



HAL
open science

Improving the use of plant genetic resources to sustain breeding programs' efficiency

Dimitri Sanchez, Sarah Ben Sadoun, Tristan Mary-Huard, Antoine Allier, Laurence Moreau, Alain Charcosset

► To cite this version:

Dimitri Sanchez, Sarah Ben Sadoun, Tristan Mary-Huard, Antoine Allier, Laurence Moreau, et al.. Improving the use of plant genetic resources to sustain breeding programs' efficiency. Proceedings of the National Academy of Sciences of the United States of America, 2023, 120 (14), pp.e2205780119. 10.1073/pnas.2205780119 . hal-04215595

HAL Id: hal-04215595

<https://hal.science/hal-04215595v1>

Submitted on 22 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Improving the use of plant genetic resources to sustain breeding programs' efficiency

Dimitri Sanchez^a , Sarah Ben Sadoun^a , Tristan Mary-Huard^{ab} , Antoine Allier^c , Laurence Moreau^a , and Alain Charcosset^{a,1}

Edited by Susan McCouch, Cornell University, Ithaca, NY; received June 8, 2022; accepted September 13, 2022

Genetic progress of crop plants is required to face human population growth and guarantee production stability in increasingly unstable environmental conditions. Breeding is accompanied by a loss in genetic diversity, which hinders sustainable genetic gain. Methodologies based on molecular marker information have been developed to manage diversity and proved effective in increasing long-term genetic gain. However, with realistic plant breeding population sizes, diversity depletion in closed programs appears ineluctable, calling for the introduction of relevant diversity donors. Although maintained with significant efforts, genetic resource collections remain underutilized, due to a large performance gap with elite germplasm. Bridging populations created by crossing genetic resources to elite lines prior to introduction into elite programs can manage this gap efficiently. To improve this strategy, we explored with simulations different genomic prediction and genetic diversity management options for a global program involving a bridging and an elite component. We analyzed the dynamics of quantitative trait loci fixation and followed the fate of allele donors after their introduction into the breeding program. Allocating 25% of total experimental resources to create a bridging component appears highly beneficial. We showed that potential diversity donors should be selected based on their phenotype rather than genomic predictions calibrated with the ongoing breeding program. We recommend incorporating improved donors into the elite program using a global calibration of the genomic prediction model and optimal cross selection maintaining a constant diversity. These approaches use efficiently genetic resources to sustain genetic gain and maintain neutral diversity, improving the flexibility to address future breeding objectives.

genetic diversity | bridging | breeding programs | optimal cross selection | genomic selection

Modern plant selection can be traced back to the 19th century with pioneering work by the de Vilmorin family (1). They initiated the switch from single plant selection to that of families, i.e., genealogical selection. This led to a significant gain in heritability, one key driver of genetic gain in the breeder's equation (2). It cleared the way for numerous methodological developments up to recent advances in genomic prediction. Numerous selection cycles in main annual crops have led to high genetic progress, which has contributed from 50 to 75% of productivity gains (3). Nevertheless, these advances have been accompanied by a loss of genetic diversity available for the farmers and the breeders (4–6). This may cause production instabilities in areas with adverse abiotic and biotic environmental conditions. It also may slow down genetic gain in breeding programs (2). In addition, several studies demonstrated that in the absence of specific constraints on diversity maintenance, recent selection methods such as genomic selection (GS) could accelerate the loss of genetic diversity by selecting highly related individuals (7–9). Diversity is also key to the evolution of breeding programs toward new selection objectives, such as those needed to anticipate climate change (10–12). Therefore, a suitable management of genetic diversity is needed. Recent advances in genotyping and statistical genetics methods offer new possibilities to achieve this objective.

Genetic diversity depletion within breeding populations can be slowed down using specific constraints. Methodologies to do so emerged initially in animal species for which the breeding population is directly used in production. In this case, excessive loss of diversity reduces production due to inbreeding depression, in addition to lowering additive genetic variance and therefore long-term genetic gain. A first approach is to impose restrictions on the number of progeny selected from any family (13). A more efficient one is to consider relatedness between the selected individuals and their contribution to the next generation (14, 15). These elements are incorporated in the optimized cross selection (OCS) index definition. OCS enables to select lists of parental crosses maximizing the mean expected breeding values of the progeny while constraining diversity to be above a predetermined decline trajectory. This was applied first using pedigree inferred relatedness and later extended to genotypic data. It has been shown that this approach can increase

Significance

The increase and stability of crop production call for further genetic progress. To maintain it, breeders need to preserve genetic diversity in their programs, although it is depleted by the selection. During the last century, large collections of genetic resources have been organized, representing significant diversity. However, using them for variety development has remained an unsolved problem due to their performance gap with the elite varieties. Advances in genotyping and statistical methods now allow their efficient use through genomic prediction-based bridging schemes and diversity monitoring. Our study provides a demonstration of a suite of new tools incorporated into new breeding schemes to effectively resurrect donor accessions, enabling them to contribute to the improvement of quantitative traits in elite populations.

Author contributions: D.S., S.B.S., T.M.-H., A.A., L.M., and A.C. designed research; D.S. and T.M.-H. performed research; D.S. and T.M.-H. contributed new reagents/analytic tools; D.S., S.B.S., A.A., L.M., and A.C. analyzed data; and D.S., S.B.S., T.M.-H., A.A., L.M., and A.C. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹To whom correspondence may be addressed. Email: alain.charcosset@inrae.fr.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205780119/-/DCSupplemental>.

Published March 27, 2023.

long-term genetic gain, with only minor short-term penalty (16, 17). Recently, Allier et al. (18) refined OCS in the context of GS schemes by developing the usefulness criterion parental contribution (UCPC). UCPC evaluates the interest of a list of crosses based on the expected genetic gain [UC (19)] and parental contributions (PC) of the selected individuals in the next generation. Simulations showed that the use of UCPC increased both short- and long-term genetic gains compared to OCS (20). In addition, other criteria were developed, such as the optimal haploid value (21) or the optimal population value (22), which are based on the selection of crosses or of a population, respectively, predicted to give the best doubled haploid progeny. Compared to GS without constraints on diversity, these aforementioned approaches preserve substantially more genetic diversity in the population by limiting the loss of low-frequency variants. However, with realistic plant breeding population sizes, diversity erosion in closed programs appears ineluctable in the long term.

Sustaining the efficiency of plant breeding programs therefore necessitates an appropriate diversity enrichment strategy. This strategy must i) identify relevant donors, i.e., individuals carrying favorable alleles not yet, or no longer, present in the breeding population and ii) incorporate these alleles into the elite population. Diversity sources in plant breeding can be categorized according to their readiness for use in an elite program. Two main factors impact this use: their “eliteness” and their environmental adaptation (*SI Appendix, Fig. S1*). Current competitor materials adapted to the target environment are the least challenging for incorporation. The accessibility of such resources depends on regulation rules. Varieties under protection by the International Union for the Protection of New Varieties of Plants (UPOV) are accessible for breeding, which guarantees a straightforward use for inbred varieties (*SI Appendix, Fig. S1*). Patents and Plant Variety Protection (PVP) delay the access to competitor materials until patent expiration, giving to the owner a substantial advantage. Incorporation of UPOV-protected and ex-PVP materials has become a routine procedure in breeding programs. In contrast, the inclusion of landraces and lines from collections of patrimonial materials (23) remains challenging.

