Back to the future: Implications of genetic complexity for hybrid breeding strategies

Frank Technow^{*}, Dean Podlich[†] and Mark Cooper[‡]

Abstract Commercial hybrid breeding operations can be described as decentralized networks of smaller, more or less isolated breeding programs. There is further a tendency for the disproportionate use of successful inbred lines for generating the next generation of recombinants, which has led to a series of significant bottlenecks, particularly in the history of the North American and European maize germplasm. Both the decentralization and the disproportionate inbred use reduce effective population size and constrain the accessible genetic space. Under these conditions, long term response to selection is not expected to be optimal under the classical infinitesimal model of quantitative genetics. In this study we therefore aim to propose an alternative rational for the success of large breeding operations in the context of genetic complexity arising from the structure and properties of interactive genetic networks. For this we use simulations based on the *NK* model of genetic architecture. We indeed found that constraining genetic space and reducing effective population size, through program decentralization and disproportionate inbred use, is required to expose additive genetic variation and thus facilitate heritable genetic gains. These results introduce new insights into why the historically grown structure of hybrid breeding programs was successful in improving the yield potential of hybrid crops over the last century. We also hope that a renewed appreciation for “why things worked” in the past can guide the adoption of novel technologies and the design of future breeding strategies for navigating biological complexity.

1 Introduction

Pioneered by Shull (1908), hybrid breeding is credited as one of the most significant factors for the tremendous productivity increases of major field crops that enabled food production to keep pace with population growth. Hybrid breeding programs originally were centred around maximum exploitation of heterosis, a phenomenon that remains largely unexplained even after a century of research (East, 1936; Lippman and Zamir, 2007). This later evolved into the modern concept of hybrid breeding, characterized by its distinctive structuring of germplasm into heterotic groups and patterns (Melchinger and Gumber, 1998). Beyond heterotic groups, the structure of commercial hybrid breeding, particularly in major crops like maize, is characterized by the largely isolated and unique sub-heterotic patterns of the major companies (White et al., 2020) as well as a high degree of decentralization into smaller, more or less disconnected sub-programs within those (Cooper et al., 2014). Plant breeders further have a tendency for relying on only a small set of elite inbred lines for producing the next generation of recombinants (Rasmusson and Phillips, 1997), lead-

ing to a series of significant bottleneck events in the history of, for example, the North American maize germplasm (White et al., 2020). These characteristics drastically reduced the effective population size within breeding programs and are not predicted to be promising strategies under the additive, infinitesimal model of quantitative genetics (Gaynor et al., 2017). Nevertheless, consistent long-term genetic gain has been demonstrated (Duvick et al., 2004).

To better describe and quantify the observed genetic variation among hybrids, the concept of general and specific combining ability was developed early on (Sprague and Tatum, 1942). The former, commonly abbreviated as GCA, is a property of the additive effects of contributing genes and describes the average performance of all hybrids derived from an inbred. The latter, commonly abbreviated as SCA, is a non-additive residual term that describes the deviation of the performance of a particular hybrid from the expectation based on the parental GCA values.

Running efficient hybrid breeding programs requires a preponderance of additive genetic variation to maximize response to selection in the next generation of inbred lines (Falconer and Mackay, 1996) as well as the predictability of hybrid performance from the GCA of inbred lines (Reif et al., 2007). A preponderance of GCA variation also allows identification of inbreds that can serve as parents of several high performing hybrids. This greatly simplifies production of commercial seed, which is a major

^{*}Plant Breeding, Corteva agriscience, Tavistock, ON, Canada; frank.technow@corteva.com

[†]Systems and Innovation for Breeding and Seed Products, Corteva agriscience, Johnston, IA, USA

[‡]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia, QLD, Australia

challenge for many crops (Technow, 2019). Therefore, hybrid breeding programs have traditionally relied on maximizing and exploiting GCA variation (Falconer and Mackay, 1996; Melchinger, 1999).

The historically grown paradigms around hybrid breeding designs and strategies are now being challenged by innovative concepts (e.g., Gaynor et al., 2017; Wallace et al., 2018; Hickey et al., 2019; Voss-Fels et al., 2019; Seye et al., 2020) devised in the wake of technological advances such as whole genome prediction (Meuwissen et al., 2001), high-throughput phenotyping (Araus and Cairns, 2014) and genotyping (Poland and Rife, 2012), as well as gene editing (Jaganathan et al., 2018). While some of these concepts are highly speculative and might not live up to expectations (Bernardo, 2016), it is clear that the next decades will change plant breeding. However, before implementing drastic changes to breeding programs we require a theoretical and simulation framework to explore and understand the structures and strategies that have contributed to the success of long term genetic gain and germplasm improvement. From this historical basis we can evaluate novel proposals and draw lessons for design of future breeding strategies.

Empirical reports show a preponderance of additive variation in many wild, domesticated and laboratory species (Falconer and Mackay, 1996; Lynch and Walsh, 1998; Hill et al., 2008). This agrees well with published studies showing a preponderance of additive GCA over non-additive SCA variation in hybrid breeding programs (Technow et al., 2014; Laripe et al., 2017). At the same time, however, advances in plant physiology and molecular and systems biology have stimulated a renewed appreciation of the intricate interactions at the molecular, metabolic and physiological level that underlie complex traits (Hammer et al., 2006; Carlborg and Haley, 2004; Phillips, 2008; Wilkins et al., 2016; Saha et al., 2011; Jiang et al., 2017). Of particular relevance for hybrid breeding are recent studies indicating that heterosis is an emergent property of complex metabolic networks (Fivet et al., 2010, 2018; Vacher and Small, 2019).

The paradox between the complexity of the underlying biology and the simplicity of the expressed variation can of course be resolved by distinguishing between biological and statistical effects and realizing that the former cannot be inferred from the latter (Wade, 2002; Mackay, 2014; Huang and Mackay, 2016). Statistical effects of genes, as well as their aggregates such as GCA and SCA and their vari-

ances depend on the genetic background of the population in which they are evaluated, particularly on allele frequencies and linkage disequilibrium (LD) patterns (Falconer and Mackay, 1996). For example, it was shown that, regardless of the underlying genetic architecture, genetic variances in random mating populations are expected to be predominantly additive when genes are at extreme frequencies and linkage disequilibrium is high (Hill et al., 2008).

Thus, ratios of additive to non-additive variation are not intrinsic properties of biological systems but at least partly a function of allele frequencies and LD patterns and thus dependent on breeding strategies. Because of the importance of additive variation for efficient operation of breeding programs, a framework to evaluate and study breeding strategies should allow for the possibility of additivity arising from high degrees of biological complexity at the genetic level.

In this study we will use simulations based on the NK model of genetic complexity (Kauffman, 1993) to explore two *themes* representing key, historically grown, characteristics of hybrid breeding: firstly its decentralization into smaller, more or less independent sub-programs ("*decentrality* theme") and secondly the disproportional use of superior inbred lines for producing the next generation of recombinants ("*inbred usage* theme"). Our goal thereby is not to make specific recommendations for optimal structuring of programs, but rather to gain an appreciation for the properties of these structures in the context of different degrees of genetic complexity.

Material and Methods

Model of genetic complexity

The NK model, introduced by Kauffman (1993) will form the basis of the simulations. The NK model allows generation of a tunable series of models of trait genetic architecture with increasing dimensionality and complexity by varying the number of genes N (dimensionality) and the degree of interaction among them (K , complexity).

The *genetic landscape* metaphor was introduced and developed by Wright (1932) to aid conceptualizing genetic complexity in high dimensions. As a metaphor it should not be taken literally but can help to gain an intuition for the complexity and ruggedness associated with increasing values of K

Mount Fuji ($K = 1$)

The Alps (intermediate K)

The Dunes (high K)

Figure 1 Schematic visualization of genetic landscapes corresponding to different values of complexity parameter K .

(Kauffman, 1993) as well as making the rather abstract concepts discussed henceforth more tangible. At $K = 1$ (special case of additive gene action), the genetic landscape can be imagined as that of Mount Fuji, i.e., a single, clearly distinguished peak with a steady and monotonous incline to the top (Figure 1). At intermediate K levels, the landscape is characterized by multiple peaks clustered together in a certain region of genetic space. This might be visualized as akin to the European Alps, i.e., a mountainous region within an otherwise flat landscape. Finally, at high value of K , the landscape resembles a sea of dunes, i.e., a range of peaks of similar height and shape distributed more or less evenly in space.

We implemented the NK model according to the generalized approach described by Altenberg (1994), but adapted the model to accommodate diploid genomes. Here, the complex trait is described as a normalized sum of a set of “fitness components”. The value of each fitness component is computed as a function of K interacting genes drawn at random from all N genes. Following Altenberg (1994), the specific fitness values were calculated with *random functions* derived from the *ran4* pseudo-random number generator (Press et al., 1992) and are distributed uniformly between 0 and 1. For this study, the number of fitness components and the number of genes were both set to 500. Genes were biallelic and the simulated organism diploid. The complexity parameter K was varied from 1 to 15 in steps of 1 (i.e. creating genetic landscapes ranging in complexity from Mount Fuji to The Dunes; Figure 1). With the exception of $K = 1$, this parameter

was used as the rate parameter in a Poisson distribution from which the number of interacting genes was drawn independently for each fitness component. The sampled values were then truncated to fall within a range of 1 and 15. The identities of the interacting genes were drawn at random from the total set. Thus genes typically influenced multiple fitness components (i.e., act pleiotropically). For $K = 1$, each of the 500 genes was assigned to exactly one fitness component and the values of heterozygous allele configurations constrained to be midway between the homozygous configurations. Thus, $K = 1$ represents the special case of additive gene action. Using order statistics, the expected value of the maximum of two samples from a Uniform distribution between 0 and 1 is $2/3$. Thus, the expected maximum attainable fitness for the $K = 1$ special case is $2/3$.

