuals (Avalon/Glupro) (ii) a RIL population of 146 individuals (Baguette 10/Klein Chajá); (iii) 28-36 advanced lines from a breeding program. Grain yield, SF, grain number per unit area (GN), grain weight, test weight and grain protein content were determined. Regardless of the environmental conditions, simulated selection for high SF alone always resulted in GN increases (between 1.6 and 27.4%). Average GY increase observed after simulated selection for high SF (5.6%; N=20; S^2=80.1) did not differ (p=0.2) from the average GY increase observed after simulated selection for GY *per se* (12.9%; N=12; S^2=308.1). Our findings validate the use of SF as a selection criterion for increasing genetic progress of grain yield in bread wheat breeding programs.

KEYWORDS: grain yield, grain number; fruiting efficiency; breeding; selection; genetic progress

INTRODUCTION

Grain yield is one of the key targets of bread wheat (*Triticum aestivum* L.) breeding programs worldwide. Under a scenario of increased global demand for grains and limited possibilities of expanding cropping areas, breeding efforts must focus on raising annual yield increase rates (1). One way to achieve this is by using yield-related, high-throughput phenotypic variables as selection criteria.

Grain yield in wheat is determined mainly by the number of grains per unit area (GN) and, to a lesser extent, by grain weight (GW). Therefore, variations in GN best explain yield variation (2-8). However, this is a difficult trait to select for in early generations of a breeding program, in which not enough seed is available. Fischer (9) proposed that, under non-limiting growing conditions (i.e., without water or nutrient limitations and in absence of pests and diseases), GN is determined by 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 (SF) (2) or “fruiting efficiency” (10).

There is extensive evidence that GN is tightly linked to SF (2, 7, 11-14). In the reference method (9), SF determination is carried out at anthesis, but this is destructive and very time consuming. A high throughput, non-destructive method for assessing the trait in small samples at maturity has then been proposed (17). In addition, SF at maturity has been shown to have high heritability, substantial genetic variability and low genetic by environment interaction (12, 15, 16). All these features turn SF into a promising selection criterion for increasing GN (and yield) in breeding programs (4, 14-25). As a matter of fact, Alonso et al. (12) have shown that selection for high SF, either solely or in combination with selection for high grain yield,

effectively increases yield, resulting in higher and more stable yields than if selecting for high yield *per se*. These encouraging results, obtained with a single RIL population under non-limiting environmental conditions, need further validation before such selection strategies can be implemented in a breeding program. Also, the assessment of possible trade-offs that could arise as a byproduct of these schemes should be carried out. Therefore, the aim of this work was to determine the effectiveness of selection for high SF in raising grain yield and its effect on other variables, by analyzing a combination of diverse genetic materials (RILs and advanced lines) under an array of environmental conditions.

MATERIALS AND METHODS

Field experiments and plant material

Field experiments were carried out at the Balcarce (37º45' S; 58º18' W; 130 m a.s.l.) and Marcos Juárez (32°43'S; 62°06'W; 112 m a.s.l.) Agricultural Experimental Stations of the Instituto Nacional de Tecnología Agropecuaria (INTA), Argentina. Characteristics of experiments are summarized in Table 1.

Experiments were conducted under no nutrient limitations; weeds, pests and fungal diseases were chemically controlled. As described in Table 1, experiments were grouped into Sets I, II and III. In Set I, a biparental population of 90 recombinant inbred lines (RIL) derived from the cross between 'Avalon' and 'Glupro' was evaluated in experiments BAG15 and BAG16. 'Avalon' is a cultivar from the UK with high protein content, whereas 'Glupro' is a cultivar from the USA with the introgression of the *GPC-B1* gene for high protein content. In Set II, a biparental population of 146 RIL derived from the cross between 'Baguette 10' and 'Klein Chajá' (12, 15, 16) was evaluated in experiments BFE13, BFE14, BFE15 and MJFE17. Both parents are Argentinian spring bread wheat cultivars of good agronomic performance, with contrasting SF and yield components. In Set III, breeding lines derived from crosses between cultivars of diverse origin (CIMMYT, Europe, Argentina), advanced through a modified bulk selection scheme and selected for grain yield and agronomic performance in a wheat breeding program (INTA Balcarce Experimental Station), were evaluated in experiments BLE16 and BLE17 (28 lines), and BLP17 and BLE18 (36 lines).

Data collection

Grain yield was determined 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). In the former case, the plot (and harvested) area was 3.5 m^2, with an inter-plot distance of 0.3 m; in the latter case, the harvested area was 5 m^2. Grain moisture was obtained from a grain subsample of each plot to calculate dry grain yield (GY; g/m^2). Grain weight (GW; mg) was determined by counting a 1000-grain sample with an electronic counter, weighing it and dividing total weight by 1000. Grain number per square meter was then calculated as the quotient between GY and GW $\times$ 1000.

Spike fertility index was determined at maturity on a sample of 15-20 spikes drawn at random from the harvestable area of each plot prior to mechanical harvest. Spikes were cut at the lowest spikelet level, counted, weighed, and threshed. Spike chaff dry weight (g) was calculated as the difference between total spike dry weight (i.e. before threshing) and total grain weight. Grains were counted using an electronic counter. Then, SF (grains/g) was calculated as the quotient between the number of grains and spike chaff dry weight per sample (17).