The use of the diversity collections in breeding has long been limited by the genetic gap in performance with the elite programs. These collections were designated as “seed morgues” by Goodman in the 1980s (24). Nevertheless, successful introductions of favorable alleles originated from these resources were recorded for monogenic and oligogenic traits (25), and efficient marker-assisted backcross or gene pyramiding approaches have been proposed to facilitate them (26–29). Diversity enrichment strategies adapted to quantitative traits are more complex due to the polygenic nature of the traits but have also received long-term attention. Pioneering methodological work aiming at detecting useful individuals to complement elite materials was conducted by Dudley (30, 31). Cramer and Kannenberg (32) proposed a reshaped breeding program to transfer diversity from donors with an overall poor agronomic level to an elite breeding program through intermediate populations. These questions are currently renewed in the GS era (33, 34). Marker effects estimated using a genomic model calibrated with a suitable training population can help identify donor individuals that best complement an elite recipient population. Allier et al. (35) evaluated a criterion (H) using a sliding window along the genome to identify chromosomal regions where a donor outperforms the best individuals in the elite population. These regions can then be considered to target specific recombination events in the progeny of a donor × recipient cross (36). Regarding the incorporation step, Allier et al. (37) highlighted the benefit of introducing a bridging component in the selection scheme, i.e.,

selecting within the progeny of donor × elite crosses before introduction in the elite component. This appears as mandatory for the successful use of low-performance genetic resources, as they otherwise get introduced in the elite pool to sustain diversity, but with no chance of entering the pedigree of released elite varieties. This was recently confirmed by Vanavermaete et al. (38) and Breider et al. (39), who extended this idea by introducing a multicycle bridging component between the genetic resource component and the elite program. These studies showed a large long-term benefit of introducing genetic diversity, even from low-performing genetic resource collections, over closed elite schemes with an optimized diversity trajectory management.

The objective of our study is to further improve the use of genetic resource collection materials (classes + and ++ in *SI Appendix, Fig. S1*) in variety development, in a context where their overall gap in performance with elite materials increases over time. To do so, we consider as a general framework the scheme described in Fig. 1. We used the recent generic public simulator MoBPS [Modular Breeding Program Simulator (40)] to evaluate how the global efficiency of the program is affected by diversity management, genetic resources donor choice, genomic models used to select parental individuals and crosses of the bridging and elite components, with a hint in the allocation of resources to the different steps. We also analyzed the fate of donor alleles introduced in the program, the dynamics of diversity at neutral markers, and loci under selection. Our results emphasize the interest of phenotyping genetic resource collections and that of the bridging step to select donors which have the potential to unleash useful variation. They show that allocating 25% of total experimental resources to upstream bridging efforts increases mid- and long-term genetic gain and makes it possible to maintain diversity at currently neutral genomic regions, thus guaranteeing more flexibility to breed for new traits in a changing climatic and societal context.

Results

Advantages of Managing Diversity in a Closed Breeding Program.

We first evaluated as a benchmark a closed breeding program (i.e., elite population in Fig. 1) using different options to define the cross list. The first three are based on parental performances (phenotypic or predicted by a GS model) or the cross usefulness criterion (UC) which considers also the expected variance in

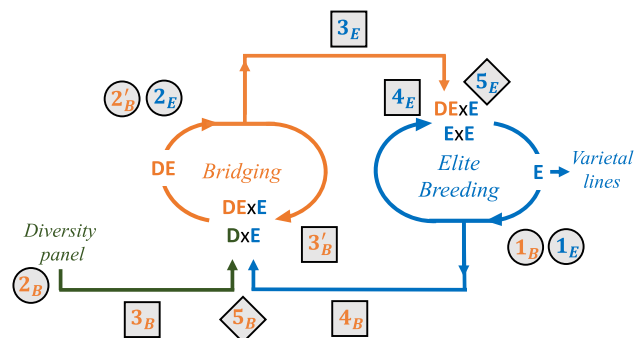


Fig. 1. Global scheme for evaluated strategies. From left to right, donor detection from a diversity panel (green), donor improvement in a bridging breeding component (orange), and elite breeding phase to select new varieties (blue). Numbers indicate decisions to be taken at each selection cycle in the bridging (B) and elite (E) components: 1) Elite parent candidate evaluation, 2 and 2') nonelite parent candidate evaluation, 3 and 3') preselection of nonelite parent candidates, 4) preselection of elite parent candidates, and 5) cross list selection. The types of decisions are recalled by the shade: parent evaluation (circle), parent preselection (square), and cross selection (diamond).

progeny. The last two (OCS and UCPC-based OCS) use optimal cross selection methods to control the loss of genetic diversity at neutral markers (He) by imposing a decline trajectory. We selected 20 crosses each year and generated 80 progenies per cross.

As expected, OCS and UCPC-based OCS with a controlled decline in diversity led to the highest long-term genetic gain, with only a short-term penalty. (*SI Appendix, Fig. S2 and Table S1*). After 60 generations, all strategies led to a low level of genetic diversity. Note that attempting to maintain the elite diversity constant was not advantageous in the short term and it was too stringent so that long-term UCPC parental selection failed. In the following, we introduced external donors to maintain diversity. UCPC-based OCS with a controlled decline in diversity in a closed program was considered as a benchmark to evaluate the benefit of different open breeding strategies

Usefulness of Donor Introductions using Bridging Strategies. Impact of GS Model Calibration and Diversity Management.

We centered our study on the introduction of diversity from gene bank collections, using different strategies listed in *SI Appendix, Table S2*. We first kept the same total breeding resources as for the closed program and attributed 25% (5 crosses per year) to a one-stage bridging component and 75% (15 crosses per year) to the elite component. This allocation of experimental resources is referred as $15C_E + 5C_B$ (C_E : elite crosses, C_B : bridging crosses). All examined strategies surpassed the closed benchmark (UCPC) after 35 generations on the genetic gain (μ) and slightly earlier on the genetic gain at the commercial level (μ_{10} ; mean of the 10 most performing elite DH lines at a given generation, which mimics the mean value of released varieties from the elite program, Fig. 2). We did not notice any penalty on μ_{10} in previous generations. A joint GS model calibration on both bridging and elite components (Joint TS) appeared superior to using specific calibration (Specific TS) for each component ($\mu = 40.92\%$ compared to $\mu = 38.42\%$ after 60 generations, *SI Appendix, Table S3*). We used a Joint TS for the rest of our study. Keeping a constant elite diversity (He constant) led to lower μ values than allowing

a linear diversity decline (He linear loss) until the generations 30 to 35, whereas we observed the reverse later. Focusing on μ_{10} , the “He constant” performed similarly as “He linear loss” until generations 30 to 35 and outperformed it later. Regarding diversity, we observed a burst in the first generations of introductions (Fig. 2B). Diversity then followed the targeted trajectory. In the case of a linear decline trajectory, it exceeded the targeted values in the long term.

Benefits of a Multiple-Stage Bridging Component on Genetic Gain. To allow a progressive improvement of donor individuals before their incorporation into the elite component, we considered a multiple-stage bridging strategy, in which we allow crosses between bridging progeny and elite individuals (DExE) in addition to donor \times elite crosses (Fig. 1). This was applied using all available bridging progenies or after intracross preselection to balance cross contribution. These alternatives were evaluated under both above-described diversity management regimes (Fig. 3 and *SI Appendix, Table S4*). We confirmed the superiority of the “He constant” for all options. Multiple-stage strategy with preselection of DE progeny appeared slightly superior than that without ($\mu = 46.80\%$ and $\mu = 45.72\%$, respectively) and surpassed the one-stage strategy ($\mu = 44.79\%$) after 60 generations. Although beneficial, especially in the long term, multiple-stage strategy had a milder effect than diversity management.