The complexity of the generated NK models was quantified following the “one-mutant neighbour” hill-climbing algorithm described by Kauffman (1993), but adapted to diploid organisms. A randomly generated genotype was used as the starting value. From there, all possible genotypes were generated that differ from the initial genotype by one allele at one of the 500 loci. Thus, a homozygous locus was changed to the heterozygous state while a heterozygous locus was changed to both alternate homozygotes. Then, the fitness values of all one-allele neighbours were evaluated according to the defined NK model and an improved genotype chosen at random from all fitter one-allele neighbours. This process was repeated until no fitter one-mutant neighbour could be found, meaning that the search reached a local or global optimum. For each level of K , 100 NK models were generated independently and a minimum of 65 searches, each starting at a random initial genotype, were conducted for each. The statistics recorded were the average number of steps until a local optimum was reached, the average Hamming genotypic distance, i.e., the normalized number of differing genome positions (Pinheiro et al., 2005), among optima and the correlation between the fitness values of the optima and the Hamming distance to the highest optimum identified (Kauffman, 1993).

The average Hamming distance between the local peaks increased from just below 0.5 at $K = 2$ to $2/3$ at around K of 6 or 7 and remained constant at this value from there on (Figure 2A). Note that with three different genotypes at each locus, $2/3$ is the expected value of the Hamming distance between

Figure 2 Relationship between the NK model complexity parameter K and (A) the average genetic Hamming distance between local peaks, (B) the average correlation between the fitness values of local peaks and their genetic Hamming distance to the highest peak and (C) the average number of steps to a local peak. Panels (A) and (B) omit results for $K = 1$, for which only a single peak exists. The dotted lines in these panels indicate values of $2/3$ and 0.0 , respectively.

randomly generated genotypes. Similarly, the correlation between the fitness values of the local peaks and their Hamming distance to the highest identified peak increased from -0.25 at $K = 2$ to zero at $K = 9$ (Figure 2B). Here, a negative correlation means that local peaks with higher fitness tend to be found near each other and clustered around the highest peak. Further, a zero correlation indicates that there is no clustering of the peaks and proximity to the highest peak. Therefore, local peaks are randomly distributed throughout the genetic landscape. Thus, somewhere between $K = 6$ and $K = 9$, the landscape shifts from one in which local peaks tend to cluster together, to one where local peaks of arbitrary height can exist anywhere in genetic space. The average number of steps until a local peak was reached decreased with K from 500 at $K = 1$ to just 167 at $K = 15$ (Figure 2C). Note that 500 is the expectation at $K = 1$, the special case of additive gene action, when starting from randomly generated genotypes, because $1/3$ of the 500 loci are already at their highest possible value, $1/3$ are one step removed (the heterozygous genotypes) and $1/3$ are two steps removed (the lower homozygotes). Thus, the complexity and ruggedness of the genetic landscapes increase further after they become uncorrelated around K of 6 to 9.

The simulated genome comprised 10 diploid chromosomes of 1 Morgan length each. Each of the chromosomes received a random subset of 50 of the 500 genes, which were distributed evenly across the chromosome. Recombination was simulated according to the Haldane mapping function with the R package “hypred” (Technow, 2013), in the version available from the supplement of Technow and Gerke (2017).

Simulation of hybrid breeding process

The simulation process is visualized in Figure 3. The starting point of the simulation was a base population of inbred lines of size 1,000. This population was simulated stochastically as described by Montana (2005) to result in an expected LD between two loci t Morgan apart equal to $r^2 = 0.5 \cdot 2^{-t/0.1}$ and with minor allele frequencies distributed uniformly between 0.35 and 0.50. The lines from the base population were then separated at random into two heterotic groups (arbitrarily labelled ‘1’ and ‘2’) and further into sub-populations within those. The size of those sub-populations depended on the scenario. One sub-population from one heterotic group was then paired with one sub-population from the other group to form sub-heterotic patterns. These population pairs will henceforth be referred to as “breeding programs”. Hybrids were produced strictly across heterotic groups, by crossing lines from one sub-population of a program with lines from the other. Breeding crosses, i.e., crosses to generate a new generation of recombinant lines were done within and among sub-populations, depending on the scenario but strictly within heterotic groups. The simulation of the breeding process described above is an approximation of the structure and evolution of long term hybrid breeding akin to what we have observed in practice; i.e. starting from an initial germplasm base, separation into distinct heterotic groups and future separation into sub-populations.

The GCA of the lines was evaluated with an incomplete mating design (Melchinger et al., 1987; Seye et al., 2020) by performing 10 crosses per line with random partners from the opposite sub-population of the same program. The performance of the resulting hybrids, as determined according to

Figure 3 Schematic visualization of simulated hybrid breeding process.

used as a metric of genetic gain. This metric reflects that commercial breeding programs release only a handful of hybrid products each cycle.

Breeding crosses among inbred lines for initiating the next recombination cycle were chosen by assigning each inbred line a usage probability, which was a product between an individual and population level relative contribution value. To determine the former, the lines within each sub-population were ranked according to their observed GCA values. Only the top lines, how many depended on the scenario, were selected as potential parents, the remainder given an individual contribution value of zero. The relative contributions of the selected lines from a given sub-population were drawn from a Dirichlet distribution. The concentration parameters of this distribution were used to modulate the relationship between selection rank and relative contribution. Further details about this will be given later when describing the setting for the “inbred usage theme”.

The population level contribution values describe the overall contribution of lines from one sub-population to the breeding crosses of another. They are thus defined anew for each target population and hence the contribution value of population ‘A’ to the crosses for population ‘B’ might be different than that to the crosses of population ‘C’. The process will be explained using the example visualized in Figure 3. Here there are three programs (labelled ‘A’, ‘B’ and ‘C’, with subscript 1 or 2 indicating the heterotic group).

The programs are ranked from highest to lowest performing according to the average performance of the selected set of experimental hybrids, as described above. Germplasm, in the form of lines used as crossing partners, is exchanged only from higher performing to lower performing programs. Specifically, the amount of crosses with lines from other programs increased from zero for the best performing program (A) to a proportion of P_{max} for the lowest performing program (C), with intermediate programs staggered equidistantly between. In the example, $P_{max} = 50\%$. Thus, program A will perform no crosses with lines from other programs, program B will use lines from other programs in 25% of its new crosses and program C in 50% of its crosses. How much of that overall proportion was derived from each of the other programs was proportional to the relative performance differences. In the example, the difference between program C and A is twice as large as that between C and B, thus, lines from program A were used in twice as many crosses

the defined NK model, was then averaged. Finally, a normally distributed noise variable with zero mean and variance equal to one third of the variance of the GCA values of that sub-population was added to those averages to represent experimental and environmental noise. The so obtained values were used as observed GCA values. Those GCA estimates were then used to predict the performance of all possible inter-group hybrids of that program. The top hybrids, how many exactly depended on the scenario, were then selected and their true performance determined according to the NK model. The average of this select group of hybrids, which represents a set of advanced experimental hybrids, was used to quantify the overall performance of the program in the current cycle. The maximum true performance of the selected hybrids from all programs was defined as the peak performance of the whole breeding operation in the current cycle and

than lines from program B (33% from A and 17% from B for a total of 50%). This process thus reflects that highly successful programs tend to exploit their own genetics while less successful programs have more of an incentive to explore superior genetics from other programs.

The relative individual contributions were then multiplied with the relative population contributions to arrive at a final relative contribution value for each line to the crosses of a given population. The actual breeding crosses were then determined by sampling the lines with probabilities proportional to their contribution values. This was done with replacement, meaning that the same cross could have been made multiple times, but excluding crosses that would result in selfings. One recombinant line was derived from each crossing through seven generations of single seed descent selfing, followed by a final doubled haploid step (Dwivedi et al., 2015) to result in fully homozygous inbred lines. This new generation of recombinants fully replaced the previous generations, i.e., a line was considered as a crossing partner in only one generation. The so obtained new recombinants then form the next breeding cycle. The simulations were conducted for 30 cycles in total and repeated independently at least 500 times for each scenario studied. All computations were conducted in the R environment for statistical computing (R Core Team, 2018).

Recorded metrics

In addition to the already described true performance of the best identified hybrid, which was used as a measure of *peak performance* in a given cycle, several other measures were recorded to describe and understand the dynamics of the system.

The *proportion of GCA to total genetic variance* (%GCA) describes the amount of exploitable additive genetic variation currently available. It was estimated using the hybrids generated for evaluating the GCA of the inbred lines. For this, the following mixed model was fitted: $h_{ij} = \mu + g_i + g_j + e_{ij}$, where h_{ij} was the true performance of the hybrids, μ the overall mean, g_i and g_j were the GCA effects of the parents from the two heterotic groups and e_{ij} a residual term. Because the true genetic performances of the hybrids were used, e_{ij} corresponds to the SCA component. The model was fitted using the R package “lme4” (Bates et al., 2015) and %GCA then calculated as $(V_{g_i} + V_{g_j}) / (V_{g_i} + V_{g_j} + V_{e_{ij}})$,

where V_{g_i} etc. were the estimated variance components. In scenarios with multiple programs, %GCA was estimated separately for each and then averaged to arrive at a single estimate for each cycle.

The *modified Rogers’ distances* (Reif et al., 2005) between the heterotic groups within each program were used as measures of heterotic group divergence. The distances were calculated for all programs and averaged to arrive at representative value for that cycle.

To describe the distribution of allele frequencies within each sub-population and hence the amount of available allelic diversity we calculated the proportion of loci with a minor allele frequency of less than 5%. This probability measures the thickness of the extreme tail of the allele frequency distribution and thus reflects the degree with which it follows a ‘U-shape’ (Hill et al., 2008). This metric was evaluated for all sub-populations in each cycle and then averaged.

As a more high-level diversity metric we considered the *effective population size* (N_e) of each sub-population. N_e was calculated according to the method described by Corbin et al. (2012) for estimating constant effective population size. The so obtained values were averaged across sub-populations.

Hybrid breeding ‘themes’

All previously described parameters, such as parameters related to the NK model and genetic architecture, parameters related to testcross evaluation, etc., were kept constant across the themes investigated.

In the *decentrality* theme we explored consequences of separating hybrid breeding programs into smaller, more or less isolated, units. We defined three distinct strategies for ‘searching’ (Podlich and Cooper, 1999) genetic space (Figure 4): a single large program (*centralized search*) to multiple smaller, fully isolated programs (*isolated search*). Between these two extremes we considered a strategy with multiple smaller programs that exchange germplasm in the form of breeding crosses (*distributed search*).

