Accelerating agriculture: Data-intensive plant breeding and the use of genetic gain as an indicator for agricultural research and development

Hugh F. Williamson [*], Sabina Leonelli

Exeter Centre for the Study of the Life Sciences, Department of Sociology, Philosophy and Anthropology, University of Exeter, Byrne House, St. German's Road, Exeter EX4 4PJ, UK

A B S T R A C T

Accelerating the rate of genetic gain has in recent years become a key objective in plant breeding for the Global South, building on the availability of new data technologies and bridging biological interest in crop improvement with economic interest in enhancing the cost efficiency of breeding programs. This paper explains the concept of genetic gain, the conditions for its emerging status as an indicator of agricultural development and the broader implications of this move, with particular emphasis on the changing knowledge-control regimes of plant breeding, the social and political consequences for smallholder farmers and climate-adaptive agriculture. We analyse how prioritising the variables used to derive the indicator when deciding on agricultural policies affects the relationship between development goals and practice. We conclude that genetic gain should not be considered as a primary indicator of agricultural development in the absence of information on other key areas (including agrobiodiversity, seed systems and the differential impact of climate change on soil, crops and communities), as well as tools to evaluate the pros and cons of the acceleration in seed selection, management and evaluation fostered by the adoption of genetic gain as a key indicator.

1. Introduction

The genomics revolution in the biosciences has led to substantive changes in the organisation of how biological research is done and how biotechnologies are produced, at practical, social and epistemic as well as technological levels (Hilgartner, 2017; Richardson & Stevens 2015). Plant science and its applied fields are no exception, and much attention has been given to the development and potential impacts of transgenic technologies and genome editing techniques such as CRISPR for food and agriculture (e.g. Helliwell et al., 2019; Montenegro de Wit, 2020). "New breeding techniques", as these technologies are frequently termed, are often analysed in terms of their novelty, disruptive potential and risk. Yet, as noted by plant scientist Caixia Gao, beyond the headlines much of the advantage offered by CRISPR and similar technology is simply the ability to produce "identical results to conventional [breeding] methods in a much more predictable, faster and even cheaper manner" (Gao, 2019, p. 275). Indeed, major debates continue to interrogate whether gene editing produces outcomes that differ from conventional breeding methods, such as chemical-induced mutagenesis, and what this may mean for regulation (Pirscher et al., 2018). At a moment when the role of genomics in plant breeding is still being defined, speed and efficiency in

the identification and production of valuable varieties are increasingly prioritised as both practical and policy objectives. The use of gene editing technologies to this end is only the tip of the iceberg, and in fact it is a strategy that poses many barriers. For international agricultural research and breeding networks focused on the Global South, limited resources present a significant obstacle to the implementation of cutting-edge technologies such as CRISPR at scale; similar concerns affect areas of the Global North which are not able to reproduce the conditions required of intensively managed crops or are peripheral in relation to large scale processing infrastructure. Less visibly, but with a broader impact, agricultural research networks are being reorganised in ways that combine older statistical and more recent data-intensive breeding methods, with wide implications for scientific research, breeding practice and agricultural systems.

In this paper, we discuss the significance of an indicator that is rapidly becoming central to the shape and direction of such reorganisations: the rate of genetic gain. Genetic gain is a statistical measure of the genetic improvement of breeding populations that was first introduced in the context of animal breeding in the early twentieth century. Genetic gain has acquired new relevance for plant breeding in light of a resurgence of interest in the possibilities afforded by quantitative genetics alongside

increasingly cheap and easy methods for genomic data collection, sharing and analysis. It is being championed as a key performance indicator for plant breeding by breeders, researchers and funders around the world, alongside an active commitment to "accelerate" rates of genetic gain as a key policy objective.