Advantage of Phenotype-Based Selection of Donors. Donors were selected above based on UCPC of all possible DxE crosses (5B in Fig. 1), using the GS model calibrated with the Joint TS. We evaluated an alternative based on OCS and phenotypic evaluation of donors. This approach was implemented i) directly on the whole population of donors or ii) after preselection of 20 donors using criterion H to target original favorable chromosome segments superior to those of the elite component. Phenotype-based selection appeared beneficial for both μ and μ_{10} (*SI Appendix, Fig. S3 and Table S5*), but this advantage decreased for the most advanced generations. With the one-

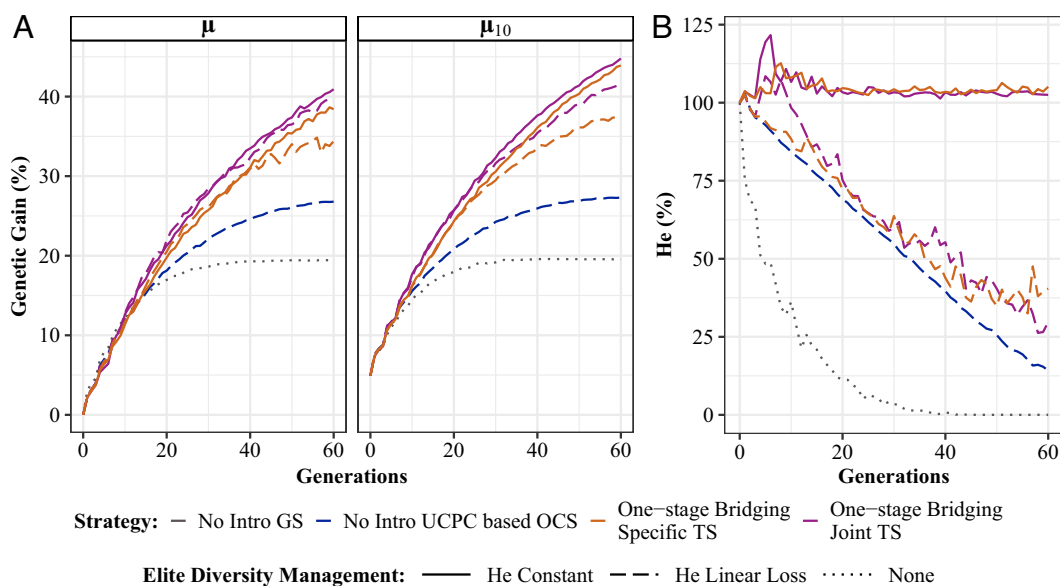


Fig. 2. Evaluation of the impact of one-stage bridging strategies under different options for the GS model calibration and the elite diversity management. (A) Mean genetic gain for elite population (μ) and for the 10 best individuals (μ_{10}) after the burn-in phase. (B) Elite neutral Nei diversity (He) in percentage of the neutral Nei diversity at the end of the burn-in phase. “No intro”: closed breeding program. “Specific TS” and “Joint TS”: choice of GS calibration model. “He Constant” and “He Linear Loss”: targeted elite diversity trajectory.

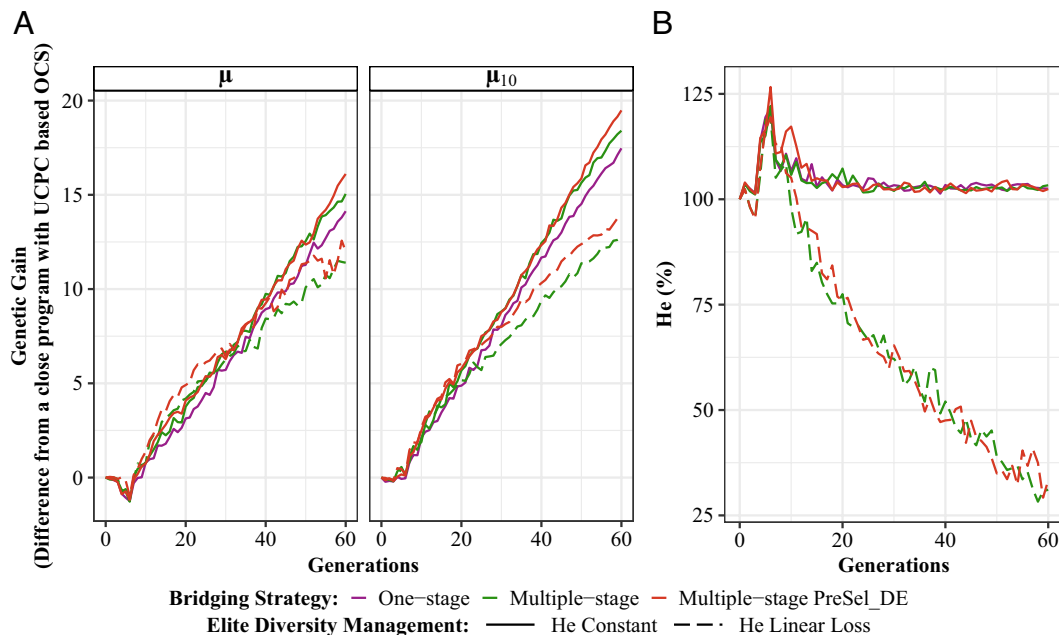


Fig. 3. Comparison of one-stage and multiple-stage bridging strategies under different options for the elite diversity management. (A) Mean genetic gain for elite population (μ) and for the 10 best individuals (μ_{10}) from the end of the burn-in phase reported as deviations from a closed program managed with UCPC-based OCS. (B) Elite neutral Nei diversity (He) in percentage of the neutral Nei diversity at the end of the burn-in phase. “One-stage”, “multiple-stage”, and “multiple-stage PreSel_DE”: bridging strategies, “PreSel_DE” indicates a preselection of DE progenies based on their performance. “He Constant” and “He Linear Loss”: targeted elite diversity trajectory.

stage bridging strategy, at generation 60, $\mu_{10} = 43.79\%$ and 45.67% when the donors are selected by GS or phenotyped, respectively. Preselecting donors based on the H criterion did not lead to a clear improvement. Note that at the beginning of the program, the phenotyping of donors reduced the advantage of the multistage bridging strategy.

Effect of Breeding Resource Allocation. We evaluated two additional experimental resource allocations between the bridging and elite component. We kept the same total number of individuals created each year as the initial allocation (1,600). For the first ($15C_E + 10C_B$), we increased from 5 to 10 the cross number made at each generation in the bridging while reducing from 80 to 40 the progeny number per cross. For the second ($10C_E + 10C_B$), we increased to 50% (10 crosses) rather than 25% (5 crosses) the experimental resources allocated to the bridging component, keeping 80 progenies by cross. $15C_E + 10C_B$ allocation had a weak positive impact on μ and μ_{10} compared to the initial allocation, whereas $10C_E + 10C_B$ allocation appeared prejudicial (SI Appendix, Fig. S4 and Table S5). The best for long-term genetic gain optimization was using the $15C_E + 10C_B$ allocation combined with a preselection of donors based on the H criterion followed by phenotype-based selection.

Dynamics of Diversity Evolution and Donor Contributions to the Elite Pool. A closed elite program (Fig. 4A and SI Appendix, Table S6) led to the fixation or quasifixation of all quantitative trait loci (QTLs) at generation 60, including a substantial proportion of QTLs fixed for the unfavorable allele (QTL^-). Strategies with diversity management delayed the QTL fixation and led to a higher proportion of QTLs fixed for the favorable allele (QTL^+) than the classical GS strategy. For strategies involving donor introductions (Fig. 4B and SI Appendix, Table S7), the “He linear loss” diversity management allowed a rapid fixation of

favorable alleles but the QTL^- proportion remained constant. Instead, the “He constant” allowed to efficiently decrease the QTL^- proportion over time. It also preserved a larger proportion of segregating QTLs. At generation 60 with a Joint TS and a one-stage bridging, 30.62% of QTLs was still segregating with this diversity management, whereas this number dropped to 9.02% with “He linear loss”.

In scenarios with introductions, the average proportion of donor genome in the elite population increased over time to reach between 37.4% and 56.7% at generation 60 (SI Appendix, Table S8). Using a Joint TS instead of a Specific TS reduced the number of different donors incorporated in the elite population (N_{ID}) (28.4 instead of 45.6), but a higher proportion of them had progenies up to the generation 60. Selecting donors based on their phenotype had an even stronger effect: N_{ID} dropped but more than three-quarters of the introduced donors had contributed up to generation 60 (SI Appendix, Table S8). Phenotype-based selection of donors, and to a lesser extent the use of a Joint TS, selected first donors in the best true breeding value (TBV) quintile and then progressively donors with lower TBVs (Fig. 5).