The centralized search was characterized by a single program consisting of one sub-population per heterotic group. The size of each was 500, for a total of 1,000 lines generated in each cycle. The number of lines selected to contribute to the next generation was 125 per sub-population. The relative individual contributions of these lines decreased proportionally with their performance ranks. The number

Figure 4 Schematic visualization of the three general search strategies explored in the *decentrality* theme.

of selected experimental hybrids was 125. The isolated search strategy comprised five programs, each with one sub-population per heterotic group. The sub-population size was 100 of which 25 were selected. Also here, the relative individual contributions of the lines were proportional to their performance ranks. The number of experimental hybrids selected per program was 25.

Figure 5 Distributions of relative individual contributions of selected inbred lines considered in the *inbred usage* theme.

In the distributed search we considered three levels of Pmax: 25%, 50% and 75%. The number of programs as well as lines and hybrids created and selected for each followed those of the isolated search. Note, thus, that the total number of lines and hybrids were the same across all scenarios as was the selection intensity.

In the *inbred usage* theme we explored the consequences of different degrees of imbalance in the relative contributions of the selected inbred lines

to the next generation. The different scenarios explored correspond to the distributed search strategy with Pmax = 50%. Only the relative usage of inbred lines was varied. As described above, the observed relative contributions were drawn from a Dirichlet distribution with concentration parameter chosen in a way to result in a certain average relationship between relative contribution and performance rank. Three scenarios were considered (Figure 5). In the *balanced* usage scenario, all selected inbreds contributed equally on average, in the *proportional* scenario, the relative contribution declined proportional with the performance rank of the lines. In the *disproportional* scenario, contributions halved with every 5 ranks, meaning that the highest performing line will contribute twice as much to the next generation as the 5th ranked line. The increasing imbalance in contributions can be quantified as $1/b'b$ (with b being the vector of relative contributions), which is an estimate of the effective number of contributing lines (Boichard et al., 1997). For the balanced scenario, this was 25 and thus equal to the actual number of selected lines within each sub-population. It decreased to 19.1 and 13.6 for the proportional and disproportional scenarios, respectively.

Results

Decentrality theme

Which strategy achieved the highest peak performance depended on the value of the complexity parameter K , with the centralized strategy being superior at low $K < 5$, the distributed strategy at intermediate K and the isolated strategy at high values of K above eight (Figure 6A). The differences between the strategies tended to increase with increasing K . The centralized and distributed search

Figure 6 Relationship between the NK model complexity parameter K (average number of interacting genes) and peak genetic performance in the last cycle for the strategies explored in the decentrality theme: (A) comparing the isolated, distributed and centralized strategies and (B) the different values of P_{max} within the distributed strategy. The curve of the distributed strategy in (A) is an average across the three P_{max} scenarios within it.

strategies came very close to reaching the theoretical maximum peak performance at the special case of additivity ($K = 1$), but the isolated strategy remained considerably below that. Within the distributed strategy, the highest P_{max} value of 75% was superior at K values below eight and the lowest P_{max} of 25% at high K (Figure 6B). The case of $P_{max} = 50\%$ had peak performance in between the two extremes, but more similar to $P_{max} = 75\%$. All P_{max} scenarios achieved virtually identical peak performance at $K = 1$.

For brevity, trajectories across cycles are shown only for K values of 1, 6 and 15, representing the additive, multi-peaked but clustered and fully uncorrelated landscapes, respectively (Figure 1). Results for all values of K are available as supplemental information (File S1). At $K = 1$, the centralized search strategy had the highest peak performance in all cycles, closely followed by the three versions of the distributed search (Figures 8A, B, C). The peak

performance of the isolated search strategy was considerably lower than that of the other strategies as it increased at a lower rate and seemed to reach a plateau at around cycle 20. At $K = 6$, the isolated search strategy achieved the highest peak performances in the earlier cycles but was overtaken by the distributed search strategies later. Those had very similar peak performances until the last few cycles when the version with P_{max} of 25% fell behind. The centralized search had the lowest peak performances throughout, with the differences to the other strategies being particularly large between the intermediate cycles 15–20. Finally, at $K = 15$, only the isolated search had a sizable increase in peak performance cycle over cycle. The distributed search strategies showed an increase only in the last few cycles and the centralized strategy did not increase peak performance at all.

As expected %GCA was equal to one for all scenarios at $K = 1$ (Figures 8D, E, F). At $K = 6$, %GCA started at just below 10% and increased from there with each cycle. The rate of increase was greatest for the isolated strategy which reached almost 100% in the final cycles. The centralized search strategy had the slowest increase and was still below 50% in the final cycle. The distributed search strategies were intermediate between these two extremes. The increase was steepest for $P_{max} = 25\%$ case, which translated to it having a markedly higher %GCA than the $P_{max} = 50\%$ case and $P_{max} = 75\%$ case during intermediate cycles 15–20. However, all three converged to a similar value of around 80% in the final cycle. At $K = 15$, %GCA started at zero and only the isolated strategy saw a marked increase in early cycles. The distributed search strategies saw an increase in %GCA noticeably above zero only in the final cycles and the centralized strategy remained at zero throughout.

The percent of loci with $MAF < 0.05$ increased over cycles for all strategies and complexity levels (Figures 8G, H, I). In all cases, the increase over cycles was strongest for the isolated search strategy, where it reached close to 100% in the final cycles and weakest in the centralized search strategy. The curves for the three P_{max} levels of the distributed search strategy were similar to each other and intermediate compared to the two other strategies. The differences between the strategies increased with K because the increase in the proportion of loci at extreme frequencies slowed for the distributed and centralized strategies with increasing K . At the highest levels of complexity, only between 30% and

Figure 7 Modified Rogers Distance between heterotic groups as a function K , averaged across programs, in the connectivity theme. The curve of the distributed strategy represents the average across the three Pmax levels.

40% of loci showed a $MAF < 0.05$ in the different distributed strategies and less than 20% in the centralized strategy.

The modified Rogers Distance between heterotic groups increased over cycles for all strategies (Figures 8J, K, L). For all levels of complexity, this distance was highest for the isolated strategy and lowest for the centralized strategy, with the three versions of the distributed search having similar values that were intermediate to the two extremes (Figure 7).

The N_e differences among the strategies remained largely constant across cycles and levels of K . For the sake of brevity results will only be reported for cycle 15 and $K = 7$. The estimated N_e for each sub-population for the isolated search strategy was 20.0, for the three versions of the distributed search it was 23.7 (Pmax = 25%), 31.3 (Pmax = 50%) and 35.4 (Pmax = 75%), respectively, and for the centralized search strategy 98.3.

Inbred usage theme

For brevity sake, only results for K of 1, 6 and 15 are shown (results for all values of K are provided in supplemental file S2). Again, which inbred usage scenario achieved the highest peak performance depended on the complexity level K (Figure 9). At the additive case of $K = 1$, all strategies achieved very similar peak performances close to the theoretical maximum of $2/3$. Until $K = 8$, the highest peak performances were reached with proportional usage of selected inbred lines. For $K > 8$, disproportional use of inbred lines resulted in the highest peak per-

formances. Balanced use generally resulted in the lowest peak performance, except for $K < 4$, where this strategy was slightly ahead of the disproportional usage strategy. The differences between the strategies tended to increase with K .

The cycle over cycle increase in peak performance was initially higher the more disproportional the use of the inbreds (Figures 10A, B, C). However, except for the highest level of complexity, this did not result in the highest maximum performance for this scenario, because the increase started to level off in the last five to ten cycles. Scenarios with proportional and balanced use of inbreds therefore had the highest peak performance at $K = 1$, though the differences were small. At the intermediate level of $K = 6$, the disproportional use scenario was overtaken by the proportional use scenario in the last cycles. The differences between these two were small, however. Finally, at $K = 15$, only the disproportional use scenario achieved a sizable increase in peak performance.

At $K = 1$, %GCA stayed constant at one for all scenarios, as expected. At $K = 6$, %GCA increased most strongly for disproportional use, followed by proportional and balanced use (Figures 10D, E, F). Reaching above 90% for the former, and above 80% and 50% for the latter two, respectively. At $K = 15$, %GCA remained near zero for the balanced and proportional use scenarios throughout. For the disproportional use scenario, it remained at zero as well until cycle ten and increased from there to almost 60%.

For all strategies and values of K , the percent of loci with a $MAF < 0.05$ increased from its initial value of zero (Figures 10G, H, I). The increase over cycles was strongest for disproportional inbred use, for which it reached close to 100% at K of 1 and 6. The proportional usage strategy had the second strongest increase and the balanced strategy the weakest. As was the case in the decentrality theme, the differences between the strategies tended to increase with K . At the highest level of $K = 15$, the proportional and balanced strategies stayed below 40%, whereas the disproportional usage strategy reached close to 80%.

The modified Rogers distance between heterotic groups increased over cycles in all scenarios (Figures 10J, K, L). Throughout it was highest for disproportional use, followed by proportional and balanced use. Overall, the distance was greatest for the intermediate complexity level of $K = 6$.

N_e , again reported only for cycle 15 at $K = 7$, was

Figure 8 Evolution of metrics over cycles in the decentrality theme for scenarios with K of 1 (left column), 6 (middle) and 15 (right).

Figure 9 Relationship between the *NK* model complexity parameter *K* and peak genetic performance in the last cycle for the strategies explored in the the inbred usage theme.

23.2, 31.3 and 44.6, for the disproportional, proportional and balanced usage scenarios, respectively.

Discussion

The objective of this study was to explore properties of the historically grown structure of commercial hybrid breeding programs, particularly in maize, and aid the understanding of why these structures successfully generated significant amounts of genetic gain in the past and thereby impact global food security. The infinitesimal framework (Barton et al., 2017), in which traits are described as the sum of a large number of genes all having additive, context independent effects of similar magnitude, is our starting point. However, we seek extensions to account for the empirical observations that a) there are results observed from operating a long-term breeding effort that are not consistent with or not easily explained within the infinitesimal model framework (Rasmusson and Phillips, 1997) and b) reflect the reality of a highly complex trait biology (Hammer et al., 2006). Therefore, we are motivated to consider the influence of complexity of trait genetic architecture on breeding strategies from the perspective of a long-term commercial breeding program (Duvick et al., 2004; Cooper et al., 2014).