Test weight (TW) was measured using a Schopper chondrometer on a 250 g grain sample and expressed as kg/hl. Grain protein content (GPC; %) was determined indirectly by NIRS using a TEC NIR 256 Inlab NIR in 100 g grain samples, previously cleaned and dried at 60°C to constant weight.

Selection for high spike fertility index and responses to selection in grain yield and other traits

Data from three sets of experiments (Table 1) were used to evaluate responses in GY and other traits after simulated selection for high SF. At each set, the 10% top lines with the highest SF values were selected, based on SF data generated in experiments BAG16 (Set I), BFE13, BFE14, BFE15 and MJFE17 (Set II) and BLE16, BLE17, BLP17 and BLE18 (Set III). Their performance was evaluated based on data generated in the remaining experiments of each set. In order to compare the effects of selection for high SF with those for high GY *per se*, a similar approach was followed by simulating selection for high GY *per se* (10% intensity) and assessing responses to selection.

Responses to selection were calculated for yield and other traits as the difference between the mean values of the 'selected group' and the general mean of the population at each testing experiment. Response to selection was then expressed as a percentage of the population mean.

Statistical analysis

Grain yield was analyzed with a linear fixed effects model including experiments and genotypes and genotype by experiment interaction effects nested in sets of experiments. T-test was used to compare responses to selection (in %) between selection strategies (i.e., for high SF vs. for high GY *per se*).

RESULTS

Experiments included in Sets I to III were carried out under different environmental conditions (i.e. different locations, years, growing conditions, plot size, etc.; Table 1). In addition, conspicuous environmental differences between crop seasons due to temperature, radiation and rainfall were observed (Fig. 1). This resulted in significant differences in average grain yield between experiments; differences between genotypes were also evident across experiments (Table 2).

As a consequence of the observed environmental differences and the genetic constitution of the populations under study, conspicuous phenotypic variation was observed for all traits (Fig. 2).

Very different population means were also observed across experiments of each set, reflecting environmental variation between years and/or locations (Table 3). In each set of experiments, simulated selections for high SF were carried out with data generated in one experiment in a small sample of 15-20 spikes, and the response to selection assessed with data of GY and other traits generated at the plot scale in the same and/or in other experiments. A summary of these results is presented in Table 3. Selection of the top 10% lines with the highest SF in experiments BAG16 (Set I), BFE13, BFE14, BFE15 and MJFE17 (Set II) and BLE16 and BLE17, BLP17 and BLE18 separately (Set III) always increased GN, and also increased GY except for a few cases of slight decrease in Set III. Overall, GY and GN respectively increased by 5.6 and 13.4% as a result

of simulated selection for high SF (Table 3), as compared with the population means. On the other hand, this selection strategy consistently lowered GW (overall decrease of 5.3%, Table 3). As for TW, response to selection for high SF was slightly positive in Set I and slightly negative in Set II, but not significant. Grain protein content, only determined in Set I, was decreased by 6.9% as a result of selection for high SF.

Selection responses after simulated selection by high GY *per se* are presented in Table S1. Both selection strategies (i.e., by high SF and by high GY *per se*) were compared in terms of the average response to selection attained in GN and GY (Table 4). As a result, both yielded substantially similar average responses (P>0.05) in GN and GY, although the variances were several fold higher when using GY *per se* as the selection criterion.

DISCUSSION

Grain yield increase is of paramount importance in every bread wheat breeding program. However, annual increase rates have been steadily declining and would hardly meet the present and future demand. In a recent review, Fischer and Rebetzke (20) proposed focusing on a number of breeding targets for accelerating grain yield genetic gains, among which they included fruiting efficiency/spike fertility index (SF). When evaluating several selection strategies under different environmental conditions, Alonso et al. (12) found that simulated selection for high SF, either solely or along with selection for high grain yield, always resulted in higher and more stable yields than if selecting for high yield alone. In the present study, we show that, regardless of the genetic material or the environmental conditions, selection for high SF alone virtually always increases grain yield in bread wheat.

Independently of the set of experiments, selection for high SF always resulted in GN increases (between 1.6 and 27.4%) and in GY increases in most cases, except for a few in Set III (Table 3). These latter cases, however, occurred in the population with the least number of individuals (N=28/36; Table 3). The top 10% lines with the highest SF values amounted to only 3-4, which could have increased sampling error. Nevertheless, overall results suggest that one can select lines with high SF on the basis of data collected from a very small spike sample and sometimes in a different environment than the one in which yield is determined (even including lower

yield potential locations as it is Marcos Juárez as compared with Balcarce, in this study) (27) and expect, at the minimum, a meager increase in yield and/or GN.

