

# Genetic Similarity Between Donor And Recurrent Parents Can Reduce The Number of Backcross Generation In Marker-Assisted Backcross

**Elton Otacílio de Souza**

Longping High-Tech

**Gustavo Vitti MÔro** (✉ [gv.moro@unesp.br](mailto:gv.moro@unesp.br))

São Paulo State University (Unesp) <https://orcid.org/0000-0001-7709-7814>

**Gabriel Belfort Rodrigues**

Longping High-Tech

**Ivan Schuster**

Longping High-Tech

---

## Research Article

**Keywords:** Zea mays L, marker-assisted backcross, single nucleotide polymorphism markers, genetic diversity, trait introgression

**Posted Date:** September 27th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-915422/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Euphytica on April 8th, 2022. See the published version at <https://doi.org/10.1007/s10681-022-03009-9>.

# Abstract

The marker assisted backcross (MABC) method is the most used to obtain transgenic hybrid and transgenic inbred lines in maize with few backcrosses (BC) generations. It is possible that the use of donor parents with greater genetic similarity with recurrent parents can further reduce the number of BC generation to recover recurrent parent genome. The objective of this study was to evaluate the influence of genetic distance between parents and the percentage of recurrent parent genome recovery, as well as the similarity of BC plants and recurrent parent. Nine maize BC populations were evaluated, with genetic distances between donors and recurrent parents ranging from 0.238 to 0.499. In the backcross generations, molecular markers were used to identify the plants with the highest percentage of recurrent genome recovery and with greater similarity to the recurrent parent. There was no difference in the recurrent genome recovery among populations after three BC generations. In the first two BC generation the similarity between BC plants and recurrent parents was positively correlated with the similarity between donor and the populations recurrent parent. BC populations with higher similarity among parents could be finished with two BC generations, and BC populations with lower similarity could just be finished after three generation of MABC. The use of donor parents with higher similarity with recurrent parent can reduce one BC generation in MABC approach.

## Introduction

The backcross (BC) method aims to introgression one or few genes in a variety or elite inbred line, with subsequent recovery of the recurrent parent genome. The method is usually used to correct elite genotypes in the traits in which they are deficient, by crossing with genotypes with traits that are desired to introduce, called donor genotypes. More recently, the BC method has been used for the introduction of transgenic events in conventional elite genotypes.

BC method is the main method used to introduce transgenic events in corn. In 2019, 16.3 million ha of transgenic corn were planted in Brazil, corresponding to 93% of the corn area in the country (Isaaa 2019). This demonstrates the importance of BC method in corn breeding programs.

One of the limitations of BC method is the long time necessary to complete the process. In some cases, the elite genotype used as a recurrent parent becomes obsolete at the end of the BC program (Mesquita et al. 2005). Using molecular markers to select the plants with the highest proportion of recurrent genome in each BC generation can reduce the number of BC generations (Hospital 2001; Bouchez et al. 2002; Wang et al 2020).

The probability of success of a BC program depends, in large part, on the amount of germplasm from the donor parent that needs to be eliminated in the BC process. Thus, the use of donor parents with more genetic similarity to recurrent parents could, in theory, facilitate the recurrent genome recovery in these programs. The genetic similarity between donor and recurrent parents can mean that some chromosomal segments in the potential donor can be identical by descent or at least identical in state with the recurrent

parent. These similar chromosomal segments, in essence, don't need to be recovered in BC projects. With less background genome to be recovered, it is possible to accelerate the recurrent parent germplasm recovery (Peng et al. 2013).

Although the advantage of using donor parents genetically more similar to recurrent parents can be demonstrated theoretically or by simulation, there are no studies using real data that prove this advantage. The present study aimed to evaluate the influence of the genetic distance between donor and recurrent parents in the recurrent genome recovery and in genetic similarity between backcrossing plants and the recurrent parent, in a MABC program.

## Material And Methods

Nine BC populations conducted in the LongPing High-Tech research unit in Cravinhos-SP were evaluated for the realization of this study. To avoid conflicts of interest by using transgenic traits, the target gene was a corn native gene (*Rf4*). But the results represent what happen when considering a transgene and are, therefore, directly applied to backcross programs for introgression of transgenic events.

All populations were obtained with the same donor parent (DP1) containing the male fertility restoration gene *Rf4*. The nine recurrent parents (RP) had the recessive allele *rf4*, and were named RP1, RP2, RP3, RP4, RP5, RP6, RP7, RP8 and RP9. The recurrent parents and their BC populations were classified in three groups: Group1: RP genetically close to DP; Group2: RP with intermediate similarity/distance to DP; and Group3: RP genetically distant to DP. In each group three populations were evaluated (Table 1).