Emergence of additivity

As a representation of genetic complexity we chose the *NK* model framework developed by Kauffman

(1993), which allows exploration of the full continuum from complete additivity to deep and almost intractable genetic complexity. For reference, the *NK* models used in this study, corresponded to the Mount Fuji landscape at $K = 1$ (Figure 1) and to the 'Alps' landscape from $K = 2$ to K of 7 or 8. After this the genetic models transitioned from the multi-peaked but correlated 'Alps' landscape to the uncorrelated landscape represented by the 'Dunes' metaphor (Figures 1 and 2).

In complex genetic landscapes, additive genetic variance, the sine qua non of genetic gain, is not a constant factor of trait biology (i.e., deducible from the molecular properties of genes) but rather emerging from the interplay of biology and natural or artificial properties of population structure (Wade, 2002; Cooper et al., 2005). In particular, additivity emerges in response to a constraining of the dimensionality of genetic space, or, in other words, by limiting genetic diversity. In practice, such constraints in dimensionality are achieved through fixation or near fixation of genes (Wade, 2002; Hill et al., 2008). This process is illustrated in Figure 11 for a simple epistatic network consisting of two genes. Thus, as genetic complexity increases, the breeder needs practical ways to reduce this complexity to a manageable level that allows genetic progress. This study explored two particular practical approaches that have been adopted within commercial hybrid breeding, particularly in maize. With the availability of genomics and novel thinking about genetic complexity, we can now study the genetic implications of these practical approaches, many of which were devised and adopted prior to the availability of a theoretical and empirical framework to study their effects.

Two processes in particular accelerate such constraint, namely the creation of population bottlenecks and the subdivision of larger populations into more or less independent 'demes' (Katz and Young, 1975; Goodnight, 1995). Equivalent processes in the context of plant breeding programs are the degree of connectivity between breeding programs and the relative use of superior inbred lines in breeding crosses for producing the next generation, both 'themes' were explored in this study.

Decentrality theme

Classical quantitative genetics infinitesimal theory was used to design and optimize commercial hybrid breeding strategies, in combination with em-

Figure 10 Evolution of metrics over cycles in the inbred usage theme for scenarios with K of 1 (left column), 6 (middle) and 15 (right).

Figure 11 Illustration of a two-locus epistatic network in which neither the A nor the B locus exhibit any additive variation when all alleles are common (A) but collapses to a perfectly additive system in which the substitution effect of the A locus explains 100% of the variation when the B₂ allele becomes rare (B). Note that the substitution effect of the A locus would be reversed in sign when the B₁ allele became rare instead. Colours green, yellow and red represent high, intermediate and low phenotypic values, respectively.

772 pirical experience of what worked and what did not
773 (Hallauer et al., 2010). Yet, even though the infinitesimal
774 model implies optimality of a single, homogeneous
775 population, there were discussions about the
776 relative merits of large centralized vs. decentralized
777 breeding programs early on (Baker and Curnow,
778 1969). Later on, Podlich and Cooper (1999) explored
779 this problem on the basis of Sewall Wright's *shifting*
780 *balance theory*, (Wright, 1931, 1977; Wade and Good-
781 night, 1998). The shifting balance theory describes
782 an evolutionary process in which genetic drift re-
783 sulting from population subdivision enables ran-
784 dom movements across genetic space (i.e., against
785 selection gradients) and also converts epistatic to
786 additive genetic variation through constraining ge-
787 netic space as described above. This then en-
788 ables local adaptation in complex genetic landscapes
789 which is followed by differential migration from
790 higher to lower performing sub-populations and
791 thus 'spreading' of superior gene complexes across
792 the whole *meta population*.

793 While this theory remains controversial as a
794 model of natural evolution (Coyne et al., 1997), there
795 are remarkable similarities between meta popula-
796 tions in the context of the shifting balance theory
797 and the population structure of large commercial

798 breeding operations. The latter also do not oper-
799 ate as a centralized unit but rather as a decentral-
800 ized network of smaller programs with the most
801 successful germplasm being shared across (Cooper
802 et al., 2014). The same seems to be the case at the
803 industry level, with the major commercial breed-
804 ing operations being based on unique and distinct
805 germplasm backgrounds, with only occasional ex-
806 change of elite material, e.g., through ex-PVP lines
807 (Mikel and Dudley, 2006; White et al., 2020). As a
808 result of this decentralization, plant breeding pro-
809 grams are also characterized by having low effective
810 population sizes (Cowling, 2013), which makes
811 them more susceptible to genetic drift.

812 Here, we expanded on the work of Baker and
813 Curnow (1969) and Podlich and Cooper (1999) by
814 exploring breeding population structures with differ-
815 ing degrees of decentrality (Figure 4), ranging
816 from a large centralized program with high N_e to a
817 isolated set of smaller programs with low N_e , with a
818 series of scenarios with decentralized but connected
819 programs with N_e values in between these two ex-
820 tremes. We indeed found that strategies resulting
821 in low within program N_e through increased decen-
822 tralization and isolation became increasingly supe-
823 rior in terms of peak hybrid performance, as ge-
824 netic complexity K increased, while a centralized
825 strategy with high N_e was superior in less complex
826 landscapes. These results thus confirm the find-
827 ings of Podlich and Cooper (1999) that a decentral-
828 ized search strategy is superior in complex genetic
829 landscapes. Increasing isolation and decentraliza-
830 tion and the associated N_e reduction led to quicker
831 increases over cycles and higher overall values of
832 %GCA (Figure 8). At the highest levels of com-
833 plexity, only complete isolation generated amounts
834 of GCA variation sufficient for making genetic im-
835 provements cycle over cycle. The better ability to ex-
836 pose additivity in the form of GCA variation of the
837 more isolated and decentralized strategies was ex-
838 pected as per the discussion at the beginning of this
839 section outlining the relationship between amounts
840 of additive variation and constraint of genetic
841 space. This explains the clear advantages in terms
842 of genetic peak performance of the isolated strategy
843 at the highest levels of K .

844 The corollary of the constraint of genetic
845 space of course is a more rapid decline in genetic di-
846 versity and susceptibility to genetic drift, which ulti-
847 mately limits the selection potential of the programs.
848 Indeed, for values of K below eight, which marked
849 the switch from an uncorrelated to multi-peaked but

correlated genetic landscape (Figure 2), decentralized programs with increasing rates of germplasm exchange became superior. Accordingly, having a large centralized program became the optimal strategy at lower values of K . Here, the genetic landscape was simple enough to not require severe constraint of genetic space to expose sufficient amounts of GCA variation. The genetic drift experienced by small, isolated programs then unnecessarily led to the fixation of unfavourable alleles. This was most apparent at $K = 1$ where all variation is additive by definition and a decentralized strategy is not expected to have any advantage (Rathie and Nicholas, 1980). Here the isolated strategy led to fixation of almost all loci from cycle twenty onward and to a stalling of genetic gain significantly below the theoretically achievable maximum (Figure 8).

The establishment of genetically divergent heterotic groups has always been a central tenant of hybrid breeding (Melchinger and Gumber, 1998). Originally, optimal exploitation of heterosis was the main driver of their establishment (East, 1936). Later, however, maximization of GCA vs. SCA variation was identified as an important secondary feature of heterotic groups (Melchinger and Gumber, 1998). While this is well established for dominant gene action (Reif et al., 2007; Fischer et al., 2009), there are also indications for the conservation of favourable epistatic patterns that are disrupted when lines from different heterotic groups are recombined (Bernardo, 2001). Often, heterotic groups are established from populations that evolved in isolation for a long time. One of the best examples for this is the Dent by Flint heterotic pattern in maize which is prevalent in Central Europe and is comprised of populations that evolved in separation for centuries (Rebourg et al., 2003). Heterotic groups are thus a different and additional form of constraint of genetic space through historically grown genetic isolation. In our simulations, the different heterotic groups were originated from the same base population, yet we still observed a significant degree of genetic differentiation evolve over cycles (Figure 8), as expected in recurrent, reciprocal selection regimes (Labate et al., 1999; Longin et al., 2013). A portion of this differentiation can be attributed to genetic drift (Gerke et al., 2015), as evidenced by the non-zero genetic distance at $K = 1$, where all effects are additive and increasing genetic differentiation between heterotic groups would have no effect on the proportion of GCA variance. However, the genetic differentiation was considerably higher for

$K > 1$ (Figure 7), indicating that there indeed was a selection advantage to increased heterotic group divergence in complex genetic landscapes. This was particularly clear for the isolated scenario, where heterotic patterns could form uninterrupted within programs. For the centralized and decentralized strategies, the differentiation was maximal at lower values of K , because %GCA, and hence the effectiveness of recurrent, reciprocal selection, declined afterwards.

Inbred usage theme

The history of North American and European maize germplasm can be described as a succession of key inbreds that were heavily used in breeding crosses and had a distinct and lasting impact on germplasm (Mikel and Dudley, 2006; Technow et al., 2014; White et al., 2020). Those inbreds owe their success either to the outstanding general combining ability relative to their peers at the time, such as was case for the important North American line B73 (Mikel and Dudley, 2006) or their unique adaptation to specific climatic conditions, such as the European Flint lines F2 and F7 (Messmer et al., 1992; Böhm et al., 2014). The highly disproportionate importance of successful inbreds led to a significant reduction in genetic diversity (Rasmusson and Phillips, 1997; White et al., 2020), particularly relative to the source populations from which they were derived (Böhm et al., 2017). However, this constraint also might be responsible for the emergence of additive genetic variation from complex gene action through the so called founder or bottleneck effect (Goodnight, 1988; Cheverud and Routman, 1996; Naciri-Graven and Goudet, 2003; van Heerwaarden et al., 2008). We indeed observed that %GCA increased faster over cycles and reached higher values overall the more uneven the use of selected parents in breeding crosses (Figure 10), with the exception of $K = 1$, where all variance is additive by definition. At the highest degrees of landscape complexity only disproportionate use of inbreds, resulting in very low N_e , succeeded in generating amounts of %GCA sufficient for genetic improvements. Like in the decentrality theme, the higher values of %GCA of the disproportional use strategy translated into superior peak performances only at the values $K > 8$, i.e., after the landscape transitioned from multi-peaked but correlated to uncorrelated. Before that, balanced and particularly proportional use, both having higher N_e , achieved superior peak performances.