Regarding the substantial variation in the magnitude of the responses to selection observed across sets and experiments within sets (Table 3 and Table S1), several possible underlying causes could be considered. On the one hand, variations might have occurred due to sampling errors, sample size, and different genetic and genetic by environment interaction effects of the traits under evaluation. In this sense, additional data sets comprising a wider spectrum of germplasm and environments/years of evaluation should be incorporated in the analysis in order to reduce these errors. On the other hand, the heritability of each trait as well as the degree of association between the trait used for selection and the response trait might have been other sources of variation in the observed responses to selection. For instance, in the case of the RIL population of Set II, the heritability values of the traits (h^2_{SF}=0.84; h^2_{GY}=0.28; h^2_{GN}=0.35; Alonso et al. 2018) and the correlations between the traits, both genetic ($r_{SF\text{-}GY}$=0.36, $r_{SF\text{-}GN}$=0.55; Alonso et al. 2018) and phenotypic ($r_{SF\text{-}GY}$=0.29, $r_{SF\text{-}GN}$=0.44; Table S2), provide an idea of how these relationships could affect the responses. Spike fertility index, used here as a selection criterion, has high heritability with low environmental influence, which makes it fairly stable across successive generations. The response to selection is expected to be positively correlated with the degree of association between the selection and evaluation traits; indeed, in this study, the response was better in GN than in GY. In turn, the year-to-year variations in the response to selection in these traits are in line with their heritabilities (low to moderate) and with their great sensitivity to environmental effects. Therefore, selecting for high SF would give stability to the response in GN and GY, whereas these latter traits' low/moderate values of both heritability and genetic/phenotypic correlation with SF would increase variability in such response.

Some trade-offs were observed: in general, GW, GPC and TW tended to decrease with selection for high SF. Nevertheless, genotypes with all high SF, GY, GW, GPC and TW were detected (data not shown); therefore, these trade-offs could be avoided by concurrent selection for all traits. Furthermore, all three populations used in the present study had a relatively low number of individuals if considering the scale of an actual breeding program. Provided that the prerequisite of existence of genetic variability for the traits is met (as it was the case in the present study; Fig. 2), increasing population numbers is likely to yield even better results when selecting for high SF. This way, while increasing GY and GN, one could concurrently maintain

desirable GW, TW and GPC values. Phenotypic correlations between variables were generally positive (Table S2) except for the ones between GN and GW (-0.18) and SF and GW (-0.2). These low values, however, confirm the idea that the negative relationship between these variables is not complete.

Independently of the genetic constitution of the breeding material and/or the environmental conditions, simulated selections for high spike fertility index alone, as determined in a small spike sample, appear to result in consistently increased grain number per unit area (and yield) in bread wheat. As virtually similar average responses to selection in GN and GY were attained when using SF or GY *per se* as selection criteria, selection for SF alone has the potential of increasing the throughput and efficiency of the breeding process, while lowering costs. This could translate into a massive screening of lines for SF in small plots and subsequent multi-environment yield trials that only include a subset of the lines initially screened. This turns SF into a very promising selection target in breeding programs aimed at increasing grain yield. Notwithstanding, the effectiveness of carrying out selection for high SF in spaced plants, as it would be the case of early generations of a breeding program, remains yet to be established.

DATA AVAILABILITY

The dataset of the study is available from the authors upon reasonable request.

REFERENCES

1. CIMMYT. Food security. Available from: https://www.cimmyt.org/news/food-security/. Accessed 2019 Sept 15.

2. Abbate P, Andrade F, Lazaro L, Bariffi J, Berardocco H, Inza V, et al. Grain yield increase in recent Argentine wheat cultivars. Crop Sci. 1998;38(5):1203-9. doi:10.2135/cropsci1998.0011183X003800050015x.

3. Borrás L, Slafer GA, Otegui ME. Seed dry weight response to source–sink manipulations in wheat, maize and soybean: a quantitative reappraisal. Field Crops Res. 2004;86(2-3):131-46. doi:10.1016/j.fcr.2003.08.002.

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

5. Ferrante A, Cartelle J, Savin R, Slafer GA. 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. 2017;203:114-27. doi: 10.1016/j.fcr.2016.12.028.

6. Lynch JP, Doyle D, McAuley S, McHardy F, Danneels Q, Black LC, et al. The impact of variation in grain number and individual grain weight on winter wheat yield in the high yield potential environment of Ireland. European journal of agronomy. 2017;87:40-9. doi: 10.1016/j.eja.2017.05.001.

7. Shearman V, Sylvester-Bradley R, Scott R, Foulkes M. Physiological processes associated with wheat yield progress in the UK. Crop Sci. 2005;45(1):175-85. doi:10.2135/cropsci2005.0175.

8. Slafer GA, Andrade FH, Satorre EH. Genetic-improvement effects on pre-anthesis physiological attributes related to wheat grain-yield. Field Crops Res. 1990;23(3-4):255-63. doi: 10.1016/0378-4290(90)90058-J.

9. Fischer R.A. Wheat. In: Smith W, Banks S, editors. Proceedings of symposium on potential productivity of field crops under different environments. IRRI, Los Baños, Philippines1984. p. 129–54.

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

11. Acreche MM, Briceño-Félix G, Sánchez JAM, Slafer GA. Physiological bases of genetic gains in Mediterranean bread wheat yield in Spain. European journal of agronomy. 2008;28(3):162-70. doi: 10.1016/j.eja.2007.07.001.

12. Alonso MP, Mirabella N, Panelo JS, Cendoya MG, Pontaroli AC. Selection for high spike fertility index increases genetic progress in grain yield and stability in bread wheat. Euphytica. 2018;214(7):112. doi: 0.1007/s10681-018-2193-4.

13. Foulkes J, Rivera C, Trujillo E, Sylvester-Bradley R, Reynolds M. Achieving a step-change in harvest index in high biomass wheat cultivars. In: Reynolds M, Molero G, Mollins J, Braun H, editors. Proceedings of the International TRIGO (Wheat) Yield Potential Workshop 2015; 2015 March 24-26; Obregón, Sonora, Mexico: CENEB, CIMMYT; p. 31-5.