Discussion

Limits of Closed Breeding Programs. Maintaining genetic diversity in a breeding program is essential to guarantee sustainable genetic gain for targeted traits. Reduction in genotypic costs now allows to do so by implementing efficient molecular marker-based methods, such as those using optimal cross selection. We confirmed that they lead to a higher long-term genetic gain than selecting parents based on their phenotype or genomic prediction and intermating them randomly (37). The advantage of these methods comes from their ability to limit the involuntary fixation of negative alleles at QTLs due to drift or linkage with favorable alleles (Fig. 4). We developed optimization algorithms to facilitate

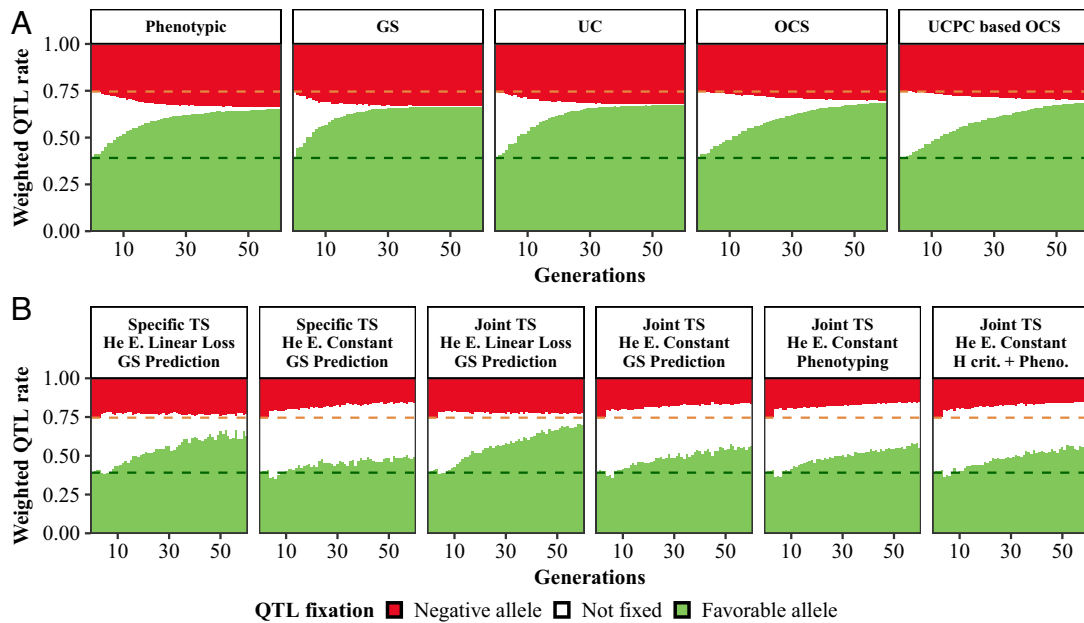


Fig. 4. Evolution of fixation and polymorphism maintenance at QTLs in closed breeding programs (A) and in elite component of one-stage bridging programs (B). At each generation, QTLs are classified into 3 groups according to their status in elite population: fixed for the negative allele (red), fixed for the favorable allele (green), and not fixed (white). Weighted QTL rate of each category is computed as the sum of absolute effects of corresponding QTL, divided by the sum of absolute effects of all QTLs. Dashed lines indicate the values at the end of the burn-in phase. “Specific TS” and “Joint TS”: choice of GS calibration model. “He Constant” and “He Linear Loss”: targeted elite diversity trajectory. “GS prediction”, “Phenotyping”, and “H crit. + Pheno”: donors selection strategies. For bridging programs, 25% of breeding resources are allocated to the bridging component.

their implementation. Nevertheless, our results and other long-term simulation studies (37, 41, 42) showed that, with realistic plant breeding settings, a genetic gain plateau is reached in all cases. Thus, introducing relevant sources of diversity is necessary to avoid it.

Beneficial Introduction of Genetic Resources through Improved Bridging Schemes. Diversity sources can be categorized according to their readiness for incorporation into an elite program (SI Appendix, Fig. S1). In this study, we focused on the use of adapted materials preserved in gene banks, created more than

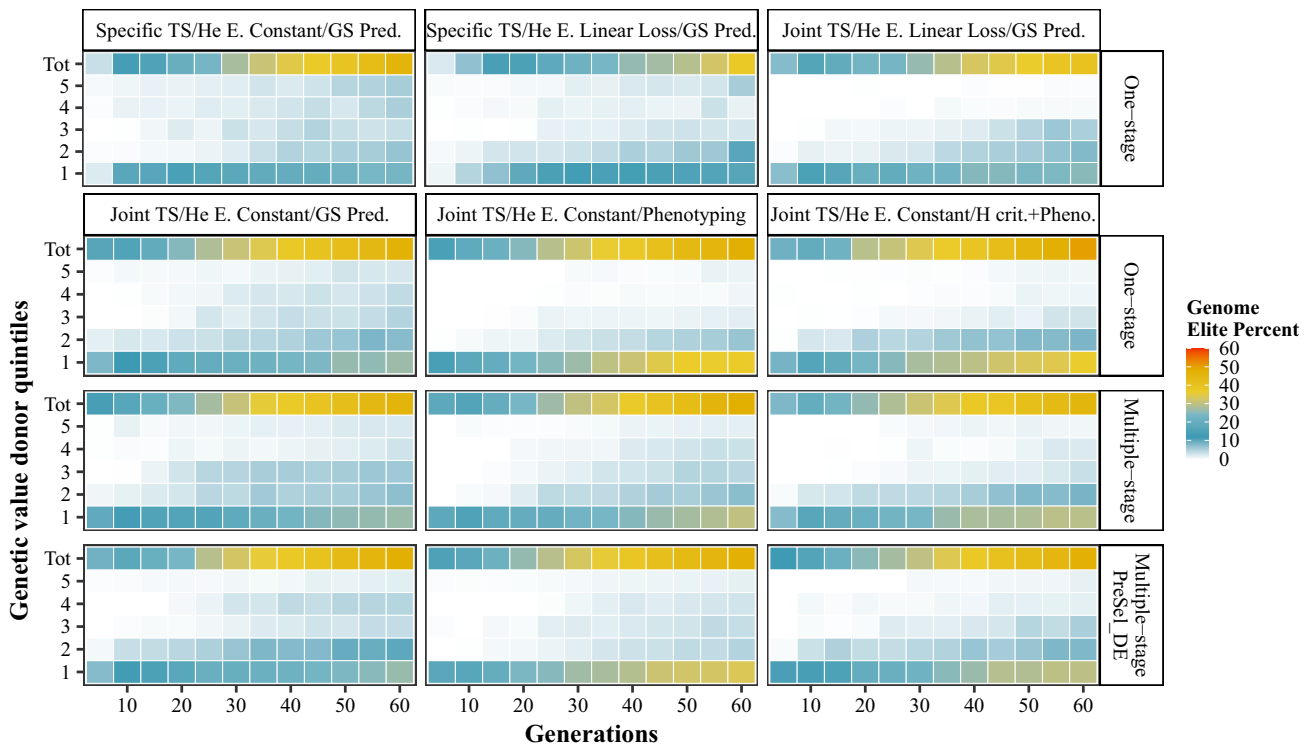


Fig. 5. Relative contributions of introduced donors to the elite pool. Donors are ranked according to their genetic value quintiles, from 1 to 5 for the highest and lowest performing, respectively. Total contribution is also displayed (Tot). Contribution to elite genome is calculated based on pedigree information. “Specific TS” and “Joint TS”: choice of GS calibration model. “He Constant” and “He Linear Loss”: targeted elite diversity trajectory. “GS prediction”, “phenotyping”, and “H crit. + Pheno”: donors selection strategies. “One-stage”, “multiple-stage”, and “Loop NoPreSel_DE”: bridging strategies. For bridging programs, 25% of breeding resources are allocated to the bridging component.