Initially the crosses between donor parent with all recurrent parents were performed to obtain the F<sub>1</sub> generation. The plants of the F<sub>1</sub> generation were crossed again with the respective recurrent parents. In BC<sub>1</sub>, BC<sub>2</sub> and BC<sub>3</sub> generations, plants of each population were initially selected based on the analysis of the presence of *Rf4* gene using a specific molecular marker for this gene. The plants containing the *Rf4* gene were genotyped with 1,065 single nucleotide polymorphism (SNP) markers to estimate the recurrent genome recovered and the genetic similarity with recurrent parents. In BC<sub>1</sub> and BC<sub>2</sub> generations, the five plants that possessed the highest genetic similarity with the recurrent parents, after selection for *Rf4* gene, were crossed again with the respective recurrent parents. In BC<sub>3</sub> generation the five selected plants were used for the finalization of the converted inbred lines.

For DNA extraction one 5 mm diameter leaf disk was collected from each plant. The DNA was extracted using the Fast ID Genomic kit (Genetic ID – Fairfield, Iowa, USA), following the instructions of the manufacturer. Initially, the DNA samples were used for the selection of the *Rf4* gene, using specific primers and probes, and the TaqMan® method (Holland et al. 1991). The plants containing the *Rf4* gene were genotyped with a panel of 1,065 SNP markers using tGBS (target Genotyping by Sequence) approach. For this genotyping, the Ion GeneStudio S5 System platform (Thermo Fisher – Waltham, Massachusetts, USA) was used, following the manufacturer's instructions. DNA samples from donors and recurrent parents of each population was also genotyped.

Once in the genotyping panel there were monomorphic and polymorphic markers between donor and recurrent parents, two comparison parameters between BC generations plants and recurrent parents were made: the genetic similarity and the recurrent genome recovery. The genetic similarity (S) was estimated using the expression:

$$S = \frac{\text{Number of homozygous SNPs equal to the recurrent parent}}{\text{total number of SNPs}}$$

The recurrent genome recovery in each BC plant was estimated in the same way as genetic similarity but using only the polymorphic SNP markers between donor and recurrent parents. In this case it is considered that the recurrent and donor parents have zero similarity. The recurrent genome recovery between BC<sub>1</sub>, BC<sub>2</sub> and BC<sub>3</sub> plants and the recurrent parents, based only on polymorphic markers, represents the amount of recurrent genome that was divergent between donor and recurrent parents and that was already recovered in these plants. The similarity and the recurrent genome recovery were expressed in percentage, multiplying the estimated value by 100.

The genetic similarity analyses were performed with the aid of the Tassel 5 program (Bradbury et al. 2007). The three groups of population (Table 1) were compared by Analysis of Variance. The genetic similarity and the recurrent genome recovery were used as variables. The three groups were the treatments and individual population within groups where the repetitions in a completely randomized model. The averages of each group were compared by the Tukey test with significance of 5%. The relationship of genome recovery or genetic similarity of BC plants and the genetic similarity between the RP and DP was made by using Pearson correlation. The correlations significances were evaluated by Mantel test with 5,000 simulations, and t-test. The Analyses of Variance, average test and correlations were performed with the aid of the Genes software (Cruz 1997).

## Results

### *Recurrent genome recovery*

After the first BC generation the average of recurrent genome recovery of all BC<sub>1</sub> populations was 78.29%. In group 1 the average of recurrent genome recovery was on 78.49%, in group 2 77.36% and in group 3 79.03% (Figure 1). The difference in the recurrent genome recovery observed between groups 2 and 3 was significant.

By the backcross theory, in the BC<sub>1</sub> generation the proportion of the recurrent recovered genome is expected to be 75% (Borem and Miranda 2005) for all populations. The proportion expected by the backcrossing theory is based on infinite size populations, and deviations from that expected can be observed according to the number of samples evaluated. In the BC<sub>1</sub> generation, as there was no selection, these differences from the expected may have been observed due to the sample size. As real

populations are always limited in size regarding to theoretical populations or ideal populations, these deviations could be expected.

The difference in the proportion of the recurrent genome recovery in groups 2 and 3 was significant, but the difference of this proportion between groups 1 and 3 was not significant, indicating that there is no relationship between the initial genetic similarity between parents and the proportion of the recurrent genome recovery in the BC<sub>1</sub> generation. This is also demonstrated by the lack of significant correlation between the average of recurrent genome recovery of BC<sub>1</sub> plants with the genetic distances among parents in each population (Table 2). The deviation of the recurrent genome proportion observed in group 2 regarding group 3 can be associated with the lowest number of plants evaluated in group 2 (Table 1).