14. Terrile II, Miralles DJ, González FG. 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. 2017;201:86-96. doi: 10.1016/j.fcr.2016.09.026.

15. Martino DL, Abbate PE, Cendoya MG, Gutheim F, Mirabella NE, Pontaroli AC. Wheat spike fertility: inheritance and relationship with spike yield components in early generations. Plant Breeding. 2015;134(3):264-70. doi: 10.1111/pbr.1226.

16. Mirabella N, Abbate P, Ramirez I, Pontaroli A. Genetic variation for wheat spike fertility in cultivars and early breeding materials. The Journal of Agricultural Science. 2016;154(1):13-22. doi: 10.1017/S0021859614001245.

17. Abbate PE, Pontaroli AC, Lázaro L, Gutheim F. A method of screening for spike fertility in wheat. The Journal of Agricultural Science. 2013;151(3):322-30. doi: 10.1017/s0021859612000068.

18. Fischer R. Understanding the physiological basis of yield potential in wheat. The Journal of Agricultural Science. 2007;145(2):99. doi: 10.1017/S0021859607006843.

19. Fischer R.A. Wheat physiology: a review of recent developments. Crop and Pasture Science. 2011;62(2):95-114. doi: 10.1071/CP10344.

20. Fischer R.A., Rebetzke G. Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: a review. Crop and Pasture Science. 2018;69(5):439-59. doi: 10.1071/CP17409.

21. González F, Terrile II, Falcón M. Spike fertility and duration of stem elongation as promising traits to improve potential grain number (and yield): variation in modern Argentinean wheats. Crop Sci. 2011;51(4):1693-702. doi: 10.2135/cropsci2010.08.0447.

22. Lázaro L, Abbate PE. Cultivar effects on relationship between grain number and photothermal quotient or spike dry weight in wheat. The Journal of Agricultural Science. 2012;150(4):442-59. doi: 10.1017/s0021859611000736.

23. Foulkes MJ, Slafer GA, Davies WJ, Berry PM, Sylvester-Bradley R, Martre P, et al. Raising yield potential of wheat. III. Optimizing partitioning to grain while maintaining lodging resistance. J Exp Bot. 2010;62(2):469-86. doi: 10.1093/jxb/erq300.

24. García GA, Serrago RA, González FG, Slafer GA, Reynolds MP, Miralles DJ. Wheat grain number: identification of favourable physiological traits in an elite doubled-haploid population. Field Crops Res. 2014;168:126-34. doi: 10.1016/j.fcr.2014.07.018.

25. Slafer GA, Elia M, Savin R, García GA, Terrile II, Ferrante A, et al. Fruiting efficiency: an alternative trait to further rise wheat yield. Food and Energy Security. 2015;4(2):92-109. doi: 10.1002/fes3.59.

26. Alonso M, Abbate P, Mirabella N, Merlos F, Panelo J, Pontaroli A. Analysis of sink/source relations in bread wheat recombinant inbred lines and commercial cultivars under a high yield potential environment. European Journal of Agronomy. 2018;93:82-7. doi 10.1016/j.eja.2017.11.007.

27. Menéndez FJ, Satorre EH. Evaluating wheat yield potential determination in the Argentine Pampas. Agricultural systems. 2007;95(1-3):1-10. doi: 10.1016/j.agsy.2007.03.004.

SUPPORTING INFORMATION

Supporting information (Tables S1 and S2) is available online.

Tables

*Table 1. Description of field experiments carried out in this study, as organized in Sets of exps. (experiments) I to III. Genetic material included in each set, N (number of individuals), exp. name, location, year, exp. design (RCBD=randomized complete block design; AD=augmented design), No. reps. (number of replications), growing conditions, plot dimensions and data collected (SF=spike fertility index; GY=grain yield; GN=grain number per unit area; GW=grain weight; TW=test weight; GPC=grain protein content). Data from exps. BFE13, BFE14 and BFE15 have been partially published by Alonso et al. (*2018 a,b*).*

Set of exps.	Genetic material	N	Exp. name	Location	Year	Exp. design	No. reps.	Growing conditions	Plot dimensions				Data collected					
									Area (m^2)	No. rows	Row length (m)	Inter-row spacing (m)	SF	GY	GN	GW	TW	GPC
I	RIL population (Avalon/Glupro)	90	BAG15	Balcarce	2015	RCBD	2	Rainfed	3.5[a]	3	5	0.2		X	X	X	X	X
		90	BAG16	Balcarce	2016	RCBD	2	Rainfed	3.5[a]	3	5	0.2	X	X	X	X	X	X
II	RIL population (Baguette 10/ Klein Chajá)	146	BFE13	Balcarce	2013	RCBD	2	Irrigated	3.5[a]	3	5	0.2	X	X	X	X	X	
		146	BFE14	Balcarce	2014	RCBD	2	Irrigated	7	7	5	0.2	X	X	X	X	X	
		146	BFE15	Balcarce	2015	RCBD	2	Irrigated	7	7	5	0.2	X	X	X	X	X	
		146	MJFE17	Marcos Juárez	2017	AD	1	Rainfed	0.9	3	1	0.3	X					
III	Advanced breeding lines	28	BLE16	Balcarce	2016	AD	1	Rainfed	7	7	5	0.2	X	X	X			
		28	BLE17	Balcarce	2017	AD	1	Irrigated	7	7	5	0.2	X	X	X			
		36	BLP17	Balcarce	2017	AD	1	Rainfed	7	7	5	0.2	X	X	X	X		
		36	BLE18	Balcarce	2018	RCBD	2	Irrigated	7	7	5	0.2		X				

[a]Inter-plot spacing: 0.3 m

Table 2. Analysis of variance for grain yield across nine experiments. All experiments from Sets I to III (Table 1) were included in the analysis except for MJFE17, which plot size (0.9 m^2) was deemed too small for accurate grain yield determination.