20 y before the elite program was started. *SI Appendix, Fig. S5* illustrates that this corresponds to a large genetic gap between donors and elite material, which increases over time as the elite program progresses. We show a clear benefit of enriching diversity using such genetic resources through a bridging component, i.e., by preselecting within the progeny of donor \times elite crosses prior to their incorporation into the elite program. This confirms previous findings (37). This advantage can be obtained with no additional cost for programs that already have implemented GS by reallocating a moderate fraction of the elite program resources to the bridging component. This strategy comes with no or very limited penalty in the short term and a beneficial impact on genetic gain at commercial level after 10 generations with our simulation settings. This delay is expected to be reduced with a lower initial diversity level in the elite component and a higher donor richness in favorable alleles.

Using only 25% of experimental resources for implementing the bridging appeared more efficient than using 50%. For given resources allocated to the bridging program, we recommend preserving a substantial number of progeny per donor \times elite cross rather than increasing the number of crosses. This is required to estimate accurately enough the variation generated by each cross to select valuable donors and also train appropriately the genomic prediction model for the newly introduced alleles.

Identification and Preservation of Useful Variations. Our results further lead to two important recommendations to ensure successful introduction.

Preselection of genetic resources should be performed based on phenotypic data rather than based on the GS model trained on the breeding program. This was true especially at the beginning of the program, when only a limited fraction of the genetic resource collection was integrated into the bridging program. It links to the well-known property of GS to favor individuals related to the training set (7–9) due to the impossibility for this approach to detect the interest of alleles not segregating within the training set (43). Phenotype-based selection of diversity donors reduced the number of donors entering the bridging component, but a higher proportion of these contributed to varietal release (*SI Appendix, Table S8*). This importance of phenotyping donors supports initiatives devoted to the phenotypic characterization of genetic resource collections, complemented by genomic prediction model training to facilitate prescreening (34, 44). This characterization should focus on performance (yield potential), tolerance to abiotic stresses and diseases, phenology, and also defaults like lodging to limit as much as possible later corrective efforts. This requires international coordination to have access to large trialing networks. Such initiatives, including public–private partnership are being deployed for several crops (45, 46) and must be generalized. A specific focus should be made on genotype \times environment interaction to address production stability and adaptation to climate change, with a characterization of potential donors in diverse environmental conditions. Combination of field network and phenotyping platform with GS and ecophysiological modeling (47) should prove instrumental in detecting genetic resources bringing new adaptive features.

Note that phenotype-based selection of donors hinders the possibility to predict the variance of donor \times elite crosses. In this case, crosses can be selected based on their average performance and their contribution to the diversity inferred from genotypic data, using OCS algorithms. This approach proved efficient but may be improved by calibrating a specific GS model for donor selection based on the genetic resource panel itself. This option may be further improved by adding representative individuals

from the elite germplasm and also individuals not selected (to represent the whole range of elite alleles) (48). Optimizing such strategies should be conducted using specific simulations which were beyond the scope of the present study.

GS models calibrated with a same global training set (Joint TS) should be used for evaluating all types of candidate crosses. With respect to genetic gain, this option appeared superior to selecting the crosses within the bridging and elite components with specific models (Specific TS). It also led to a higher incorporation rate of DE progeny in the elite germplasm (*SI Appendix, Table S8*). This suggests that one major interest of the bridging component of the program is to create and phenotype intermediate individuals between donors and elites that make it possible to evaluate contrasts between alleles of the elite pool and newly introduced alleles.

Despite an efficient phenotype-based selection of donors, we observed that only part of the donors introduced in the bridging component had a progeny selected for incorporation into the elite component (*SI Appendix, Table S8*). This might be explained by differences in the variability of the DE progeny depending on the donor. Such differences were observed in nested association mapping population in maize and were poorly predicted by the genetic distance between parents for traits with a complex genetic architecture (49, 50). GS-based prediction of variances would be highly beneficial but, as stated above, GS cannot anticipate the potential of donors with a low relatedness to the elite pool. One important aspect of the bridging crosses is thus to evaluate experimentally such variation, which necessitates sufficient progeny number for each donor \times elite cross (40 or 80 in our settings).

Advantage of Rapidly Recycling Improved Donor Progeny within a High Diversity Elite Component.

For donors with low performance, several cycles of crossing to the elite are needed to generate progeny competitive with elite lines. Pedigree examination showed that, with UCPC-based OCS, the elite diversity constraint allowed to preserve a part of introduced DE progeny and improve it by several crosses in the elite component (*SI Appendix, Fig. S6*). During this phase, crosses between DE progenies can also reveal complementarity between different donors. Vanavermaete et al. (38) showed an interest to have additional rounds of crosses to elite individuals for DE progenies within the bridging component. To understand the dynamics of the DE progeny improvement, we evaluated scenarios which combine the recycling of DE progenies in the bridging component and an elite diversity management. These scenarios did not lead to a substantial improvement of genetic gain in comparison with a one-stage bridging program despite a strong performance differentiation between elite and genetic resources (*SI Appendix, Fig. S5*). We observed that DE progenies were crossed to new E individuals essentially in the elite component, and their recycling in the bridging component was rarely used (*SI Appendix, Fig. S6 and Table S9*). An alternative diversity management with a weaker constraint on the elite diversity promoted the ExE crosses in the elite component at the expense of DExE crosses. It led to more frequent cycling of DE progeny in the bridging but without improving the final genetic gain (*SI Appendix, Table S9*). This confirms the impact of the diversity management on DE progeny improvement dynamics. Further investigations could be useful to find an optimal diversity management to ensure DE progeny improvement and maximize genetic gain in the elite component.

Benefit of Maintaining a Global Diversity at Different Scales.

Maintaining diversity is crucial to rapidly adjust programs to changing breeding targets in response to climatic and societal changes. For open breeding programs, we showed that application

of the UCPC-based OCS method can maintain diversity at neutral loci while increasing the frequency of the favorable alleles at QTLs (Fig. 4). Fixation of favorable alleles at QTLs is expected to lead to a loss of diversity at closely linked neutral markers, generating a tension between genetic gain and diversity maintenance. Our simulation settings involved 1,000 QTLs, so that most regions could not be considered “neutral”. Constant global diversity is therefore ensured by i) maximizing the diversity in regions with small cumulative effects for the trait of interest or ii) breaking LD in regions with larger effects so that a given favorable allele is carried by different haplotypes. These phenomena favor the occurrence of recombination events which can generate new haplotypic combinations at QTLs.

Our results underline the importance of the introduction of genetic diversity in individual breeding programs to ensure the release of original varieties and diversify the portfolio. Efforts to measure and monitor diversity of varieties deployed in a given agricultural landscape should be considered as a further step, using indices reviewed in Khoury et al. (4). Public policies that would encourage diversity maintenance at the landscape level are expected to have a positive impact on food security thanks to a better resilience to unpredictable climatic conditions. For example, considering genetic distance to existing varieties as a bonus in registration process may conduct breeders to consider explicitly the originality of released varieties as a target in their programs, providing a further incentive to use genetic resources collections.

Conclusion and Future Work. Our results support that the introduction of diversity from gene banks, which has long been questioned by breeders, can now be reconsidered using new tools and breeding strategies. They confirm the interest of the bridging strategies to introduce resources from collection materials. Our approach should be particularly useful to counterbalance the effect of genetic drift for breeding programs conducted with a small effective population size. This is the case of programs targeting market niches, which have limited resources and/or limited access to external elite resources, calling for incorporations from diversity collections. It also should be beneficial for larger breeding programs with restricted access to external elite germplasm or when programs of all competitors converge toward a same genetic pool with limited diversity (see *Discussion* above on diversity at the landscape level). More generally, we recommend performing a systematic assessment of the joint evolution of genetic gain, genetic variance, and genetic diversity, to detect situations, in which diversity erosion has a negative impact on genetic gain and calls for a reenrichment. (35) This approach should also help to estimate empirically the level of diversity to be targeted in the elite program.