The five plants of each BC<sub>1</sub> population with the highest recurrent parent genome recovery was selected for the next backcrossing generation. Considering only the selected plants, the average of recurrent genome recovery mean among all populations was 86.73%, and in group 1 the average was 87.29%, in group 2 was 85.38% and, in group 3, the average was 87.51% (Figure 1). The difference between the recurrent genome recovery from the selected plants and the average of recurrent genome recovery observed in whole population is the effect of using molecular markers in MABC to accelerate genome recovering in BC program.

The average of recurrent genome recovery in the selected plants from group 1 and group 3 differed significantly from the average of recurrent genome recovery of the plants selected from group 2. This difference is not related to the genetic distances among the parents, since the correlation between the genetic distance and the recurrent genome recovery was not significant (Table 2). The total number of plants evaluated by markers on group 2 was 253, which is 5.6% lower than group 1 and 8.6% lower than group 3. As the number of selected plants was fixed in each population, there were different selection intensities. The lowest proportion of recurrent genome in group 2 is related to the lowest selection intensity and the lowest average of the recurrent genome on this group (Figure 1).

In the BC<sub>2</sub> generation the average of recurrent genome recovery among all populations was 92.57%. In group 1 the average was 93.57%, in group 2 was 91.86%, and in group 3 was 92.27% (Figure 1). The differences of the average of recurrent genome recovery among the three groups in the BC<sub>2</sub> generation were not significant. The correlation between the average of recurrent genome recovery in the BC<sub>2</sub> generation and the genetic distances among parents were also not significant (Table 5).

The average of recurrent genome recovery in the BC<sub>2</sub> generation (92.57%) was only slightly lower than it would be expected for the BC<sub>3</sub> generation (93.75%). This was possible because the plants used for the BC that gave origin to the BC<sub>2</sub> (BC<sub>1</sub> selected plants) already had recovery genome proportion equivalent to the expected for the BC<sub>2</sub> generation (Table 3).

The average of recurrent genome recovery in the five selected plants of all populations was 96.18% in BC<sub>2</sub>. In group 1 the average of recovery genome in the selected plants was 96.56%, in group 2 was

96.00% and in group 3 was 95.99% (Figure 1). The differences of the average of recurrent genome recovery among the three population groups were not significant. The correlation of the recurrent genome recovery with the genetic distances among the parents was also not significant (Table 2).

In the BC<sub>3</sub> generation the average of recurrent genome recovery in plants of all populations was of 96.97%. In group 1 it was 96.90%, in group 2 was 96.89% and, in group 3 was 97.14% (Figure 1). The differences among these averages were not significant and the correlation between the recurrent genome recovery and the genetic distances among the parents was also not significant (Table 2).

Taking only the five plants with the highest recurrent genome recovery of each population, the average of proportion of the recurrent genome recovery was 98.92% in BC<sub>3</sub>. In group 1 the average of recurrent genome recovery was 98.69%, in group 2 was 98.77%, and in group 3 was 99.29%. These averages was not statistically different and were not correlated with the genetic similarities among the parents.

In the three groups of populations, it is possible to observe a great variability in the recurrent genome recovery among plants in the BC<sub>1</sub> generation, and gradual decrease of variability as they advance the BC generations (Figure 1). This variability among plants within the BC populations is the basis for the markers assisted selection in the BC programs to accelerate the recurrent genome recovery (Guimarães et al. 2016, Wang et al. 2007, Peng et al. 2013).

#### *Genetic similarity with recurrent genome*

The recurrent genome recovery implies the amount of the genome that was divergent between the donor and recurrent parents and that became similar to the recurrent genitor genome in each BC generation (Peng et. al. 2013), expressed in percentage. In this case, only the divergent part of the genome is considered in the similarity evaluation, which means to say that the original similarity between the donor and recurrent parents was zero. For this reason, it is not expected that the genetic distance among the parents has influence in the recurrent genome recovery, as was observed in this study.

However, it is to be expected that if the recurrent and donor parents are genetically more similar, the total donor genome recovery, and not percentage, could be faster. Nonetheless, to observe this, it is necessary to evaluate the genetic similarity between the BC plants and the recurrent parent considering the whole genome, and not just the proportion of the genome that is different between recurrent and donor parents. For this reason, it was also estimated in each backcrossing generation the genetic similarity of the plants with the recurrent parent considering the whole genome, including the monomorphic markers.