Maintenance of diversity

In our simulations, the constraintment of genetic space and reduction of N_e through decentralization and isolation or disproportionate use of inbred lines, while necessary for exposing additive genetic variation, led to a rapid fixation of alleles and a slowing of genetic gain in later cycles. This was partly a consequence of genetic drift caused by the low N_e (Cowling, 2013). However, the reduction of N_e was also caused in part by the effects of selection, particularly once the majority of the genetic variation was additive. This has not generally happened in commercial breeding programs, where genetic gain continues apace (Rasmusson and Phillips, 1997; Fischer et al., 2008; Duvick et al., 2004; Pfeiffer et al., 2019). Several factors that maintain diversity in practical programs were not included in the simulation model. For example, the simulation implicitly assumed that the environment and management conditions remained constant across all cycles, whereas both change more or less rapidly in reality. Changing selection environments imply changing selection targets and trajectories (Messina et al., 2011; Hammer et al., 2009), which reduce the pressure on particular alleles or allele complexes and thus slow or prevent fixation. Long-term selection experiments have shown that selection response can be maintained even in isolated and narrow populations (Dudley and Lambert, 2010; Durand et al., 2010, 2015). Several hypotheses were proposed for these surprising results, including epistasis (Carlborg et al., 2006), de novo genetic mutations, particularly when magnified through effects on epistatic complexes (Rasmusson and Phillips, 1997; Durand et al., 2010), creation of heritable epigenetic variation (Hauben et al., 2009), activity of transposable elements (Dubin et al., 2018), as well as the presence of 'cryptic genetic variation' through phenomena such as canalization (Gibson and Dworkin, 2004). Of these, only epistasis was present in our simulations. While highly speculative, these biological phenomena might explain the presence of abundant genetic variation and continued genetic gain in largely isolated and genetically narrow commercial plant breeding programs.

Applications of the NK for plant breeding

The NK model, originally developed by the theoretical biologist Stuart Kauffman to study evolution in complex genetic landscapes, has found applications

for modelling complex systems in disparate fields such as business administration (Csaszar, 2018), organizational learning theory (Lazer and Friedman, 2007), infrastructure design (Grove and Baumann, 2012) and physics (Qu et al., 2002). Following the example of (Podlich and Cooper, 1999), we here used the NK model to represent genetic complexity navigated by commercial hybrid breeding operations to study the effect of the degree of isolation between programs as well as the degree of imbalance in inbred usage, both key aspects of breeding strategies. We propose that this model can serve as an ideal starting point to study other aspects of hybrid breeding strategies. For example, Cooper and Podlich (2002) proposed an extension to the NK model that adds an environmental dimension and thus allows modelling concepts related to genotype by environment interaction (Cooper and DeLacy, 1994), yield stability (Piepho, 1998; Tollenaar and Lee, 2002), product placement (Messina et al., 2018) and the target population of environments (Comstock, 1977). These so called $E(NK)$ models represent different environments or management practices through a series of more or less similar genetic landscapes. This of course adds considerable complexity to the already complex static landscapes studied here and poses interesting dilemmas. For example, rapidly exposing additive variation, e.g., through isolation, might be even more important than in static landscapes because local optima have to be exploited quickly before they disappear once the environment shifts, for example through changes in management such as the historical increases in plant population for commercial maize production (Hammer et al., 2009). However, retaining allelic diversity, which hampers the exposing of additivity, is required to enable renewed adaptation to the changed environmental conditions.

A high degree of genetic complexity also implies a high degree of context dependency of genetic effects. Observe, for example, that the additive substitution effect of the 'A' locus in Figure 11 changes sign when the B_1 allele instead of the B_2 allele becomes rare. This has consequences on the persistence of accuracy of estimated QTL effects and genomic prediction models and can be addressed through iteratively updating training populations for genetic model parameterization (Podlich et al., 2004; Wolc et al., 2016; Forneris et al., 2017). The NK model framework can help address questions about the frequency with which this has to happen and whether data from previous generations

can be used. Recently, approaches were proposed that attempt to capture those context dependencies through biological models representing the interdependencies underlying the traits of interest (Technow et al., 2015). Such models are only approximations of the true biological complexity. However, Cooper et al. (2005), using the NK framework, have shown that even incomplete knowledge of biological networks can improve predictability of genetic effects and genetic gain. The context dependency of genetic effects, i.e., the effects being neither universally positive or negative (Wade, 2002), also has implications on innovative proposals for using CRISPR-Cas9 gene editing (Jaganathan et al., 2018; Gao et al., 2020) to either target recombination to create superior hypothetical linkage groups (Brandariz and Bernardo, 2019) or even the large scale “editing away” of deleterious mutations (Wallace et al., 2018). Finally, this framework might help devise strategies for the efficient introduction of exotic or ancient germplasm (Yu et al., 2016; Böhm et al., 2017), which evolved not just in a very different environmental, but also a different genetic context from the current elite breeding germplasm.

Back to the future

The structure of commercial plant breeding programs, particularly in major crops like maize, is characterized by a large degree of decentralization with exchange of successful germplasm within companies (Cooper et al., 2014), while isolation is the norm among companies (Mikel and Dudley, 2006). Plant breeders further have a tendency for relying on only a small set of elite inbred lines for producing the next generation (Rasmusson and Phillips, 1997), leading to a series of significant bottleneck events (White et al., 2020). All of these features lead to a drastically reduced effective population size and are not predicted to be promising strategies under the additive, infinitesimal model of quantitative genetics. Yet commercial hybrid breeding has delivered incredible amounts of genetic gain over the last century, and has thus contributed to food security and resource conservation (Duvick, 1999). Here we postulated that the described structure of plant breeding programs, with its constraint of genetic space, is in fact necessary for enabling the exploration and exploitation of genetic variation in complex genetic landscapes and that the success of a breeding program is not only determined by its germplasm per se, but by the structures that allow

it to evolve. The breeding program structures described here grew historically and we do not claim that it was set up with this intention. However, by doing “what worked”, breeders in preceding generations might have nonetheless been able to take advantage of the process described and postulated in this study. Understanding why these historic structures “worked” will be critical for designing breeding programs that can tackle the challenges of the century ahead.

References

- Altenberg, L. (1994). Evolving better representations through selective genome growth. In *Proceedings of the First IEEE Conference on Evolutionary Computation IEEE World Congress on Computational Intelligence ICEC-94*, pp. 182–187. IEEE.
- Araus, J. L. and J. E. Cairns (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* 19, 52–61.
- Baker, L. H. and R. N. Curnow (1969). Choice of Population Size and Use of Variation Between Replicate Populations in Plant Breeding Selection Programs 1. *Crop Sci* 9, 555–560.
- Barton, N. H., A. M. Etheridge, and A. Véber (2017). The infinitesimal model: Definition, derivation, and implications. *Theor Pop Biol* 118, 50–73.
- Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *J Stat Soft* 67(1), 1–48.
- Bernardo, R. (2001). breeding potential of intra- and interheterotic group crosses in maize. *Crop Sci* 41, 68–71.
- Bernardo, R. (2016). Bandwagons I, too, have known. *Theor Appl Genet* 129, 2323a–2332.
- Böhm, J., W. Schipprack, V. Mirdita, H. F. Utz, and A. E. Melchinger (2014). Breeding potential of european flint maize landraces evaluated by their testcross performance. *Crop Sci* 54, 1665–1672.
- Böhm, J., W. Schipprack, H. F. Utz, and A. E. Melchinger (2017). Tapping the genetic diversity of landraces in allogamous crops with doubled haploid lines: a case study from European flint maize. *Theor Appl Genet* 130, 861–873.
- Boichard, D., L. Maignel, and É. Verrier (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genet Sel Evol* 29, 5–23.
- Brandariz, S. P. and R. Bernardo (2019). Predicted genetic gains from targeted recombination in elite biparental maize populations. *Plant Genom* 12, 10.3835/plantgenome2018.08.0062.
- Carlborg, O. and C. S. Haley (2004). Epistasis: too often neglected in complex trait studies? *Nat Rev Genet* 5, 618–625.
- Carlborg, O., L. Jacobsson, P. Åhlgren, P. Siegel, and L. Andersson (2006). Epistasis and the release of genetic variation during long-term selection. *Nat Genet* 38, 418–420.