Source of variation	DF	Mean Square	F	P-value
Experiment	8	5155433	530.03	<0.0001
Genotype (set)	236	33324	3.43	<0.0001
Genotype*experiment (set)	499	19126	1.97	<0.0001
Error	693	9727		

Table 3. Response to selection for high spike fertility index (SF) (i.e., selection of the top 10% lines with the highest SF values at each of three sets of experiments carried out between 2013 and 2018 in Balcarce and Marcos Juárez, Argentina) in grain yield (GY), grain number per unit area (GN), grain weight (GW), test weight (TW) and grain protein content (GPC).

Set of experiments	Selection Experiment	Testing Experiment		Trait					
				SF (grains/g)	GY (g/m^2)	GN (grains/m^2)	GW (mg)	TW (kg/hl)	GPC (%)
I	BAG16	BAG16	Population mean (N=90)	70.6	313.7	9760	32.3	77.1	14.5
			Mean of 10% lines (N=9) with highest SF in exp. BAG16	89.7	396.2	12431	31.6	79.0	13.8
			Response to selection for high SF (%)		**26.3**	**27.4**	**-2.0**	**2.5**	**--4.9**
		BAG15	Population mean (N=90)		515.5	12971	39.9	79.9	16.0
			Mean of 10% lines (N=9) with highest SF in exp. BAG16		599.5	15176	39.8	82.4	14.6
			Response to selection for high SF (%)		**16.3**	**17.0**	**-0.4**	**3.2**	**-9.0**
II	MJFE17	BFE13	Population mean (N=146)		719.5	22092	33.0	79.4	
			Mean of 10% lines (N=15) with highest SF in exp. MJFE17		721.0	24313	30.6	77.6	
			Response to selection for high SF (%)		**0.2**	**10.1**	**-7.2**	**-2.2**	
		BFE14	Population mean (N=146)		367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest SF in exp. MJFE17		419.0	12592	34.1	72.3	
			Response to selection for high SF (%)		**14.0**	**26.2**	**-8.8**	**-2.0**	
		BFE15	Population mean (N=146)		749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest SF in exp. MJFE17		752.0	19387	39.3	79.4	
			Response to selection for high SF (%)		**0.4**	**9.1**	**-7.8**	**-1.2**	
	BFE13	BFE13	Population mean (N=146)	98.4	719.5	22092	33.0	79.4	
			Mean of 10% lines (N=15) with highest SF in exp. BFE13	112.8	772.4	25615.5	31.9	79.3	
			Response to selection for high SF (%)		**7.3**	**16.0**	**-3.5**	**-0.13**	
		BFE14	Population mean (N=146)		367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest SF in exp. BFE13		397.2	10811	37.5	74.0	

Set of experiments	Selection Experiment	Testing Experiment		Trait					
				SF (grains/g)	GY (g/m^2)	GN ($grains/m^2$)	GW (mg)	TW (kg/hl)	GPC (%)
			Response to selection for high SF (%)		**8.1**	**8.3**	**0.3**	**0.3**	
		BFE15	Population mean (N=146)		749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest SF in exp. BFE13		790.8	19626	40.4	79.7	
			Response to selection for high SF (%)		**5.6**	**10.5**	**-5.0**	**-0.9**	
	BFE14	BFE13	Population mean (N=146)		719.5	22092	33.0	79.4	
			Mean of 10% lines (N=15) with highest SF in exp. BFE14		763.0	24166	31.8	78.3	
			Response to selection for high SF (%)		**6.0**	**9.1**	**-3.6**	**-1.4**	
		BFE14	Population mean (N=146)	92.0	367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest SF in exp. BFE14	109.4	430.5	12259	35.4	73.5	
			Response to selection for high SF (%)		**17.2**	**22.9**	**-5.5**	**-0.4**	
		BFE15	Population mean (N=146)		749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest SF in exp. BFE14		778.7	19310	40.5	79.4	
			Response to selection for high SF (%)		**4.0**	**8.7**	**-4.9**	**-1.1**	
	BFE15	BFE13	Population mean (N=146)		719.5	22092	33.0	79.4	
			Mean of 10% lines (N=15) with highest SF in exp. BFE15		745.8	24528	31.4	78.8	
			Response to selection for high SF (%)		**3.6**	**11.0**	**-4.8**	**-0.8**	
		BFE14	Population mean (N=146)		367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest SF in exp. BFE15		408.5	12294	33.8	72.9	
			Response to selection for high SF (%)		**11.2**	**23.2**	**-9.6**	**-1.2**	
		BFE15	Population mean (N=146)	89.4	749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest SF in exp. BFE15	103.9	783.4	20006	39.5	79.5	
			Response to selection for high SF (%)		**4.6**	**12.6**	**-7.3**	**-1.1**	
III	BLE16	BLE16	Population mean (N=28)	81.5	498.9	14755			
			Mean of 10% lines (N=3) with highest SF in exp. BLE16	92.9	457.1	15962			