After the first backcross generation, the average of genetic similarity between the plants of all BC<sub>1</sub> populations with the recurrent parents was 92.15%. In BC<sub>1</sub>, the average of genetic similarity from group 1 with the recurrent parents was 94.01%, from group 2 was 92.34% and from group 3 was 90.10% (Figure 2). The differences among the average of genetic similarity from each group were statistically significant, demonstrating the genetic similarity effect between the parents and the genetic similarity of plants in BC<sub>1</sub>

with the recurrent parent (Figure 2). The correlation between the average of genetic similarity of each BC<sub>1</sub> population and the genetic similarity between the donor and recurrent parents was 0.97 (Table 2).

Considering only the five selected plants in all populations, the average of genetic similarity with the recurrent parents was 95.11% in BC<sub>1</sub> (Figure 2). In group 1 the average of similarity was 96.42%, in group 2 was of 94.98% and in group 3 was of 93.92%. Considering only the selected plants, the average of similarity of group 1 populations with the recurrent parents was statistically superior to the average of similarity in groups 2 and 3. The correlation between average of similarities of the selected plants in each population with the genetic similarities among the parents of each backcrossing population was 0.917.

In the BC<sub>2</sub> generation the average of genetic similarity of all populations with the recurrent parents was 97.15%. In group 1 the average of similarity was 98.22%, in group 2 was 97.08%, and in group 3 was 96.14% (Figure 2). In BC<sub>2</sub> generation, the average of similarity was statistically different only between group 1 and group 3 (Figure 2). After two BC generation the similarity between the BC plants and the recurrent parents became closer between the groups. BC populations from group 3 were equal to group 2 populations, and BC populations from group 2 were equal group 1. In BC<sub>2</sub> generation the genetic similarity between the BC plants and recurrent parents were also correlated positively ( $r=0.94$ ) with the genetic similarity among the parents (Table 2).

Among the five plants with higher similarity the recurrent parent plants in each BC<sub>2</sub> population, the average of similarity was 98.40. In group 1 the average of similarity was 98.95%, in group 2 was 98.48% and, in group 3, was 97.77% (Figure 2). Among the selected plants there was also significant difference between the similarity of BC plants with the recurrent parents between groups 1 and 3. The correlation between the similarity of these selected plants with the recurrent parents and the genetic similarities among the parents of each population was not significant (Table 2).

In the BC<sub>3</sub> generation the average of genetic similarity from all populations with the recurrent parents were 98.61%. In group 1 average of similarity was 98.84%, in group 2 was 98.60%, and in group 3 was 98.39% (Figure 2). The differences among this average were not significant, and the correlation between BC the similarities of plants and the recurrent parents with the genetic similarity among the population parents was also not significant (Table 3).

Considering the five selected plants in each population, the average of similarity with the recurrent parents was 99.44%. In group 1 this average was 99.40%, in group 2 was 99.37%, and in group 3 the average of similarity with the recurrent parents were 99.55%. These average of similarity with the recurrent parents did not differ statistically and were not correlated with the genetic similarity among the parents.

## Discussion

### *Genome recovery parameter*

Using plants with the highest proportion of the recurrent genome recovery to generate the next BC generation, the proportion of the recurrent genome recovery in the next generation will be higher than that expected, and with this the number of BC generations for the recurrent genome recovery can be lower (Hospital 2001; Bouchez et al. 2002; Wang et al 2020). Once the variability within the populations is has been reduced at each BC generation, it is possible to use fewer plants in the initial BC generations and more plants in the final generations, for the optimization of resources associated with higher recurrent genome recovery.

In this study the five plants with the highest proportion of genome recovery, in each population and in each generation, were selected to be used in the next BC generations. The number of selected plants depends on the number of seeds that is desired to obtain, and this number is related to the number of genes to be selected before the selection of the recurrent genome recovery. In a backcrossing program, for each additional gene to be selected for introgression in the elite germplasm the population size is multiplied by two. In addition, it is necessary to consider some more plants in the case of there is any problem with the selected plants, and these do not produce the number of seeds expected.

The recurrent genome recovery in the BC<sub>1</sub> generation was slightly superior to the theoretically expected. The selected plants with the use of molecular markers in the BC<sub>1</sub> generation already had the proportion of the recurrent genome recovery equivalent to that that would be expected in the BC<sub>2</sub> generation (Table 3). Consequently, the average of recurrent genome recovery in the BC<sub>2</sub> generation was already equivalent to the expected in the BC<sub>3</sub> generation. The selected plants in the BC<sub>2</sub> generation already presented recurrent genome equivalent to the expected in the BC<sub>4</sub> generation. The average of recurrent genome recovery in the BC<sub>3</sub> generation did not differ much from the average of selected plants in the BC<sub>2</sub> generation, but the average of recovery of the BC<sub>3</sub> selected plants was closer to the expected in BC<sub>6</sub> than that expected in BC<sub>5</sub> (Table 3).