- 1152 Cheverud, J. M. and E. J. Routman (1996). Epistasis as a source of
1153 increased additive genetic variance at population bottlenecks.
1154 *Evolution* 50, 1042–1051. 1206
- 1155 Comstock, R. E. (1977). Quantitative genetics and the design of
1156 breeding programs. In *Proceedings of the International Confer-*
1157 *ence on Quantitative Genetics*, Ames, IA, pp. 16–21. Iowa State
1158 University Press. 1207
- 1159 Cooper, M. and I. H. DeLacy (1994). Relationships among
1160 analytical methods used to study genotypic variation and
1161 genotype-by-environment interaction in plant breeding multi-
1162 environment experiments. *Theor Appl Genet* 88, 561–572. 1208
- 1163 Cooper, M., C. D. Messina, D. Podlich, L. R. Totir, A. Baumgarten,
1164 N. J. Hausmann, D. Wright, and G. Graham (2014). Predicting
1165 the future of plant breeding: complementing empirical evalua-
1166 tion with genetic prediction. *Crop Pasture Sci* 65, 311. 1209
- 1167 Cooper, M. and D. W. Podlich (2002). The E(NK) model: Extend-
1168 ing the NK model to incorporate gene-by-environment inter-
1169 actions and epistasis for diploid genomes. *Complexity* 7, 31–47. 1210
- 1170 Cooper, M., D. W. Podlich, and O. S. Smith (2005). Gene-to-
1171 phenotype models and complex trait genetics. *Aust J Agric*
1172 *Res* 56, 895–918. 1211
- 1173 Corbin, L. J., A. Y. H. Liu, S. C. Bishop, and J. A. Woolliams
1174 (2012). Estimation of historical effective population size using
1175 linkage disequilibria with marker data. *J Anim Breed Genet* 129,
1176 257–270. 1212
- 1177 Cowling, W. A. (2013). Sustainable plant breeding. *Plant*
1178 *Breed* 132, 1–9. 1213
- 1179 Coyne, J. A., N. H. Barton, and M. Turelli (1997). Perspective:
1180 a critique of Sewall Wright’s shifting balance theory of evolu-
1181 tion. *Evolution* 51, 643–671. 1214
- 1182 Csaszar, F. A. (2018). A note on how NK landscapes work. *J Org*
1183 *Design* 7, 10.1186/s41469-018-0039-0. 1215
- 1184 Dubin, M. J., O. Mittelsten Scheid, and C. Becker (2018). Trans-
1185 posons: a blessing curse. *Curr Opin Plant Biol* 42, 23–29. 1216
- 1186 Dudley, J. W. and R. J. Lambert (2010). 100 Generations of Selec-
1187 tion for Oil and Protein in Corn. In *Plant Breeding Reviews*, pp.
1188 79–110. John Wiley & Sons, Ltd. 1217
- 1189 Durand, E., M. I. Tenaillon, X. Raffoux, S. ThÃl’pot, M. Falque,
1190 P. Jamin, A. Bourgaïs, A. Ressayre, and C. Dillmann (2015).
1191 Dearth of polymorphism associated with a sustained response
1192 to selection for flowering time in maize. *BMC Evol Biol* 15,
1193 10.1186/s12862-015-0382-5. 1218
- 1194 Durand, E., M. I. Tenaillon, C. Ridel, D. Coubriche, P. Jamin,
1195 S. Jouanne, A. Ressayre, A. Charcosset, and C. Dillmann
1196 (2010). Standing variation and new mutations both contribute
1197 to a fast response to selection for flowering time in maize in-
1198 bred. *BMC Evol Biol* 10, 10.1186/1471-2148-10-2. 1219
- 1199 Duvick, D. (1999). Heterosis: feeding people and protecting natu-
1200 ral resources. In J. Coors and S. Pandey (Eds.), *The genetics and*
1201 *exploitation of heterosis in crops*, pp. 19–29. Madison, WI: CSSA. 1220
- 1202 Duvick, D., J. Smith, and M. Cooper (2004). Long-term selection
1203 in a commercial hybrid maize breeding program. In J. Janick
1204 (Ed.), *Plant Breeding Reviews*, pp. 109–152. Hoboken, NJ: John
1205 Wiley & Sons, Inc. 1221
- Dwivedi, S. L., A. B. Britt, L. Tripathi, S. Sharma, H. D. Upad-
hyaya, and R. Ortiz (2015). Haploids: constraints and oppor-
tunities in plant breeding. *Biotechnol Adv* 33, 812–829. 1222
- East, E. M. (1936). Heterosis. *Genetics* 21, 375–397. 1223
- Falconer, D. S. and T. F. C. Mackay (1996). *Introduction to Quanti-*
tative Genetics (4 ed.). London: Pearson. 1224
- Fiévet, J. B., C. Dillmann, and D. de Vienne (2010). Systemic prop-
erties of metabolic networks lead to an epistasis-based model
for heterosis. *Theor Appl Genet* 120, 463–473. 1225
- Fiévet, J. B., T. Nidelet, C. Dillmann, and D. de Vienne
(2018). Heterosis is a systemic property emerging from
non-linear genotype-phenotype relationships: evidence from
in vitro genetics and computer simulations. *Front Genet* 9,
10.3389/fgene.2018.00159. 1226
- Fischer, S., J. Möhring, H. P. Maurer, H.-P. Piepho, E.-M. Thiemt,
C. C. Schön, A. E. Melchinger, and J. C. Reif (2009). Impact
of genetic divergence on the ratio of variance due to specific
vs. general combining ability in winter triticales. *Crop Sci* 49,
2119–2122. 1227
- Fischer, S., J. Möhring, C. C. Schön, H.-P. Piepho, D. Klein,
W. Schipprack, H. F. Utz, A. E. Melchinger, and J. C. Reif
(2008). Trends in genetic variance components during 30 years
of hybrid maize breeding at the University of Hohenheim.
Plant Breed 127, 446–451. 1228
- Forneris, N. S., Z. G. Vitezica, A. Legarra, and M. Pérez-Enciso
(2017). Influence of epistasis on response to genomic se-
lection using complete sequence data. *Genet Sel Evol* 49,
10.1186/s12711-017-0340-3. 1229
- Gao, H., M. J. Gadlage, H. R. Lafitte, B. Lenderts, M. Yang,
M. Schroder, J. Farrell, K. Snopek, D. Peterson, L. Feigenbutz,
S. Jones, G. S. Clair, M. Rahe, N. Sanyour-Doyel, C. Peng,
L. Wang, J. K. Young, M. Beatty, B. Dahlke, J. Hazebroek,
T. W. Greene, A. M. Cigan, N. D. Chilcoat, and R. B. Meeley
(2020). Superior field performance of waxy corn engineered
using CRISPR-Cas9. *Nat Biotechnol* 38, 10.1038/s41587-020-
0444-0. 1230
- Gaynor, R. C., G. Gorjanc, A. R. Bentley, E. S. Ober, P. Howell,
R. Jackson, I. J. Mackay, and J. M. Hickey (2017). A two-part
strategy for using genomic selection to develop inbred lines.
Crop Sci 57, 2372–2386. 1231
- Gerke, J. P., J. W. Edwards, K. E. Guill, J. Ross-Ibarra, and M. D.
McMullen (2015). The genomic impacts of drift and selection
for hybrid performance in maize. *Genetics* 201, 1201–1211. 1232
- Gibson, G. and I. Dworkin (2004). Uncovering cryptic genetic
variation. *Nat Rev Genet* 5, 681–690. 1233
- Goodnight, C. J. (1988). Epistasis and the effect of founder events
on the additive genetic variance. *Evolution* 42, 441–454. 1234
- Goodnight, C. J. (1995). Epistasis and the increase in additive
genetic variance: implications for phase 1 of Wright’s shifting-
balance process. *Evolution* 49, 502–511. 1235
- Grove, N. and O. Baumann (2012). Complexity in the telecommu-
nications industry: When integrating infrastructure and ser-
vices backfires. *Telecomm Policy* 36, 40–50. 1236