In plant breeding for the public domain and the Global South, this commitment is being spearheaded by the Consultative Group for International Agricultural Research (CGIAR), arguably the most influential and extensive research network for agricultural research in the world (Curry & Lorek, under review), and the Bill and Melinda Gates Foundation (BMGF), currently a major funder of the CGIAR and of international research focused on development more broadly (McGoey, 2016). Similar objectives are also well-established in advanced commercial plant breeding programmes (e.g. Byrum et al., 2017; Eathington et al., 2007). We leave the latter aside in this paper however, in order to focus on the specific changes to public plant breeding in the international arena, where resources are limited and the structure of objectives is explicitly oriented towards a wider range of development goals beyond commercial growth.

Through an analysis of relevant scientific literature and international initiatives fostering data-intensive agriculture, we identify a set of proposals for the use of genetic gain as a performance indicator, which prioritise the achievement of greater genetic gain in seed systems and agriculture as a necessary step towards food security and sustainable agricultural development. We highlight the challenges that this set of proposals holds for climate-adaptive agriculture and the sustainable management of seed systems. In conclusion, we point out that genetic gain, despite its usefulness as one of the main indicators for agronomic performance, is not sufficient to assess the pace and progress of plant breeding for agricultural development goals. Rather, the tendency to prioritise and incentivise speed of breeding needs to be complemented by incentives to ensure that breeding outputs are adaptive to diverse agroecological systems and breeding programmes recognise the multiple values of agricultural biodiversity.

To this aim, the paper is structured as follows. Section two begins the paper with an overview of the concept of genetic gain and its status as an indicator. In section 2.2 we outline the reasons why genetic gain has become important in contemporary plant breeding and the wider field of discourse in which its value has been located. In the third section, we outline the evolving context of international plant breeding within which genetic gain is utilised, focussing on a range of modernisations aimed to make plant breeding more data-driven, and a related set of changes in the knowledge-control regimes (Hilgartner, 2017) that structure international breeding networks. The fourth section returns to the concept of genetic gain, and shows how elements used as variables to compute the indicator are being prioritised as targets for investment, which in turn affects whether and how greater genetic gain contributes to major goals such as developing climate-adaptive agriculture. The fifth section investigates the implications of accelerating genetic gain for seed systems, focussing on how those commitments to greater speed are directly linked to visions for greater commercialisation of agriculture in the Global South. In the conclusion, we broaden our gaze to the place of agronomic indicators within the wider knowledge-control regimes of agricultural development, and ask what questions the above analysis poses to this field. We make some suggestions for implementing the use of indicators for agricultural development in ways that are comprehensible and responsive to stakeholders, and acknowledge a wider range of goals and values related to agroecological diversity and sustainability.

Methodologically, this article is primarily based on review and analysis of the relevant scientific literature on genetic gain in plant breeding. The analysis is also informed by collaborative research experience in this domain, taking place between 2018 and 2022, and including: participation in international plant science and agricultural data networks; informal discussions with breeders, data managers and others within the CGIAR; and presence at a range of conferences where relevant material has been presented and discussed. The article contributes to the existing historical and philosophical literature on

statistical techniques in agricultural research (e.g. Berry, 2015; Derry, 2015; Maat, 2008; Parolini, 2015a, 2015b; Theunissen, 2020), focussing on the use of a specific indicator in relation to broader transformations occasioned by data-intensive methods in this domain (Williamson et al., 2021; Williamson & Leonelli, 2022a).

2. Genetic gain: an old indicator of new relevance

2.1. Genetic gain as an indicator

The rate of genetic gain (ΔG), also known as the response to selection, has been defined as "the improvement in average genetic value in a population or the improvement in average phenotypic value due to selection within a population over cycles of breeding" (Rutkoski, 2019a, p. 219). Put another way, it is a measure of the change in the population average for a given trait or index (set) of several traits that is attributable to heritable genetic effects, and therefore to the selection decisions made within a breeding programme. Notably, the rate of genetic gain does not require or produce direct knowledge about the genetic architecture of any individual or the population. Rather, it uses statistical methods to account for the interaction of genotypic effects with environmental and error effects in phenotypic evaluation data collected from plants or animals in the field.