In average, the plants selection with the highest proportion of recurrent genome in each generation equals the gain in the recurrent genome recovery of an additional BC generation. Thus, each marker-assisted BC generation is equivalent to two BC conventional generations. In this study, the selected plants in BC<sub>1</sub> were equivalent to the BC<sub>2</sub> generation, the selected plants in BC<sub>2</sub> were equivalent to the BC<sub>4</sub> generation and the selected plants in BC<sub>3</sub> were equivalent to the BC<sub>6</sub> generation (Table 3).

All the evaluated populations could be finalized with three marker-assisted backcrossing generations, with more than 98% of recurrent genome recovery. Seven of the nine evaluated populations presented at least one plant in the BC<sub>3</sub> generation with more than 99% of the recurrent genome recovery. The exceptions were the RP1xDP1 population of group 1, which had maximum recovery of 98.28%, and the DP6xRP1 population of group 2, which had maximum recovery of 98.25%.

The variability within the BC populations reduced rapidly from one BC generation to another (Figure 2). In BC<sub>1</sub> generation, without selection, the populations derived from more similar parents presented lower variability than the populations derived from less similar parents. The selection by molecular markers contributed to reduce the variability within BC populations, because among the selected plants the variability is lower, with higher average. Because there is greater variability among the plants derived from less similar parents, the gain by markers assisted selection is higher in these populations. In BC<sub>2</sub> generation the similarity between BC plants and recurrent parents was closer among the groups, and in the BC<sub>3</sub> generation the genetic similarity among BC plants was equal for all the populations, independent of the genetic similarity among the parents.

By the results observed in both, the recurrent genome recovery and in the evaluation of genetic similarity between the BC plants and the recurrent parents, it was possible to recover the recurrent parent genotype with three marker-assisted BC generations. The genetic similarity of the selected BC plants at the end of three MABC generations was superior to 99%, and the recurrent genome recovery close to 99%, independent of the genetic similarity among the parents.

In all three populations from group 1, we obtained plants with genetic similarity to recurrent parent above 99% with just two generations of BC. The lowest similarity considering each five selected plants individually in these three populations was 98.74%. These three populations could have been finalized in the BC<sub>2</sub> generation.

Two of the three populations from group 2 also had average similarity of the five selected plants with the recurrent parent above 98.5% and could be finalized with two BC generations. However, one population from group 2 and all three populations from group 3 could not be finished on BC<sub>2</sub>. This demonstrates that the use of donor parents with high genetic similarity with recurrent parents can reduce the number of BC generations from three to two in a MABC program.

Considering the evaluation of recurrent genome recovery, taking in account just the percentage of divergent genome recovered in each BC generation, it was not possible to observe any difference among the groups. And there could be no recommendation of finalizing some populations after two BC generations, even with some populations already having plants very close to recurrent parents.

In the recurrent genome recovery evaluation, it is considered only the genome regions that is contrasting among the parents used in the initial crossing. However, the recurrent genome recovery evaluation without considering the genetic similarity based on the evaluation of the whole genome, and not only in the contrasting regions, can lead to interpretation errors, and the result of the BC program may not be as expected.

Considering the example of two BC populations, one with genetic similarity among the parents of 0.80 and another with genetic similarity of 0.50. Also considering the corn genome with around 1,700 cM. The parents of the first population differ in 340 cM, while the parents of the second population differ in 850

cM. When the first population has recovered 98% of the recurrent genome, it will have recovered 332.2 cM, remaining 6.8 cM of the donor parent in the converted plants. When the second population has recovered 98% of the recurrent genome, it will have recovered 833 cM, remaining 17 cM of the donor parent in these plants, i.e., 2.5 times more donor parent germplasm than in the first population.

The genetic similarity evaluation between the BC populations plants with recurrent genitor offers more accurate information for the BC programs management and defining the moment to end the BC cycles and to finish the inbred line conversion. For this, it is necessary to use a number of markers that acknowledge the whole genome, so that the genetic similarities are estimated with high accuracy.

The use of donor parents genetically closer to the recurrent parents can reduce the number of BC generations necessary to recover the amount of recurrent desired genome, as it is the case of group 1 populations used in this study.