- 1259 Hallauer, A. R., M. J. Carena, and J. B. M. Filho (2010). *Quantitative Genetics in Maize Breeding* (3 ed.). Handbook of Plant
1260 Breeding. New York: Springer-Verlag. 1312
- 1261 1313
- 1262 Hammer, G., M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. van
1263 Eeuwijk, S. Chapman, and D. Podlich (2006). Models for
1264 navigating biological complexity in breeding improved crop
1265 plants. *Trends Plant Sci* 11, 587–593. 1314
- 1266 1315
- 1267 1316
- 1268 1317
- 1269 1318
- 1270 1319
- 1271 1320
- 1272 1321
- 1273 1322
- 1274 1323
- 1275 1324
- 1276 1325
- 1277 1326
- 1278 1327
- 1279 1328
- 1280 1329
- 1281 1330
- 1282 1331
- 1283 1332
- 1284 1333
- 1285 1334
- 1286 1335
- 1287 1336
- 1288 1337
- 1289 1338
- 1290 1339
- 1291 1340
- 1292 1341
- 1293 1342
- 1294 1343
- 1295 1344
- 1296 1345
- 1297 1346
- 1298 1347
- 1299 1348
- 1300 1349
- 1301 1350
- 1302 1351
- 1303 1352
- 1304 1353
- 1305 1354
- 1306 1355
- 1307 1356
- 1308 1357
- 1309 1358
- 1310 1359
- 1311 1360
- 1312 1361
- 1313 1362
- 1314 1363
- 1315 1364
- 1316 1365
- 1317 1366
- 1318 1367
- 1319 1368
- 1320 1369
- 1321 1370
- 1322 1371
- 1323 1372
- 1324 1373
- 1325 1374
- 1326 1375
- 1327 1376
- 1328 1377
- 1329 1378
- 1330 1379
- 1331 1380
- 1332 1381
- 1333 1382
- 1334 1383
- 1335 1384
- 1336 1385
- 1337 1386
- 1338 1387
- 1339 1388
- 1340 1389
- 1341 1390
- 1342 1391
- 1343 1392
- 1344 1393
- 1345 1394
- 1346 1395
- 1347 1396
- 1348 1397
- 1349 1398
- 1350 1399
- 1351 1400
- 1352 1401
- 1353 1402
- 1354 1403
- 1355 1404
- 1356 1405
- 1357 1406
- 1358 1407
- 1359 1408
- 1360 1409
- 1361 1410
- 1362 1411
- 1363 1412
- 1364 1413
- 1365 1414
- 1366 1415
- 1367 1416
- 1368 1417
- 1369 1418
- 1370 1419
- 1371 1420
- 1372 1421
- 1373 1422
- 1374 1423
- 1375 1424
- 1376 1425
- 1377 1426
- 1378 1427
- 1379 1428
- 1380 1429
- 1381 1430
- 1382 1431
- 1383 1432
- 1384 1433
- 1385 1434
- 1386 1435
- 1387 1436
- 1388 1437
- 1389 1438
- 1390 1439
- 1391 1440
- 1392 1441
- 1393 1442
- 1394 1443
- 1395 1444
- 1396 1445
- 1397 1446
- 1398 1447
- 1399 1448
- 1400 1449
- 1401 1450
- 1402 1451
- 1403 1452
- 1404 1453
- 1405 1454
- 1406 1455
- 1407 1456
- 1408 1457
- 1409 1458
- 1410 1459
- 1411 1460
- 1412 1461
- 1413 1462
- 1414 1463
- 1415 1464
- 1416 1465
- 1417 1466
- 1418 1467
- 1419 1468
- 1420 1469
- 1421 1470
- 1422 1471
- 1423 1472
- 1424 1473
- 1425 1474
- 1426 1475
- 1427 1476
- 1428 1477
- 1429 1478
- 1430 1479
- 1431 1480
- 1432 1481
- 1433 1482
- 1434 1483
- 1435 1484
- 1436 1485
- 1437 1486
- 1438 1487
- 1439 1488
- 1440 1489
- 1441 1490
- 1442 1491
- 1443 1492
- 1444 1493
- 1445 1494
- 1446 1495
- 1447 1496
- 1448 1497
- 1449 1498
- 1450 1499
- 1451 1500
- 1452 1501
- 1453 1502
- 1454 1503
- 1455 1504
- 1456 1505
- 1457 1506
- 1458 1507
- 1459 1508
- 1460 1509
- 1461 1510
- 1462 1511
- 1463 1512
- 1464 1513
- 1465 1514
- 1466 1515
- 1467 1516
- 1468 1517
- 1469 1518
- 1470 1519
- 1471 1520
- 1472 1521
- 1473 1522
- 1474 1523
- 1475 1524
- 1476 1525
- 1477 1526
- 1478 1527
- 1479 1528
- 1480 1529
- 1481 1530
- 1482 1531
- 1483 1532
- 1484 1533
- 1485 1534
- 1486 1535
- 1487 1536
- 1488 1537
- 1489 1538
- 1490 1539
- 1491 1540
- 1492 1541
- 1493 1542
- 1494 1543
- 1495 1544
- 1496 1545
- 1497 1546
- 1498 1547
- 1499 1548
- 1500 1549
- 1501 1550
- 1502 1551
- 1503 1552
- 1504 1553
- 1505 1554
- 1506 1555
- 1507 1556
- 1508 1557
- 1509 1558
- 1510 1559
- 1511 1560
- 1512 1561
- 1513 1562
- 1514 1563
- 1515 1564
- 1516 1565
- 1517 1566
- 1518 1567
- 1519 1568
- 1520 1569
- 1521 1570
- 1522 1571
- 1523 1572
- 1524 1573
- 1525 1574
- 1526 1575
- 1527 1576
- 1528 1577
- 1529 1578
- 1530 1579
- 1531 1580
- 1532 1581
- 1533 1582
- 1534 1583
- 1535 1584
- 1536 1585
- 1537 1586
- 1538 1587
- 1539 1588
- 1540 1589
- 1541 1590
- 1542 1591
- 1543 1592
- 1544 1593
- 1545 1594
- 1546 1595
- 1547 1596
- 1548 1597
- 1549 1598
- 1550 1599
- 1551 1600
- 1552 1601
- 1553 1602
- 1554 1603
- 1555 1604
- 1556 1605
- 1557 1606
- 1558 1607
- 1559 1608
- 1560 1609
- 1561 1610
- 1562 1611
- 1563 1612
- 1564 1613
- 1565 1614
- 1566 1615
- 1567 1616
- 1568 1617
- 1569 1618
- 1570 1619
- 1571 1620
- 1572 1621
- 1573 1622
- 1574 1623
- 1575 1624
- 1576 1625
- 1577 1626
- 1578 1627
- 1579 1628
- 1580 1629
- 1581 1630
- 1582 1631
- 1583 1632
- 1584 1633
- 1585 1634
- 1586 1635
- 1587 1636
- 1588 1637
- 1589 1638
- 1590 1639
- 1591 1640
- 1592 1641
- 1593 1642
- 1594 1643
- 1595 1644
- 1596 1645
- 1597 1646
- 1598 1647
- 1599 1648
- 1600 1649
- 1601 1650
- 1602 1651
- 1603 1652
- 1604 1653
- 1605 1654
- 1606 1655
- 1607 1656
- 1608 1657
- 1609 1658
- 1610 1659
- 1611 1660
- 1612 1661
- 1613 1662
- 1614 1663
- 1615 1664
- 1616 1665
- 1617 1666
- 1618 1667
- 1619 1668
- 1620 1669
- 1621 1670
- 1622 1671
- 1623 1672
- 1624 1673
- 1625 1674
- 1626 1675
- 1627 1676
- 1628 1677
- 1629 1678
- 1630 1679
- 1631 1680
- 1632 1681
- 1633 1682
- 1634 1683
- 1635 1684
- 1636 1685
- 1637 1686
- 1638 1687
- 1639 1688
- 1640 1689
- 1641 1690
- 1642 1691
- 1643 1692
- 1644 1693
- 1645 1694
- 1646 1695
- 1647 1696
- 1648 1697
- 1649 1698
- 1650 1699
- 1651 1700
- 1652 1701
- 1653 1702
- 1654 1703
- 1655 1704
- 1656 1705
- 1657 1706
- 1658 1707
- 1659 1708
- 1660 1709
- 1661 1710
- 1662 1711
- 1663 1712
- 1664 1713
- 1665 1714
- 1666 1715
- 1667 1716
- 1668 1717
- 1669 1718
- 1670 1719
- 1671 1720
- 1672 1721
- 1673 1722
- 1674 1723
- 1675 1724
- 1676 1725
- 1677 1726
- 1678 1727
- 1679 1728
- 1680 1729
- 1681 1730
- 1682 1731
- 1683 1732
- 1684 1733
- 1685 1734
- 1686 1735
- 1687 1736
- 1688 1737
- 1689 1738
- 1690 1739
- 1691 1740
- 1692 1741
- 1693 1742
- 1694 1743
- 1695 1744
- 1696 1745
- 1697 1746
- 1698 1747
- 1699 1748
- 1700 1749
- 1701 1750
- 1702 1751
- 1703 1752
- 1704 1753
- 1705 1754
- 1706 1755
- 1707 1756
- 1708 1757
- 1709 1758
- 1710 1759
- 1711 1760
- 1712 1761
- 1713 1762
- 1714 1763
- 1715 1764
- 1716 1765
- 1717 1766
- 1718 1767
- 1719 1768
- 1720 1769
- 1721 1770
- 1722 1771
- 1723 1772
- 1724 1773
- 1725 1774
- 1726 1775
- 1727 1776
- 1728 1777
- 1729 1778
- 1730 1779
- 1731 1780
- 1732 1781
- 1733 1782
- 1734 1783
- 1735 1784
- 1736 1785
- 1737 1786
- 1738 1787
- 1739 1788
- 1740 1789
- 1741 1790
- 1742 1791
- 1743 1792
- 1744 1793
- 1745 1794
- 1746 1795
- 1747 1796
- 1748 1797
- 1749 1798
- 1750 1799
- 1751 1800
- 1752 1801
- 1753 1802
- 1754 1803
- 1755 1804
- 1756 1805
- 1757 1806
- 1758 1807
- 1759 1808
- 1760 1809
- 1761 1810
- 1762 1811
- 1763 1812
- 1764 1813
- 1765 1814
- 1766 1815
- 1767 1816
- 1768 1817
- 1769 1818
- 1770 1819
- 1771 1820
- 1772 1821
- 1773 1822
- 1774 1823
- 1775 1824
- 1776 1825
- 1777 1826
- 1778 1827
- 1779 1828
- 1780 1829
- 1781 1830
- 1782 1831
- 1783 1832
- 1784 1833
- 1785 1834
- 1786 1835
- 1787 1836
- 1788 1837
- 1789 1838
- 1790 1839
- 1791 1840
- 1792 1841
- 1793 1842
- 1794 1843
- 1795 1844
- 1796 1845
- 1797 1846
- 1798 1847
- 1799 1848
- 1800 1849
- 1801 1850
- 1802 1851
- 1803 1852
- 1804 1853
- 1805 1854
- 1806 1855
- 1807 1856
- 1808 1857
- 1809 1858
- 1810 1859
- 1811 1860
- 1812 1861
- 1813 1862
- 1814 1863
- 1815 1864
- 1816 1865
- 1817 1866
- 1818 1867
- 1819 1868
- 1820 1869
- 1821 1870
- 1822 1871
- 1823 1872
- 1824 1873
- 1825 1874
- 1826 1875
- 1827 1876
- 1828 1877
- 1829 1878
- 1830 1879
- 1831 1880
- 1832 1881
- 1833 1882
- 1834 1883
- 1835 1884
- 1836 1885
- 1837 1886
- 1838 1887
- 1839 1888
- 1840 1889
- 1841 1890
- 1842 1891
- 1843 1892
- 1844 1893
- 1845 1894
- 1846 1895
- 1847 1896
- 1848 1897
- 1849 1898
- 1850 1899
- 1851 1900
- 1852 1901
- 1853 1902
- 1854 1903
- 1855 1904
- 1856 1905
- 1857 1906
- 1858 1907
- 1859 1908
- 1860 1909
- 1861 1910
- 1862 1911
- 1863 1912
- 1864 1913
- 1865 1914
- 1866 1915
- 1867 1916
- 1868 1917
- 1869 1918
- 1870 1919
- 1871 1920
- 1872 1921
- 1873 1922
- 1874 1923
- 1875 1924
- 1876 1925
- 1877 1926
- 1878 1927
- 1879 1928
- 1880 1929
- 1881 1930
- 1882 1931
- 1883 1932
- 1884 1933
- 1885 1934
- 1886 1935
- 1887 1936
- 1888 1937
- 1889 1938
- 1890 1939
- 1891 1940
- 1892 1941
- 1893 1942
- 1894 1943
- 1895 1944
- 1896 1945
- 1897 1946
- 1898 1947
- 1899 1948
- 1900 1949
- 1901 1950
- 1902 1951
- 1903 1952
- 1904 1953
- 1905 1954
- 1906 1955
- 1907 1956
- 1908 1957
- 1909 1958
- 1910 1959
- 1911 1960
- 1912 1961
- 1913 1962
- 1914 1963
- 1915 1964
- 1916 1965
- 1917 1966
- 1918 1967
- 1919 1968
- 1920 1969
- 1921 1970
- 1922 1971
- 1923 1972
- 1924 1973
- 1925 1974
- 1926 1975
- 1927 1976
- 1928 1977
- 1929 1978
- 1930 1979
- 1931 1980
- 1932 1981
- 1933 1982
- 1934 1983
- 1935 1984
- 1936 1985
- 1937 1986
- 1938 1987
- 1939 1988
- 1940 1989
- 1941 1990
- 1942 1991
- 1943 1992
- 1944 1993
- 1945 1994
- 1946 1995
- 1947 1996
- 1948 1997
- 1949 1998
- 1950 1999
- 1951 2000
- 1952 2001
- 1953 2002
- 1954 2003
- 1955 2004
- 1956 2005
- 1957 2006
- 1958 2007
- 1959 2008
- 1960 2009
- 1961 2010
- 1962 2011
- 1963 2012
- 1964 2013
- 1965 2014
- 1966 2015
- 1967 2016
- 1968 2017
- 1969 2018
- 1970 2019
- 1971 2020
- 1972 2021
- 1973 2022
- 1974 2023
- 1975 2024
- 1976 2025
- 1977 2026
- 1978 2027
- 1979 2028
- 1980 2029
- 1981 2030
- 1982 2031
- 1983 2032
- 1984 2033
- 1985 2034
- 1986 2035
- 1987 2036
- 1988 2037
- 1989 2038
- 1990 2039
- 1991 2040
- 1992 2041
- 1993 2042
- 1994 2043
- 1995 2044
- 1996 2045
- 1997 2046
- 1998 2047
- 1999 2048
- 2000 2049
- 2001 2050
- 2002 2051
- 2003 2052
- 2004 2053
- 2005 2054
- 2006 2055
- 2007 2056
- 2008 2057
- 2009 2058
- 2010 2059
- 2011 2060
- 2012 2061
- 2013 2062
- 2014 2063
- 2015 2064
- 2016 2065
- 2017 2066
- 2018 2067
- 2019 2068
- 2020 2069
- 2021 2070
- 2022 2071
- 2023 2072
- 2024 2073
- 2025 2074
- 2026 2075
- 2027 2076
- 2028 2077
- 2029 2078
- 2030 2079
- 2031 2080
- 2032 2081
- 2033 2082
- 2034 2083
- 2035 2084
- 2036 2085
- 2037 2086
- 2038 2087
- 2039 2088
- 2040 2089
- 2041 2090
- 2042 2091
- 2043 2092
- 2044 2093
- 2045 2094
- 2046 2095
- 2047 2096
- 2048 2097
- 2049 2098
- 2050 2099
- 2051 2100
- 2052 2101
- 2053 2102
- 2054 2103
- 2055 2104
- 2056 2105
- 2057 2106
- 2058 2107
- 2059 2108
- 2060 2109
- 2061 2110
- 2062 2111
- 2063 2112
- 2064 2113
- 2065 2114
- 2066 2115
- 2067 2116
- 2068 2117
- 2069 2118
- 2070 2119
- 2071 2120
- 2072 2121
- 2073 2122
- 2074 2123
- 2075 2124
- 2076 2125
- 2077 2126
- 2078 2127
- 2079 2128
- 2080 2129
- 2081 2130
- 2082 2131
- 2083 2132
- 2084 2133
- 2085 2134
- 2086 2135
- 2087 2136
- 2088 2137
- 2089 2138
- 2090 2139
- 2091 2140
- 2092 2141
- 2093 2142
- 2094 2143
- 2095 2144
- 2096 2145
- 2097 2146
- 2098 2147
- 2099 2148
- 2100 2149
- 2101 2150
- 2102 2151
- 2103 2152
- 2104 2153
- 2105 2154
- 2106 2155
- 2107 2156
- 2108 2157
- 2109 2158
- 2110 2159
- 2111 2160
- 2112 2161
- 211