Genetic gain provides an advantage over direct measures of plant or animal performance, such as changes in yield quantity, by factoring in the extent to which such performance is due to consistent genetic characteristics of the population vis-a-vis external factors like field management or environmental conditions (e.g. Bell et al., 1995). Realising improvements in genetic gain for a trait therefore means obtaining reliable increases in performance across the range of environmental conditions that the population has been selected and tested in. This range can be adapted to capture a wide diversity of different environmental conditions, including those that characterise marginal and low-input agricultural environments. Ostensibly, this improves breeders' ability to target and breed more precisely for environments that are, for example, less agronomically standardised or are subject to greater climatic fluctuations. This has made genetic gain particularly appealing to those who want to improve the responsiveness of plant breeding to climate change, a topic we return to below and in section 4.[1]

The rate of genetic gain is calculated through a mathematical model known as the breeder's equation. There are various ways of composing the breeder's equation, but one contemporary version (from Cobb et al., 2019) is:

$$\Delta G = \frac{(\sigma_a)(i)(r)}{L}$$

Here the rate of genetic gain (ΔG) is calculated from additive genetic variance (σ_a), selection intensity (i) and selection accuracy (r), divided by the years per breeding cycle (L). Other variants may use slightly different variables. To provide some broad definitions of the parameters, additive genetic variance (σ_a) is the degree of variation for a trait in the population that is due to genetic rather than environmental effects or error, and that isn't directly attributable to interactive effects between genes (such as dominance or epistasis). Selection accuracy (r) is the square root of heritability (h^2). Heritability refers to the proportion of total variation in the phenotype that can be attributed to genetic variation, and expresses the reliability of phenotypic values as guides to the value of the population for breeding.[2] Taking the square root expresses heritability as a

---

[1] We are grateful to Paul B. Thompson for drawing our attention to observations in this paragraph.

[2] The use of measures of heritability has been widely critiqued, especially in the human sciences (e.g. Sarkar, 1998, pp. 71–100), but remains fundamental to animal and plant breeding, the latter focused on physiological traits such as yield and resistances to pests and pathogens.

nonlinear rather than a linear function, such that changes to heritability produce a variable rate of return to genetic gain. Selection intensity (*i*) refers to the percentage of a population selected in a breeding programme. Given a distribution of values for a given trait across the population, if one wants to improve the average it is necessary to select some percentage of the top performing individuals to advance to the next generation. Selecting, for example, only the top 2% will shift the average substantially more than selecting the top 40%. However, this comes at the cost of potentially narrowing genetic diversity, unless population size is increased.[3]

Variances and heritabilities are calculated in relation to a heterogeneous ensemble of data acquired from multiple generations of crops grown in field trials or other experimental sites, which may include historical or predicted data as well as purposefully collected data. Traits from the crops in question, including for instance flowering time and maximum plant height, are measured, calibrated to specific standards in order to facilitate comparisons among traits, and inputted into numerical models that yield an overarching estimate of variance for each trait.[4] These results are in turn processed further to produce an estimate for variance in each trait that may be statistically and mathematically compatible with the model. Only at this stage can an estimate of each variable used in the overarching model for genetic gain be calculated; and the model plays a crucial role in evaluating which of these factors should be prioritised relative to the others.[5]

It is not our purpose here to document in detail the technical complexity of processes involved in extracting a calculation of genetic gain from crop data. The reason we insist on such complexity is to underscore how genetic gain is not simply a measurement, but is better thought of as an indicator. As defined by Mary Morgan, indicators are "numbers that are not conceived as direct measurements of the concepts they relate to, but are understood to be indicators for characteristics relevant for those concepts" (2020, p. 113). Indicators thus derive from the assemblage of several different data series, each of which is taken to track one dimension of a complex, multi-dimensional phenomenon. Each characteristic of the phenomenon in question will have a separate data series, which is typically construed and processed to be tractable and legible by as many audiences as possible. When assembling the data series to develop an indicator, the highly processed data contained in the data series are standardised even further, in order to be aggregated and combined as required. Processing data to develop an indicator therefore involves numerous modifications to the original datasets, including practices of selection, calibration, clustering and abstraction. These modifications make an indicator into something different from a simple "representation" of the phenomenon being studied. As Morgan puts, it, indicators "offer numerical information about some aspect of one target in relation to the goal, but they are far from measuring or representing the target, let alone the overall goal that needs to be represented" (2020, p. 115). We contend that this is precisely the case with genetic gain: It is a composite of different variables and highly curated – but heterogeneous – datasets, and therefore a relatively abstract and idealised type of measure, albeit one that is put to work in breeding practice. This becomes especially important in relation to the uses of genetic gain as an evaluative tool, given that similar indicator values may belie significant