With three marker-assisted BC generations the effect of the genetic distance between the donor and recurrent parents is nullified, and all the populations can reach the genetic similarity close to or above 99% with the recurrent parent.

Even considering genetic similarity close to 99%, it is important to observe that there can still be significant differences among the BC plants and the recurrent parent, because 1% of the corn genome can contain hundreds of genes. Therefore, it is important to maintain finishing procedure via conventional breeding techniques before choosing the best version for each converted genotype (Mumm and Walters 2001; Mumm 2013).

## Declarations

*Funding:* Not applicable

*Conflicts of interest/Competing interest:* Not applicable

*Availability of data and material:* Data is available, but material is not available

*Code availability:* Not applicable

## References

1. Borem A, Miranda GV (2005) Melhoramento de Plantas. Editora UFV, Viçosa
2. Bouchez A, Hospital F, Caussee M, Gallais A, Charcosset A (2002) Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. *Genetics* 162:1945–1959. <https://doi.org/10.1093/genetics/162.4.1945>
3. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635.

<https://doi.org/10.1093/bioinformatics/btm308>

4. Cruz CD (1997) Programa Genes - aplicativo computacional em genética e estatística. Editora UFV, Viçosa
5. Guimarães CT, Schuster I, Magalhães JV, Souza CL Jr (2016) Marcadores moleculares no melhoramento de plantas. In: Borém A, Caixeta E (eds) *Moleculares M*, Editora UFV, Viçosa, pp 129–177
6. Holland PM, Abramson RD, Watson R, Gelfand DH (1991) Detection of specific polymerase chain reaction product by utilizing the 5'—3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences* 88:7276–7280. <https://doi.org/10.1073/pnas.88.16.7276>
7. Hospital F (2001) Size of donor chromosome segments around introgressed loci and reduction of linkage drag in marker-assisted backcross programs. *Genetics* 158:1363–1379
8. Isaaa (2019) *Facts about Biotech/GM Crops in 2018: Report Reveals Biotech Crops Continue to Provide Solutions to Hunger Malnutrition and Climate Change*. <https://www.isaaa.org/resources/publications/briefs/55/executivesummary/default.asp>, ISAA Brief No.55, Accessed 08 may 2021.
9. Mesquita AGG, Guimarães CT, Parentoni SN, Paiva E (2005) Recuperação do Genitor Recorrente em Milho Utilizando Retrocruzamento Assistido por Marcadores Microsatélites. *Revista Brasileira de Milho e Sorgo* 4:275–285. <https://doi.org/10.18512/1980-6477/rbms.v4n03p%25p>
10. Mumm RH (2013) A look at product development with genetically modified crops: examples from maize. *J Agric Food Chem* 61:8254–8259. <https://doi.org/10.1021/jf400685y>
11. Mumm RH, Walters DS (2001) Quality control in the development of transgenic crop seed products. *Crop Sci* 41:1381–1389. <https://doi.org/10.2135/cropsci2001.4151381x>
12. Peng T, Sun X, Mumn RH (2013) Optimized breeding strategies for multiple trait integration: I. Minimizing linkage in single event introgression. *Mol Breeding* 33:89–104. <https://doi.org/10.1007/s11032-013-9936-7>
13. Wang J, Chapma SC, Bonnett DG, Rebetzke GJ, Crouch J (2007) Application of population genetic theory and simulation models to efficiently pyramid multiple genes via marker-assisted selection. *Crop Sci* 47:582–590. <https://doi.org/10.2135/cropsci2006.05.0341>
14. Wang N, Yuan Y, Wang H, Yu D, Liu Y, Gowda AZM, Nair SK, Hao Z, Lu Y, Vicente FS, Prasanna BM, Li X, Zhang X (2020) Applications of genotyping-by-sequencing (GBS) in maize genetics and breeding. *Nature Research* 10:16–308. <https://doi.org/10.1038/s41598-020-73321-8>

## Tables

**Table 1.** Genetic similarity between donor and recurrent parents in nine backcrossing populations, based on 1,065 SNP markers.