Set of experiments	Selection Experiment	Testing Experiment		Trait					
				SF (grains/g)	GY (g/m^2)	GN (grains/m^2)	GW (mg)	TW (kg/hl)	GPC (%)
			Response to selection for high SF (%)		**-8.4**	**8.2**			
		BLE17	Population mean (N=28)		449.6	11413			
			Mean of 10% lines (N=3) with highest SF in exp. BLE16		445.1	12122			
			Response to selection for high SF (%)		**-0.1**	**6.2**			
	BLE17	BLE17	Population mean (N=28)	84.5	449.6	11413			
			Mean of 10% lines (N=3) with highest SF in exp. BLE17	98.3	424.6	11600			
			Response to selection for high SF (%)		**-5.5**	**1.6**			
		BLE16	Population mean (N=28)		498.9	14755			
			Mean of 10% lines (N=3) with highest SF in exp. BLE17		565.8	17157			
			Response to selection for high SF (%)		**13.4**	**16.3**			
	BLP17	BLP17	Population mean (N=36)	86.6	383.9	11015	35.3		
			Mean of 10% lines (N=4) with highest SF in exp. BLP17	104.5	367.4	12195	30.3		
			Response to selection for high SF (%)		**-4.3**	**10.7**	**-14**		
		BLE18	Population mean (N=36)		609.6				
			Mean of 10% lines (N=4) with highest SF in exp. BLP17		566.6				
			Response to selection for high SF (%)		-7				
			Overall average response to selection (%)		**5.6**	**13.4**	**-5.3**	**-0.6**	**-7.0**

Table 4. T test for comparison of the average response to selection in grain number per unit area (GN) and in grain yield (GY) attained after selection for high spike fertility index (SF; Table 3) vs. for high GY per se *(Table S1).*

Variable	Selection criterion	N	Response to selection (%)		DF	T	P-value
			Mean	Variance			
GN	SF	19	13.4	50.6	24	-1.52	0.14
	GY *per se*	19	20.1	319.1			
GY	SF	20	5.6	80.1	14	-1.34	0.20
	GY *per se*	12	12.9	308.1			

Figures

Figure 1

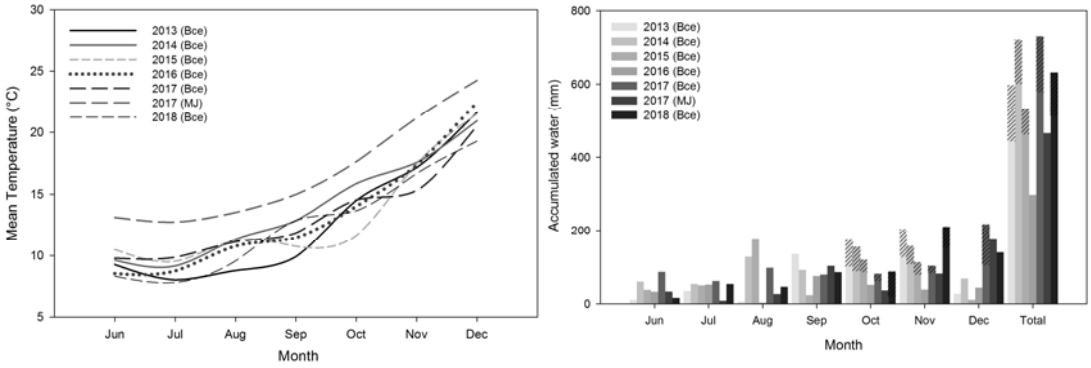

Figure 2.