variation in how such equivalence was achieved. Of particular relevance are the pragmatic constraints involved in assembling measurements on genetic traits and breeding outputs; the conceptual commitments guiding decisions around which data cluster may best represent these traits; and, most importantly, the choices around how such data should be visualised and condensed in order to maximise their informational value (cf. Leonelli, 2016). Also significant are the structural incentives that accompany these constraints, especially since indicators are often developed and adopted as tools of evaluation and governance for breeding programs. We return to this point in section 4.

2.2. Genetic gain in contemporary plant breeding

The breeder's equation was initially developed by cattle breeder and quantitative geneticist Jay L. Lush in the 1930s to monitor genetic trends in a population. The time variable was introduced by S.A. Eberhart in the 1970s as a means of evaluating the efficiency of breeding programs, thereby establishing genetic gain as a rate (Eberhart, 1970). The final numerical value calculated from the breeder's equation can be used to easily compare different breeding methods. This can be calculated retrospectively, as realised rates, or prospectively, as estimated rates. In plant breeding, retrospective rates can be determined using historical evaluation data or through an 'era' trial, where stored samples are grown out and evaluated simultaneously. Rates can also be estimated using simulated data from crop models (see Rutkoski, 2019b). Genetic gain therefore provides a convenient indicator for evaluating and, crucially, designing breeding programmes.

Despite such convenience, use of the breeder's equation has been uneven between the fields of plant and animal breeding. For animal breeders, it has long been recognised as a fundamental tool for monitoring genetic trends, reflecting the longstanding importance of statistical measures in that field to assist in evaluating working herds and flocks that can otherwise be difficult to evaluate directly (cf. Derry, 2015; Hill, 2014; Theunissen, 2020). In plant breeding, by contrast, evaluations of genetic gain have typically been limited to infrequent era trials, with little systematic attempt to monitor rates (Rutkoski, 2019b, p. 982). Recent years have seen a surge of attention to genetic gain in plant breeding, however, with experts from CGIAR proposing that the rate of genetic gain should be established as a high-level key performance indicator for public and philanthropically funded breeding programmes (Covarrubias-Pazaran, 2020). This commitment comes with significant prescriptive force: Specifically, to increase the rate of genetic gain in CGIAR and partner national agricultural research services' breeding programs, the majority of which are located in the Global South, to a minimum of 1.5% per year (Hunt, N.D.). Beyond policy circles, this is frequently expressed as a general commitment to "accelerate" rates of genetic gain and plant breeding efforts as a whole (e.g. Atlin et al., 2017; Harfouche et al., 2019; Heffner et al., 2010; Spindel & McCouch, 2016; Varshney et al., 2021), thus incorporating commonly used metaphors around the velocity of biological productivity (Landecker, 2013).

The surge in attention to genetic gain forms part of a wider discourse regarding the role of quantitative genetics in plant breeding in the twenty-first century. Put broadly, quantitative genetics concerns the statistical analysis of quantitatively varying traits and their inheritance, as in genetic variances and heritabilities. As a mode of analysis, it can be contrasted with classical and molecular genetics in that it requires little direct knowledge of actual genetic architecture or functions (although it can profitably be combined with these). Breeding methods grounded in quantitative genetic population improvement have a long history in twentieth century plant breeding (Hallauer, 2007), with some noteworthy results: For example, using recurrent selection methods on a single, closed population of maize, the Illinois Long-Term Selection Experiments have shown progressive improvement for oil and protein content with no upper limit after more than 100 years (Dudley, 2007). Despite this, in recent years quantitative genetics has been cast as a neglected corner of plant breeding, especially when compared to animal

---

[3] For good recent introductions to the various concepts, see the manuals provided by the CGIAR Excellence in Breeding platform: https://excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals For an older introduction, see Falconer and Mackay's (1996) classic textbook on quantitative genetics.