Recurrent Parent (RP)	Donor Parent (DP)	Group <sup>1</sup>	Similarity between RP and DP	Average RP and DP Similarity per Group	Num Plants BC <sub>1</sub> <sup>2</sup>	Num Plants BC <sub>2</sub> <sup>2</sup>	Num Plants BC <sub>3</sub> <sup>2</sup>
RP1	DP1	1	76.20%	74.00%	88	30	86
RP2	DP1		74.40%		91	48	48
RP3	DP1		71.40%		89	48	81
RP4	DP1	2	66.50%	68.30%	84	41	80
RP5	DP1		69.30%		89	31	85
RP6	DP1		69.10%		80	46	99
RP7	DP1	3	54.80%	54.70%	86	26	79
RP8	DP1		59.20%		95	30	82
RP9	DP1		50.10%		95	86	78

<sup>1</sup>Group 1, 2 and 3 means levels of similarity between recurrent and donor parents: 1: Greatest similarity. 2: Intermediate. 3: Greatest Distance.

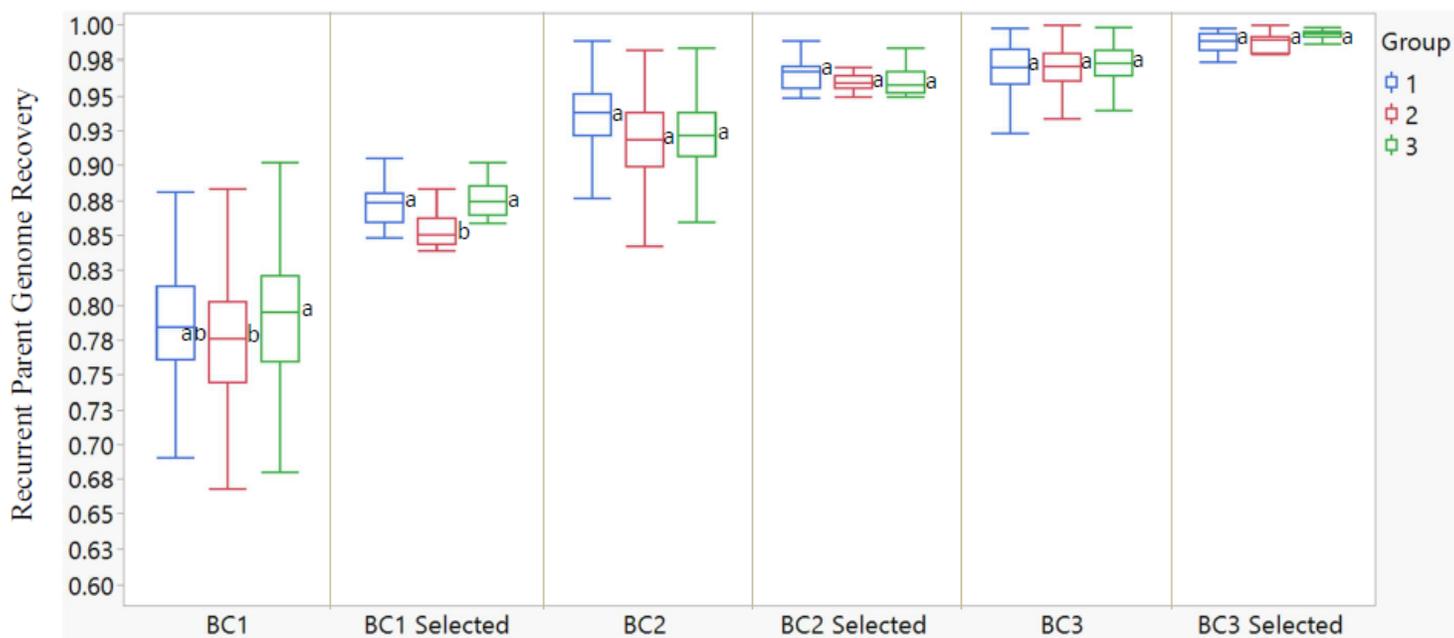
<sup>2</sup>Number of plants used in Marker-Assisted Backcrossing for similarity or genome recovering selection, after selected for target trait.

**Table 2.** Correlation of the genetic similarity between backcrossing populations parents and BC populations similarity average and with the genome recovery average in each BC generation.

Parent Similarity	BC Generation			** Significant at 1% probability by t test ++Significant at 1% probability by Mantel test based on 5,000 simulations
	BC <sub>1</sub>	BC <sub>2</sub>	BC <sub>3</sub>	
	All BC Plants			
BC and RP Similarity	0.9704 <sup>+++</sup>	0.9400 <sup>+++</sup>	0.9003 <sup>+++</sup>	<b>Table 3.</b> Observed genome recovery in marker-assisted backcrossing and equivalency with theoretical backcross genome recovery.
Genome Recovery	0.5035	0.2079	-0.3451	
	Selected BC Plants			
BC and RP Similarity	0.9170 <sup>+++</sup>	0.4545	0.4723	
Genome Recovery	0.3126	0.2437	0.5766	