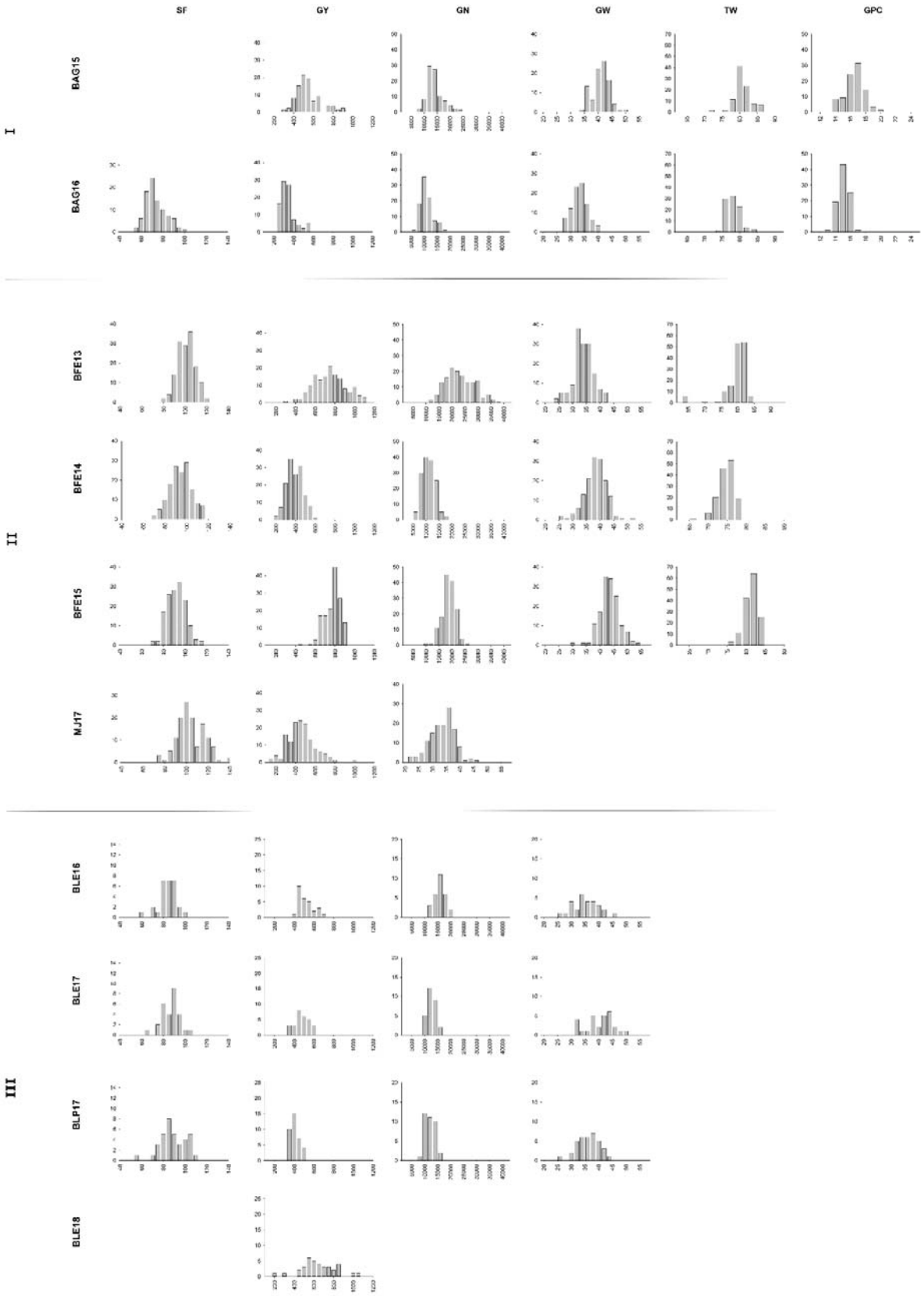
SF
GY
GN
GW
TW
GPC
I
BAG15
BAG16
II
BFE13
BFE14
BFE15
MJ17
III
BLE16
BLE17
BLP17
BLE18

Figure legends

Fig. 1. Mean temperature (left) and accumulated rainfall plus irrigation (right) during the June-December period (wheat growing season) at Balcarce (Bce) in 2013 to 2018 and in Marcos Juárez (MJ) in 2017. Stripped bars show accumulated water by irrigation.

Fig. 2. Histograms of spike fertility index (SF, grains/g), grain yield (GY, g/m^2), grain number per unit area (GN, grains/m^2), grain weight (GW, mg), test weight (TW, kg/hl) and grain protein content (GPC, %) of I: Avalon/Glupro RIL population (N=90, experiments BAG15 and BAG16); II: Baguette 10/Klein Chajá RIL population (N=146, experiments BFE13, BFE14, BFE15 and MJFE17) and III: advanced breeding lines (N=28, experiments BLE16 and BLE17; N=36, experiments BLP17 and BLE18).

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

AUTHOR CONTRIBUTIONS

ACP, MPA and NEM designed the study; ACP, MPA, NEM, JSP, MFF, and ML performed experiments at Balcarce and LSV performed the experiment at Marcos Juárez; MPA, NEM and JSP processed SF samples; NEM, MPA and ACP analyzed the data; MFF prepared all figures of the manuscript; ACP wrote the paper with input from all authors.

Table S1. Response to selection for high grain yield (GY) (i.e., selection of the top 10% lines with the highest GY values at each of three sets of experiments carried out between 2013 and 2018 in Balcarce and Marcos Juárez, Argentina) in spike fertility index (SF), grain yield (GY), grain number per unit area (GN), grain weight (GW), test weight (TW) and grain protein content (GPC).

				Trait					
Set of experiments	**Selection experiment**	**Testing Experiment**		**SF** (grains/g)	**GY** (g/m^2)	**GN** (grains/m^2)	**GW** (mg)	**TW** (kg/hl)	**GPC** (%)
I	BAG15	BAG15	Population mean (N=90)			12966	39.9	79.9	16.0
			Mean of 10% lines (N=9) with highest GY in exp. BAG15			19598	40.4	83.5	14.2
			Response to selection for high GY (%)			**51.2**	**4.6**	**4.6**	**-11.1**
		BAG16	Population mean (N=90)	70.6	313.7	9788	32.3	77.15	14.5
			Mean of 10% lines (N=9) with highest GY in exp. BAG15	81	458.2	14257	32.4	79.4	13.4
			Response to selection for high GY (%)	**14.2**	**46.0**	**45.7**	**0.3**	**2.9**	**-8.1**
	BAG16	BAG16	Population mean (N=90)	70.6		9788	32.3	77.1	14.5
			Mean of 10% lines (N=9) with highest GY in exp. BAG16	82.2		14496	33.9	80	13.3
			Response to selection for high GY (%)	**16.5**		**48.1**	**5.0**	**3.7**	**-8.5**
		BAG15	Population mean (N=90)		515.5	12971	39.9	79.9	16.0
			Mean of 10% lines (N=9) with highest GY in exp. BAG16		743.7	17752	42.3	837	14.3
			Response to selection for high GY (%)		**44.3**	**36.9**	**5.9**	**4.8**	**-10.7**
II	BFE13	BFE13	Population mean (N=146)	98.4		22092	33	79.4	
			Mean of 10% lines (N=15) with highest GY in exp. BFE13	100.2		30890	32.8	76.6	
			Response to selection for high GY (%)	**1.8**		**39.4**	**-0.7**	**-3.5**	
		BFE14	Population mean (N=146)	92	367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest GY in exp. BFE13	94.1	444	11714	38.3	74.1	
			Response to selection for high GY (%)	**2.3**	**20.9**	**17.4**	**2.4**	**0.5**	
		BFE15	Population mean (N=146)	89.4	749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest GY in exp. BFE13	90.8	772.5	18522	42.1	80.1	
			Response to selection for high GY (%)	**1.6**	**3.1**	**4.3**	**-1.1**	**-0.4**	
	BFE14	BFE13	Population mean (N=146)	98.4	719.5	22092	33	79.4	