Our results refine the roles of a bridging population. Beyond the reduction of the performance gap between diversity donors and the elite materials, which was its initial motivation, it contrasts the donor and elite alleles to calibrate efficiently GS models used to valorize the genetic variability. It also selects donors generating useful variation when crossed to the elite population. In our simulations, all candidates were phenotyped and GS models were recalibrated at each generation. A work to determine an optimal frequency of GS model recalibration would be interesting to possibly reduce phenotyping cost and accelerate breeding cycles.

Future work may also consider different trait simulation settings. Our single trait simulation setting can reflect a multitrait selection based on a linear combination but a more explicit multitrait simulation may be valuable (51). Also, we neglected the slowdown of the genetic variance erosion due to mutations and epistatic interactions (52–54). We focus here on the additivity

that reflects well autogamous breeding or allogamous breeding conducted with a tester line. However alternative GS-based schemes are currently being developed in allogamous breeding (55, 56), and dominance should be considered to manage the complementarity between breeding populations, leading to specific constraints in the monitoring of genetic diversity. Our study could also be extended to test the adequacy of bridging strategies to more exotic material introductions with maladaptation calling for a preselection for specific traits. For landraces, the extraction of improved materials such as double haploid lines allows facilitating their evaluation and preselection (57).

Finally, new breeding technologies are increasingly used in breeding programs for targeted modifications of DNA guided by knowledge of key genes and favorable allele sequences. It has been demonstrated that editing few key genes permits de novo domestication of wild accessions (58, 59). While preserving their genomic originality, this brings such resources to a level suitable for introduction in a bridging program (level “+” in *SI Appendix, Fig. S1*). Gene editing can also be considered to create de novo allelic series for genes involved in agronomical interest traits (60). Introducing such materials among donors of the bridging component (Fig. 1) is an avenue to evaluate de novo and natural alleles for their exploitation in the elite compartment. We therefore hypothesize that rather than competing with each other, approaches presented in this study should be highly synergistic with new breeding technologies.

Materials and Methods

Genetic Material and Simulated Trait Architecture. Simulations were conducted based on genotypic data of maize lines of the Amazing dent diversity panel (43), which covers founder germplasm of dent commercial breeding. Lines were genotyped with single-nucleotide polymorphisms (SNPs) from the Illumina MaizeSNP50 BeadChip (61). QC filters (line call rate ≥ 0.8 , marker call rate ≥ 0.9 , line heterozygosity rate ≤ 0.1 , and marker heterozygosity rate ≤ 0.15) lead to 338 lines and 41,495 high-quality SNPs. This panel was structured into three main groups: 57 Iodents, 82 Iowa Stiff Stalk, and 199 other dents.

Simulations were performed using R scripts (62) and the R package MoBPS (40). We simulated one quantitative trait influenced by 1000 biallelic QTLs sampled among the SNPs, with a minimal distance between two consecutive QTLs of 0.2 cM. Additive effects at QTLs were sampled from a centered Gaussian distribution with a variance of 0.05. We also sampled 2,000 SNPs as neutral markers. Genetic positions were determined using the dent consensus map (63).

Initial Elite Program. We simulated an elite breeding program (E) using the 57 Iodent lines as founders. Progeny was obtained using doubled-haploid (DH) technology. Recombination events along chromosomes were determined using a Poisson distribution, as suggested by the MoBPS user guide. We assumed 3 y to produce genotype and phenotype DH progeny. We considered overlapping generations with the selection of generation T parents from generations T-3, T-4, and T-5 to represent situations encountered in actual breeding programs. To initiate E progeny, 20 random founders were arbitrarily crossed with replacement during the first three generations, to create every year 10 families of 80 DH each. During the next 17 generations, the 50 E progeny with the best phenotypic values from the generations T-3, T-4, and T-5 were selected and randomly intermated to create 20 biparental families of 80 DH each. Phenotypic values were inferred from the genotype at QTLs and an error variance corresponding to a trait heritability of 0.7 in the founder population, mimicking an usual multitrait phenotypic evaluation with a single replicate in 4 environments and a repeatability of 0.37. Starting from the initial elite program, different breeding strategies were conducted for 60 generations, including or not including diversity donor (D individuals) introductions using a bridging component. We replicated the procedure 10 times.

Benchmark Implementation: Closed Elite Breeding Programs. Closed breeding programs were implemented after the burn-in phase, with parental cross list selection guided by either i) the mean parental phenotypic values, ii)

the mean parental GS predicted values (GEBV), iii) the cross UC, (19), iv) a classical OCS (see below, ref. 17), or v) a UCPC-based OCS. (see below, ref. 18). For the first option, 5% E progenies (i.e., 4 DH) with the highest phenotypic performance within each family of generations T-3, T-4, and T-5 were preselected as parents (i.e., 240 DH). The 20 biparental crosses with the best mean parental phenotypic values were selected to produce the next generation.

For other options, E progenies were genotyped for the 2,000 neutral markers and selection was based on a GBLUP model trained on generations T-3, T-4, and T-5 (i.e., 4,800 E progeny). To reduce computation time, genome-based estimated breeding values (GEBVs) were computed using the "direct approach" of MoPBS, which considers a fixed error variance. Marker effects were estimated by backsolving (64). We performed the same within family preselection as described above, using GEBVs instead of phenotypic values to determine potential parents. For the second and third options, we selected the 20 crosses with the best parental mean GEBVs and UC values, respectively. For a cross between lines j and k , UC was calculated as

$$UC_{jk} = \frac{GEBV_j + GEBV_k}{2} + ih\sigma_{jk} \quad [1]$$

where i is the selection intensity (here $i = 2.06$), h is the evaluation accuracy, and σ_{jk} is the expected genetic standard deviation in progeny. As in (18), we considered $h = 1$ for sake of simplicity. σ_{jk} was estimated using estimated marker effects and linkage disequilibrium between markers inferred from the genetic map, as proposed by (65). In the last two options, further management of the genetic diversity was done using OCS methods described in *SI Appendix, SI Text*. For all options, 80 DH were derived for each cross (i.e., 1,600 DH per generation).

One-Stage Bridging Implementation. We simulated different breeding programs with donor introduction through a bridging component (Fig. 1), keeping resources constant (i.e., 1,600 DH per generation). Donor individuals came from the whole diversity panel simulating a static genetic resources collection. They were first crossed to elite lines (DxE crosses) to obtain bridging progeny (DE individuals). Part of DE individuals were then selected to be introduced into the elite component (DxE crosses). Three resource allocation scenarios were considered: i) a bridging population of 5 families of 80 DH (i.e., 400 DE progeny) and an elite population of 15 families of 80 DH (i.e., 1,200 E progeny), ii) a bridging population of 15 families of 40 DH (i.e., 400 DE progeny) and an elite population of 15 families of 80 DH (i.e., 1,200 E progeny), and iii) a bridging population of 10 families of 80 DH (i.e., 800 DE progeny) and an elite population of 10 families of 80 DH (i.e., 800 E progeny). We restrained E potential parents to the 5% E progeny per family (4 DH) with the highest GEBVs in generations T-3, T-4, and T-5 (i.e., 180 DH for i and ii vs. 120 DH for iii). DE potential parents for introduction in the elite program were selected from the DE progeny (i.e., 60 DH for i and ii vs. 120 DH for iii). For these preselection steps, GEBVs were estimated with two distinct GS models calibrated, respectively, on E progeny (i.e., 2,400 or 3,600 DH) and DE progeny (i.e., 2,400 or 1,200 DH). Elite crosses were selected using a UCPC-based OCS (*SI Appendix, SI Text*) with an elite GS model ("Specific TS" in *SI Appendix, Table S2*) or a GS model calibrated on both components (summing up to 4,800 DH—"Joint TS" in *SI Appendix, Table S2*). DxE bridging crosses were selected among all possible crosses between D (i.e., 368 lines) and preselected E progeny using a bridging specific GS model or the joint GS model. This selection was performed conditionally to the elite crosses from the same generation to complement the elite component by the bridging component (*SI Appendix, SI Text*). Note that calibrating bridging-specific or joint GS models required three generations after the burn-in phase, so that they could first be used at generation 24. Before this, the bridging cross selections were performed using the elite specific GS model.