Observed BC Generation	Genome Recovey	Theoretical BC generation	Expected Genome Recovery
BC <sub>1</sub>	78.29%	BC1	75.00%
BC <sub>1</sub> selected plants	86.73%	BC2	87.50%
BC <sub>2</sub>	92.57%	BC3	93.75%
BC <sub>2</sub> selected plants	96.18%	BC4	96.88%
BC <sub>3</sub>	96.97%	BC5	98.44%
BC <sub>3</sub> selected plants	98.92%	BC6	99.22%

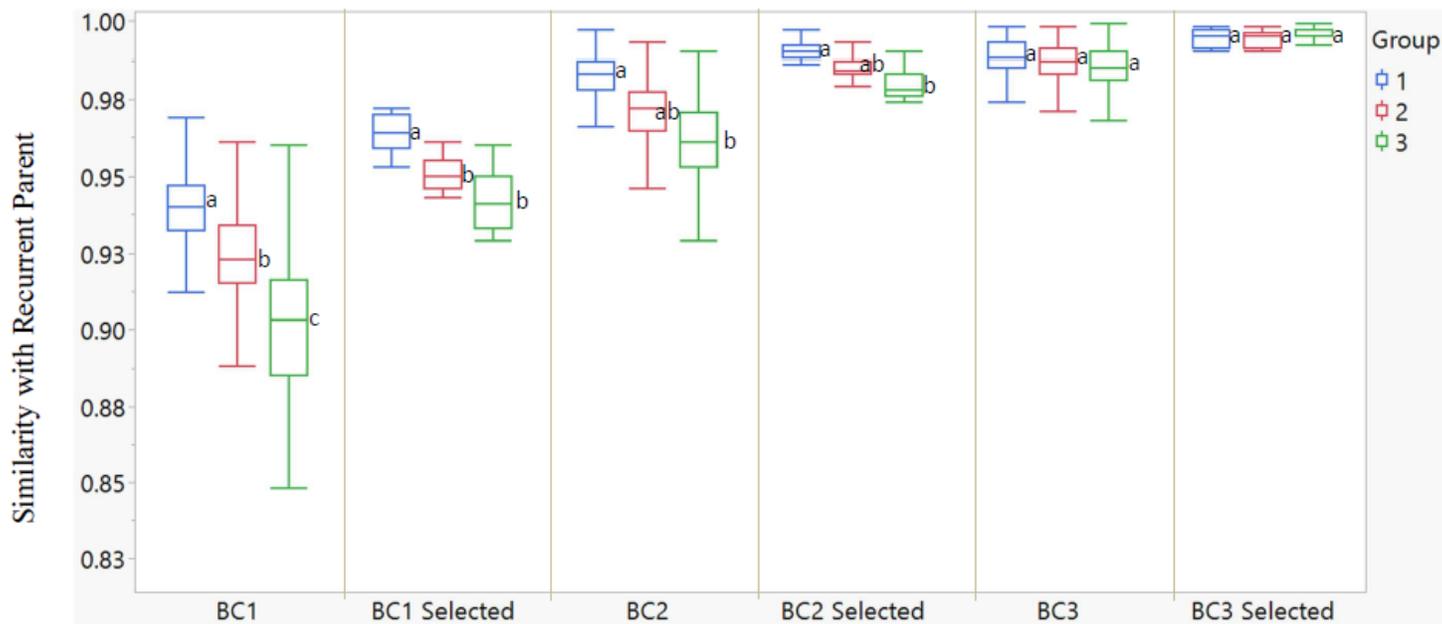
## Figures



**Figure 1**

Recurrent parent genome recovery in three Marker-Assisted Backcrossing generations. BC Selected contains the data of the five plants in each group and generation with higher recurrent genome recovery. Group 1: Average similarity between donor and recurrent parents = 74%, Group2: Average similarity between donor and recurrent parents = 68.3%, Group3: Average similarity between donor and recurrent

parents = 54.7%. Same letters in each generation means same genome recovery average by Tukey's test at 5% probability.



**Figure 2**

Similarity of corn backcrossing plants in three Marker-Assisted Backcrossing generations with recurrent parent. BC Selected contains the data of the five plants in each group and generation with higher similarity with recurrent parent. Group 1: Average similarity between donor and recurrent parents = 74%, Group 2: Average similarity between donor and recurrent parents = 68.3%, Group 3: Average similarity between donor and recurrent parents = 54.7%. Same letters in each generation means same similarity average with recurrent parent by Tukey's test at 5% probability.