[4] See Boumans and Leonelli (2020) for a discussion of the challenges encountered by plant scientists attempting to fit crop data collected from field trials into standardised formats and nomenclature used for modelling, especially when compared to empirical data inputted into economic models, where different epistemic principles are at play.

[5] See Waldmann (2001) for a classic example of calculating genetic variance in a particular plant species through a field experiment.

breeding (e.g. Hickey et al., 2017), with some commentators positioning the development of this area as the next frontier for plant breeding (Wallace et al., 2018). This is driven by the emergence of new technologies that allow a "reinvention" of quantitative genetics, in the words of plant breeder Rex Bernardo (2020), primarily through the integration of quantitative with molecular genetics. A key example of such methods is Genomic Selection, originally developed in animal breeding, which utilises genome-wide sets of molecular markers to predict the genetic value of individual plants for a given trait (Xu et al., 2020), with a range of advantages for speed and accuracy, as we discuss in section 3.1. In their analysis of this integration of quantitative and molecular genetics in an animal breeding context, Lowe and Bruce (2019) identify within it a new mode of what they term 'marker-centric' biology, in which quantitative genetics' ambivalence about genetic architecture is re-centred for the molecular age. What is distinctive about the corresponding changes in plant breeding is not simply the belated adoption of such a mode or related technologies, however, but the self-conscious discourse on reinvention that has accompanied it - and the normative significance attributed to indicators built from such markers.[6]

It is within this wider landscape of reinvention in discourse and practice that genetic gain has emerged as a focal point, not only as a working tool for breeders but also, in its capacity as a key performance indicator, as a symbolic and practical anchor around which to reorganise plant breeding. We focus here on a set of proposals for the adoption and increase of genetic gain that have been made by a broad coalition of actors working for the CGIAR and for the BMGF, including both breeders and executive figures. This includes a relatively coherent set of principles for breeding reorganisation that have been outlined across several publications (e.g. Atlin et al., 2017; Cobb et al., 2019; Santantonio et al., 2020), and the practical implementation of some of these principles in CGIAR activities. We discuss these in detail in subsequent sections. Underpinning these principles is an explicit concern for the efficiency of plant breeding. As noted above, the construction of genetic gain as a rate allows it to function as an indicator of efficiency. Its use therefore bears on funders' concerns for the cost-efficiency of breeding (e.g. Atlin, 2016). Yet the concerns for efficiency speak to much larger issues of what plant breeding may or may not be able to deliver on a global scale. A consistent theme through several of the publications noted above is the need to construct tightly integrated and optimised data-driven pipelines for breeding, stretching from more fundamental biological research and 'pre-breeding' activities, through breeding itself and on to dissemination of new crop varieties via seed systems (e.g. Cobb et al., 2019). A key driver behind this push for efficiency is the goal of making plant breeding more responsive to climate change. Alongside the advantages in targeting breeding to more diverse environmental conditions that refocusing on genetic gain may provide, improved efficiency is valued for the benefits it provides in speed—the "acceleration" mentioned above. Various commentators have drawn attention to the value of shorter breeding cycles and programmes, from initiation to new crop variety release, as a means to respond to rapidly changing climates in a timely manner, thus staving off crop failures, hunger and poverty (e.g. Atlin et al., 2017; Cairns & Prasanna, 2018; Lenaerts et al., 2019). This emphasis on climate adaptation provides one of the major rationales for the adoption of genetic gain and associated changes to plant breeding. We therefore return to it in section 4, where we address some of the implications of the nature of genetic gain as an indicator for this goal.