Multiple-Stage Bridging Implementation. To implement the potential recycling of donors within the bridging component, we adapted previous settings to allow DxE crosses in addition to DxE crosses in bridging. Two options were considered to determine DE candidate parents: i) all DE progeny from generations T-3, T-4, or T-5 (i.e., 2,400 or 1,200 DH) or ii) the preselected fraction considered for introduction in the elite component (i.e., 60 or 120 DH). We did not apply a direct constraint on the proportion of selected DxE versus DxE crosses. The same DE progeny could be selected both for introduction in the elite component and for recycling in bridging.

Preselection and Phenotyping of Donor Candidates. For some studied bridging strategies, candidate DxE crosses were selected using donor phenotypes. In this case, diversity panel lines were phenotyped only once considering the same environmental variance as for bridging and elite progeny evaluation. Phenotypes were centered and shrunk using the observed heritability to be comparable to values predicted with a GS model. As phenotype-based evaluation of donors hinders the possibility of predicting the variance of DxE progeny, crosses were selected using a classical OCS procedure (*SI Appendix, SI Text*). We also investigated strategies with a preliminary selection of donors based on an H criterion aiming at complementing E progeny from the T-5, T-4, and T-3 generations with favorable haplotypes ("H criterion + Phenotyping" in *SI Appendix, Table S2*). H is the sum of overlapping haplotype segment values and reflects the maximum DH line performance expected after several generations of intercross and selection of the donor and elite lines (see ref. 35 and *SI Appendix, SI Text* for further details). Twenty candidate donors were selected for each generation using this criterion.

Metrics used to Assess Scenario Performance. At each generation, we computed the mean genetic values of all E progeny (μ) and the 10 most performing E DH lines (μ_{10}) to mimic the mean value of released varieties from the elite program. Both μ and μ_{10} were reported as relative deviations to the average value observed at the end of the burn-in phase, expressed as percentage.

Data, Materials, and Software Availability. Genetic data are available at <https://doi.org/10.57745/58X9GT>. R functions to select optimal crosses and perform simulations are available at https://forgemia.inra.fr/dimitri.sanchez1/bridging_simulation_dent_amaizing_panel.

ACKNOWLEDGMENTS. D.S. PhD is jointly funded by the French ANRT, Lidea, and the 4 other private companies participating to the Pro-Mais ValRG project: KWS, Limagrain, MASSEEDS, and RAGT2N. We are grateful to Alain Murigneux, Stéphane Melkior, Laurent Bettinger, and Cyril Bauland for the coordination of the ValRG project and helpful discussion regarding practical breeding, as well as to Milagros Garcia and Carole Derue. We are grateful to Delphine Madur and Stéphane Nicolas for providing the Amaizing dent genotyping data used to conduct the simulations. We are grateful to Peter Rogowsky and Thomas Wiedez for helpful discussions regarding NBT and to Timothée Flutre and Jerome Enjalbert for helpful elements regarding diversity at the landscape scale. We thank the editor and anonymous reviewers for their helpful comments on earlier versions of the manuscript.

Author affiliations: ^aUniversité Paris-Saclay, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), Centre National de la Recherche Scientifique (CNRS), AgroParisTech, Génétique Quantitative et Evolution (GQE) Le Moulon, 91190, Gif sur Yvette, France; ^bUniversité Paris-Saclay, AgroParisTech, Institut National de la Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), UMR Mathématiques et Informatique Appliquées (MIA) Paris-Saclay, 91120, Palaiseau, France; and ^cSyngenta, Saint-Sauveur 31790, France

- J. Gayon, D. Zallen, The role of the Vilmorin Company in the promotion and diffusion of the experimental science of heredity in France, 1840–1920. *J. History Biol.* **31**, 241–262 (1998).
- J. L. Lush, *Animal breeding plans* (Collegiate Press Inc, 1937).
- C. Welckler, Physiological adaptive traits, a reservoir of alleles for maize genetic progress under challenging conditions. *Nat. Commun.*, in press.
- C. K. Khoury *et al.*, Crop genetic erosion: Understanding and responding to loss of crop diversity. *New Phytol.* **233**, 84–118 (2022).
- A. Allier *et al.*, Assessment of breeding programs sustainability: Application of phenotypic and genomic indicators to a North European grain maize program. *Theor. Appl. Genet.* **132**, 1321–1334 (2019).
- M. A. Mikel, J. W. Dudley, Evolution of North American dent corn from public to proprietary germplasm. *Crop. Sci.* **46**, 1193–1205 (2006).
- J.-L. Jannink, Dynamics of long-term genomic selection. *Genet. Sel. Evol.* **42**, 35 (2010).
- J. Rutkowski *et al.*, Genetic gain from phenotypic and genomic selection for quantitative resistance to stem rust of wheat. *Plant Genome* **8**, plantgenome2014.10.0074 (2015).
- Z. Lin *et al.*, Genetic gain and inbreeding from genomic selection in a simulated commercial breeding program for perennial ryegrass. *Plant Genome* **9**, plantgenome2015.06.0046 (2016).
- M. V. Mickelbart, P. M. Hasegawa, J. Bailey-Serres, Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. *Nat. Rev. Genet.* **16**, 237–251 (2015).
- S. L. Dwivedi *et al.*, Landrace germplasm for improving yield and abiotic stress adaptation. *Trends Plant Sci.* **21**, 31–42 (2016).
- S. McCouch *et al.*, Agriculture: Feeding the future. *Nature* **499**, 23–24 (2013).
- M. Toro, M. Pérez-Enciso, Optimization of selection response under restricted inbreeding. *Genet. Selection Evolution* **22**, 93–107 (1990).