- Pinheiro, H. P., A. de Souza Pinheiro, and P. K. Sen (2005). Comparison of genomic sequences using the Hamming distance. *J Stat Plan Inference* 130, 325–339.
- Podlich, D. and M. Cooper (1999). Modelling plant breeding programs as search strategies on a complex response surface. *Lecture Notes in Computer Science* 1585, 171–178.
- Podlich, D. W., C. R. Winkler, and M. Cooper (2004). Mapping as you go. *Crop Sci* 44, 1560–1571.
- Poland, J. A. and T. W. Rife (2012). Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* 5, 92–102.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge: Cambridge University Press.
- Qu, X., M. Aldana, and L. P. Kadanoff (2002). Numerical and theoretical studies of noise effects in the Kauffman model. *J Stat Phys* 109, 967–986.
- R Core Team (2018). *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rasmusson, D. C. and R. L. Phillips (1997). Plant breeding progress and genetic diversity from de novo variation and elevated epistasis. *Crop Sci* 37, 303–310.
- Rathie, K. A. and F. W. Nicholas (1980). Artificial selection with differing population structures. *Genet Res* 36, 117–131.
- Rebourg, C., M. Chastanet, B. Gouesnard, C. Welcker, P. Dubreuil, and A. Charcosset (2003). Maize introduction into Europe: the history reviewed in the light of molecular data. *Theor Appl Genet* 106, 895–903.
- Reif, J. C., F.-M. Gumpert, S. Fischer, and A. E. Melchinger (2007). Impact of interpopulation divergence on additive and dominance variance in hybrid populations. *Genetics* 176, 1931–1934.
- Reif, J. C., A. E. Melchinger, and M. Frisch (2005). Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. *Crop Sci* 45, 1–7.
- Saha, R., P. F. Suthers, and C. D. Maranas (2011). Zea mays iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. *PLoS ONE* 6, 10.1371/journal.pone.0021784.
- Seye, A. I., C. Bauland, A. Charcosset, and L. Moreau (2020). 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.
- Shull, G. H. (1908). The composition of a field of maize. *J Hered* 4, 296–301.
- Silva Dias, J. C. (2010). Impact of improved vegetable cultivars in overcoming food insecurity. *Euphytica* 176, 125–136.
- Sprague, G. F. and L. A. Tatum (1942). General vs. specific combining ability in single crosses of corn. *Agron J* 34, 923–932.
- Technow, F. (2013). *hybred: Simulation of genomic data in applied genetics*. R package. version 0.4.
- Technow, F. (2019). Use of f2 bulks in training sets for genomic prediction of combining ability and hybrid performance. *G3* 9, 1557–1569.
- Technow, F. and J. P. Gerke (2017). Parent-progeny imputation from pooled samples for cost-efficient genotyping in plant breeding. *PLOS ONE* 12, e0190271.
- Technow, F., C. D. Messina, L. R. Totir, and M. Cooper (2015). Integrating crop growth models with whole genome prediction through approximate bayesian computation. *PLOS ONE* 10, e0130855.
- Technow, F., T. A. Schrag, W. Schipprack, E. Bauer, H. Simianer, and A. E. Melchinger (2014). Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. *Genetics* 197, 1343–1355.
- Technow, F., T. A. Schrag, W. Schipprack, and A. E. Melchinger (2014). Identification of key ancestors of modern germplasm in a breeding program of maize. *Theor Appl Genet* 127, 2545–2553.
- Tollenaar, M. and E. A. Lee (2002). Yield potential, yield stability and stress tolerance in maize. *Field Crop Res* 75, 161–169.
- Vacher, M. and I. Small (2019). Simulation of heterosis in a genome-scale metabolic network provides mechanistic explanations for increased biomass production rates in hybrid plants. *Syst Biol Appl* 5(1), 10.1038/s41540-019-0101-8.
- van Heerwaarden, B., Y. Willi, T. N. Kristensen, and A. A. Hoffmann (2008). Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest drosophila. *Genetics* 179, 2135–2146.
- Voss-Fels, K. P., M. Cooper, and B. J. Hayes (2019). Accelerating crop genetic gains with genomic selection. *Theor Appl Genet* 132, 669–686.
- Wade, M. J. (2002). A gene’s eye view of epistasis, selection and speciation. *J Evol Biol* 15, 337–346.
- Wade, M. J. and C. J. Goodnight (1998). The theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. *Evolution* 52, 1537–1553.
- Wallace, J. G., E. Rodgers-Melnick, and E. S. Buckler (2018). On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu Rev Genet* 52, 421–444.
- White, M. R., M. A. Mikel, N. d. Leon, and S. M. Kaeppler (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. *Crop Sci* 60, 100–114.
- Wilkins, O., C. Hafemeister, A. Plessis, M.-M. Holloway-Phillips, G. M. Pham, A. B. Nicotra, G. B. Gregorio, S. K. Jagadish, E. M. Septiningsih, R. Bonneau, and M. Purugganan (2016). EGRINs (environmental gene regulatory influence networks) in rice that function in the response to water deficit, high temperature, and agricultural environments. *Plant Cell* 28, 2365–2384.
- Wolc, A., A. Kranis, J. Arango, P. Settar, J. E. Fulton, N. P. O’Sullivan, A. Avendano, K. A. Watson, J. M. Hickey, G. de los Campos, R. L. Fernando, D. J. Garrick, and J. C. M. Dekkers (2016). Implementation of genomic selection in the poultry industry. *Anim Fron* 6, 23–31.

- 1472 Wright, S. (1931). Evolution in Mendelian populations. *Genet-*
1473 *ics* 16, 97–159.
- 1474 Wright, S. (1932). The roles of mutation, inbreeding, crossbreed-
1475 ing, and selection in evolution. In *Proceedings of the Sixth Inter-*
1476 *national Congress of Genetics*, Ithaca, NY, pp. 356–366. Brooklyn
1477 Botanic Garden.
- 1478 Wright, S. (1977). *Evolution and the genetics of populations*. Chicago:
1479 University of Chicago Press.
- 1480 Yu, X., X. Li, T. Guo, C. Zhu, Y. Wu, S. E. Mitchell, K. L. Rooze-
1481 boom, D. Wang, M. L. Wang, G. A. Pederson, T. T. Tesso,
1482 P. S. Schnable, R. Bernardo, and J. Yu (2016). Genomic predic-
1483 tion contributing to a promising global strategy to turbocharge
1484 gene banks. *Nat Plant* 2, 10.1038/NPLANTS.2016.150.