Set of experiments	Selection experiment	Testing Experiment		Trait					
				SF (grains/g)	GY (g/m^2)	GN (grains/m^2)	GW (mg)	TW (kg/hl)	GPC (%)
			Mean of 10% lines (N=15) with highest GY in exp. BFE14	102.7	903	29246	32.8	76.6	
			Response to selection for high GY (%)	**4.3**	**25.5**	**6.9**	**-0.7**	**-3.5**	
		BFE14	Population mean (N=146)	92		9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest GY in exp. BFE14	101		13918	37	73.9	
			Response to selection for high GY (%)	**9.9**		**39.5**	**-1.5**	**0.2**	
		BFE15	Population mean (N=146)	89.4	749.1	17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest GY in exp. BFE14	95.2	759.5	18746	40.6	79.2	
			Response to selection for high GY (%)	**6.5**	**1.4**	**5.5**	**-4.7**	**-1.5**	
	BFE15	BFE13	Population mean (N=146)	98.4	719.5	22092	33	79.4	
			Mean of 10% lines (N=15) with highest GY in exp. BFE15	103.5	771.8	23610	33.4	80	
			Response to selection for high GY (%)	**5.2**	**7.2**	**6.9**	**1.1**	**0.8**	
		BFE14	Population mean (N=146)	92	367.4	9978	37.4	73.8	
			Mean of 10% lines (N=15) with highest GY in exp. BFE15	96.1	393	10767	37.1	74.5	
			Response to selection for high GY (%)	**4.5**	**7.0**	**7.9**	**-0.9**	**1.0**	
		BFE15	Population mean (N=146)	89.4		17763	42.6	80.3	
			Mean of 10% lines (N=15) with highest GY in exp. BFE15	95		20816	41.9	80.5	
			Response to selection for high GY (%)	**6.3**		**17.2**	**-1.5**	**0.2**	
III	BLE16	BLE16	Population mean (N=28)	81.5		14755			
			Mean of 10% lines (N=3) with highest GY in exp. BLE16	84.3		17905			
			Response to selection for high GY (%)	**3.4**		**21.3**			
		BLE17	Population mean (N=28)	84.5	449.6	11413			
			Mean of 10% lines (N=3) with highest GY in exp. BLE16	89.3	407.8	10638			
			Response to selection for high GY (%)	**5.7**	**-9.3**	**-6.8**			
	BLE17	BLE17	Population mean (N=28)	84.5		11413			
			Mean of 10% lines (N=3) with highest GY in exp. BLE17	80.5		13563			
			Response to selection for high GY (%)	**-4.8**		**18.8**			

Set of experiments	Selection experiment	Testing Experiment		Trait					
				SF (grains/g)	GY (g/m^2)	GN ($grains/m^2$)	GW (mg)	TW (kg/hl)	GPC (%)
		BLE16	Population mean (N=28)	81.5	498.9	14755			
			Mean of 10% lines (N=3) with highest GY in exp. BLE17	82.8	514	15455			
			Response to selection for high GY (%)	**1.6**	**3.0**	**4.7**			
	BLP17	BLP17	Population mean (N=36)	86.6		11015	35.3		
			Mean of 10% lines (N=4) with highest GY in exp. BLP17	77.4		12793	36.7		
			Response to selection for high GY (%)	**-10.6**		**16.1**	**4.0**		
		BLE18	Population mean (N=36)		609.6				
			Mean of 10% lines (N=4) with highest GY in exp. BLP17		639.1				
			Response to selection for high GY (%)		**4.8**				
	BLP18	BLP17	Population mean (N=36)	86.6	383.9	11015	35.3		
			Mean of 10% lines (N=4) with highest GY in exp. BLP18	84.9	389.4	11165	35.3		
			Response to selection for high GY (%)	**-1.9**	**1.4**	**1.4**	**0.0**		
			Overall average response to selection (%)	**3.9**	**12.9**	**20.1**	**0.8**	**0.8**	**-9.6**

Table S2. Pearson correlation coefficients between the variables analyzed in this study. All coefficients are significant (p<0.05) except where noted (ns).

	GY	GN	GW	TW
SF	0.29	0.44	-0.20	0.00$_{ns}$
GY		0.91	0.24	0.28
GN			-0.18	0.19
GW				0.21
TW				