14. N. Wray, M. Goddard, Increasing long-term response to selection. *Genet. Selection Evolution* **26**, 431 (1994).
15. T. H. E. Meuwissen, Maximizing the response of selection with a predefined rate of inbreeding. *J. Animal Sci.* **75**, 934–940 (1997).
16. S. A. Clark, B. P. Kinghorn, J. M. Hickey, J. H. van der Werf, The effect of genomic information on optimal contribution selection in livestock breeding programs. *Genet. Selection Evolution* **45**, 44 (2013).
17. J. A. Woolliams, P. Berg, B. S. Dagnachew, T. H. E. Meuwissen, Genetic contributions and their optimization. *J. Animal Breed. Genet.* **132**, 89–99 (2015).
18. A. Allier, L. Moreau, A. Charcosset, S. Teyssède, C. Lehermeier, Usefulness criterion and post-selection parental contributions in multi-parental crosses: Application to polygenic trait introgression. *G3 (Bethesda)* **9**, 1469–1479 (2019).
19. F. Schnell, H. Utz, *F1 Leistung und Elternwahl in der Zuchtung von Selbstbefruchtern* (Ber Arbeitstag Arbeitsgem Saatzüchtleiter, 1976).
20. A. Allier, C. Lehermeier, A. Charcosset, L. Moreau, S. Teyssède, Improving short- and long-term genetic gain by accounting for within-family variance in optimal cross-selection. *Front. Genet.* **10**, 1006 (2019).
21. H. D. Daetwyler, M. J. Hayden, G. C. Spangenberg, B. J. Hayes, Selection on optimal haploid value increases genetic gain and preserves more genetic diversity relative to genomic selection. *Genetics* **200**, 1341–1348 (2015).
22. M. Goiffon, A. Kusmec, L. Wang, G. Hu, P. S. Schnable, Improving response in genomic selection with a population-based selection strategy: Optimal population value selection. *Genetics* **206**, 1675–1682 (2017).
23. C. Wang, S. Hu, C. Gardner, T. Lübberstedt, Emerging avenues for utilization of exotic germplasm. *Trends Plant Sci.* **22**, 624–637 (2017).
24. M. M. Goodman, "Exotic Maize Germplasm: Status, Prospects, and Remedies" (1985).
25. H. T. Stalker, "Utilization of wild species for crop improvement" in *Advances in Agronomy*, N. C. Brady, Ed. (Academic Press, 1980), pp. 111–147.
26. F. Hospital, A. Charcosset, Marker-assisted introgression of quantitative trait loci. *Genetics* **147**, 1469–1485 (1997).
27. G. Charmet *et al.*, Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. *Theor. Appl. Genet.* **99**, 1143–1148 (1999).
28. B. Servin, O. C. Martin, M. Mézard, F. Hospital, Toward a theory of marker-assisted gene pyramiding. *Genetics* **168**, 513–523 (2004).
29. R. Dormatey *et al.*, Gene pyramiding for sustainable crop improvement against biotic and abiotic stresses. *Agronomy* **10**, 1255 (2020).
30. J. W. Dudley, A method of identifying lines for use in improving parents of a single cross 1. *Crop Science* **24**, crops1984.0011183X002400020034x (1984).
31. J. W. Dudley, Modification of methods for identifying inbred lines useful for improving parents of elite single crosses 1. *Crop. Sci.* **27**, crops1987.0011183X002700050024x (1987).
32. M. Cramer, L. Kannenberg, 5 Years of hope - the hierarchical open-ended corn breeding system. *Crop. Sci.* **32**, 1163–1171 (1992).
33. R. Bernardo, Genomewide selection for rapid introgression of exotic germplasm in maize. *Crop. Sci.* **49**, 419–425 (2009).
34. J. W. R. Martini, T. L. Molnar, J. Crossa, S. J. Hearne, K. V. Pixley, Opportunities and challenges of predictive approaches for harnessing the potential of genetic resources. *Front. Plant Sci.* **12**, 674036 (2021).
35. A. Allier, S. Teyssède, C. Lehermeier, A. Charcosset, L. Moreau, Genomic prediction with a maize collaborative panel: Identification of genetic resources to enrich elite breeding programs. *Theor. Appl. Genet.* **133**, 201–215 (2020).
36. R. Bernardo, Prospective targeted recombination and genetic gains for quantitative traits in maize. *Plant Genome.* **10** (2017).
37. A. Allier, S. Teyssède, C. Lehermeier, L. Moreau, A. Charcosset, Optimized breeding strategies to harness genetic resources with different performance levels. *BMC Genomics* **21**, 349 (2020).
38. D. Vanavermaete, J. Fostier, S. Maenhout, B. De Baets, Deep scoping: A breeding strategy to preserve, reintroduce and exploit genetic variation. *Theor. Appl. Genet.* **134**, 3845–3861 (2021).
39. I. S. Breider, A multi-part strategy for introgression of exotic germplasm into elite plant breeding programs using genomic selection. Research Square [Preprint] (2022) <https://doi.org/10.21203/rs.3.rs-1246254/v1> (Accessed 18 January 2022).
40. T. Pook, M. Schlather, H. Simianer, MoBPS - Modular breeding program simulator. *G3 Genes, Genomes, Genetics* **10**, 1915–1918 (2020).
41. H. De Beukelaer, Y. Badke, V. Fack, G. De Meyer, Moving beyond managing realized genomic relationship in long-term genomic selection. *Genetics* **206**, 1127–1138 (2017).
42. G. Gorjanc, R. C. Gaynor, J. M. Hickey, Optimal cross selection for long-term genetic gain in two-part programs with rapid recurrent genomic selection. *Theor. Appl. Genet.* **131**, 1953–1966 (2018).
43. S. Rio, T. Mary-Huard, L. Moreau, A. Charcosset, Genomic selection efficiency and a priori estimation of accuracy in a structured dent maize panel. *Theor. Appl. Genet.* **132**, 81–96 (2019).
44. X. Yu *et al.*, Genomic prediction contributing to a promising global strategy to turbocharge gene banks. *Nat. Plants* **2**, 16150 (2016).
45. ECPGR, *ECPGR: European Evaluation Network (EVA)* <https://www.ecpgr.cgiar.org/european-evaluation-network-eva> (Accessed 1 August, 2022).
46. BOLD, *BOLD - Crop Trust* <https://www.croptrust.org/work/projects/the-bold-project/> (Accessed 1 August 2022).
47. E. J. Millet *et al.*, Genomic prediction of maize yield across European environmental conditions. *Nat. Genet.* **51**, 952–956 (2019).
48. S. P. Brandariz, R. Bernardo, Maintaining the accuracy of genomewide predictions when selection has occurred in the training population. *Crop. Sci.* **58**, 1226–1231 (2018).
49. H.-Y. Hung *et al.*, The relationship between parental genetic or phenotypic divergence and progeny variation in the maize nested association mapping population. *Heredity* **108**, 490–499 (2012).
50. C. Lehermeier *et al.*, Usefulness of multiparental populations of maize (*Zea mays* L.) for genome-based prediction. *Genetics* **198**, 3–16 (2014).
51. D. Habier, R. L. Fernando, J. C. M. Dekkers, The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**, 2389–2397 (2007).
52. D. Houle, B. Morikawa, M. Lynch, Comparing mutational variabilities. *Genetics* **143**, 1467–1483 (1996).
53. W. G. Hill, Is continued genetic improvement of livestock sustainable? *Genetics* **202**, 877–881 (2016).
54. C. J. Goodnight, Epistasis and the effect of founder events on the additive genetic variance. *Evolution* **42**, 441–454 (1988).
55. A. I. Seye, C. Bauland, A. Charcosset, L. Moreau, Revisiting hybrid breeding designs using genomic predictions: Simulations highlight the superiority of incomplete factorials between segregating families over topcross designs. *Theor. Appl. Genet.* **133**, 1995–2010 (2020).
56. D. C. Kadam, A. J. Lorenz, "Toward redesigning hybrid maize breeding through genomics-assisted breeding" in *The Maize Genome*, J. Bennetzen, S. Flint-Garcia, C. Hirsch, R. Tuberosa, Eds. (Springer International Publishing, 2018), pp. 367–388.
57. M. Mayer *et al.*, Discovery of beneficial haplotypes for complex traits in maize landraces. *Nat. Commun.* **11**, 4954 (2020).
58. T. Li *et al.*, Domestication of wild tomato is accelerated by genome editing. *Nat. Biotechnol.* **36**, 1160–1163 (2018).
59. H. Yu *et al.*, A route to de novo domestication of wild allotetraploid rice. *Cell* **184**, 1156–1170.e14 (2021).
60. D. Rodríguez-Leal, Z. H. Lemmon, J. Man, M. E. Bartlett, Z. B. Lippman, Engineering quantitative trait variation for crop improvement by genome editing. *Cell* **171**, 470–480.e8 (2017).
61. M. W. Ganal *et al.*, A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the b73 reference genome. *PLoS One* **6**, e28334 (2011).
62. R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. (version 4.2.1)
63. H. Giraud *et al.*, Linkage disequilibrium with linkage analysis of multiline crosses reveals different multiallelic QTL for hybrid performance in the flint and dent heterotic groups of maize. *Genetics* **198**, 1717–1734 (2014).
64. H. Wang, I. Mizstal, I. Aguilar, A. Legarra, W. M. Muir, Genome-wide association mapping including phenotypes from relatives without genotypes. *Genet. Res.* **94**, 73–83 (2012).
65. C. Lehermeier, S. Teyssède, C.-C. Schön, Genetic gain increases by applying the usefulness criterion with improved variance prediction in selection of crosses. *Genetics* **207**, 1651–1661 (2017).