3. Changing regimes of plant breeding

Behind the proposal for genetic gain as a key performance indicator lie a range of proposed and ongoing changes to plant breeding, which can

broadly be classified as changes in the methods and technologies of breeding and related programmes, on the one hand, and changes to the overarching knowledge-control regime of plant breeding on the other. Following Steven Hilgartner, regimes concern the order of a "domain or activity, typically through some combination of formal rules, informal norms, material means, and discursive framings", while knowledge-control regimes specifically concern the production, regulation and use of knowledge (Hilgartner, 2017, p. 8). As also emphasised by Pestre (2003) and Cambrosio et al. (2014), regimes operate on a range of scales and to differing degrees of institutionalisation. Moreover, they are multiple, overlapping and not necessarily congruent. Examples include agreements on the dissemination of research data, academic publishing norms, researcher evaluation metrics and intellectual property regimes. We focus here on those changes to regimes that are relevant to the current and proposed establishment of genetic gain as an indicator.[7] As Hilgartner makes clear, knowledge-control regimes are co-produced at the intersection of new research agendas, governance efforts and technological change. This makes them not only complex and multi-dimensional forms of order, but subject to continual change, even while some of their elements become formalised and entrenched. This is reflected in the situation we are analysing, where material changes in breeding have been accompanied by the adoption of strict new evaluative mechanisms. At the same time, the penetration of these evaluative tools into international plant breeding is as yet uneven. In this section we will firstly review the major technological changes to breeding programmes, focussing on how different technical methods contribute to the constituent variables of genetic gain. We then examine the changing knowledge-control regimes and transnational organisation of breeding networks. Our primary focus will be on changes occurring around the CGIAR network, which incorporates public and philanthropically funded breeding programmes across fifteen large research centres spread across continents, and serves as a model for many other initiatives around the world.[8]

3.1. Modernising breeding programmes

Significant reorganisation efforts over the last decade have been focused on the modernisation of crop-specific breeding programs in the CGIAR, in large part through projects funded by the BMGF. These programs include a wide range of crops, from irrigated rice breeding at the International Rice Research Institute (IRRI) in the Philippines (Collard et al., 2019) through cassava breeding at the International Institute for Tropical Agriculture (IITA) in Nigeria (Wolfe et al., 2017).[9] A range of methods and technologies have been used to accelerate genetic gain, instituted in diverse combinations that vary depending on the biology of the particular crop species (whether it is temperate or tropical, its mode of reproduction and so on), the available resources for that crop (such as genetic marker panels and Genomic Selection models) and the existing state of programs and funding. These methods may contribute to genetic gain through enhancing genetic variance, selection intensity, heritability/selection accuracy, cycle length, or some combination of these.

---

[6] On the historical origins of quantitative genetics, see Provine (2001). For a broad history of quantitative genetics as it has been applied in animal breeding until the early twenty-first century, see Derry (2015).

[7] See also the work on indicators as tools of governance collected in Rottenburg et al. (2015).

[8] We should note that these changes are not representative of *all* CGIAR programmes, which include a wide diversity of approaches and objectives (cf. Curry & Lorek, under review), but are affecting several of the centralised initiatives around data-intensive breeding promoted within the network.

[9] These programs have been modernised through the Transforming Rice Breeding (2013–2018) and NextGen Cassava (2011–present) projects, respectively. Other projects with similar objectives include Delivering Genetic Gain in Wheat (2016–2019), managed by Cornell University and including the CGIAR's International Centre for Maize and Wheat Improvement (CIMMYT), Mexico, as a key partner.

3.1.1. Speed breeding

Manipulating the length of breeding offers the simplest means to accelerate genetic gain, by increasing the number of cycles realised over a given period of time (Cobb et al., 2019, p. 634). One way to achieve this is through "speed breeding" and related techniques such as Rapid Generation Advance that allow multiple generations of a crop to be grown in a year (Collard et al., 2017; Li et al., 2018; Watson et al., 2018). This includes the use of custom-built growing facilities, such as dedicated off-season greenhouses for temperate species, and early seed harvests (Chiurugwi et al., 2019).

3.1.2. Genomic prediction

Modelling methods such as Genomic Selection allow the traits of an individual plant to be predicted based on genome-wide marker data, prior to growing the plant and evaluating it (Voss-Fels et al., 2019; Xu et al., 2020). Such models allow selection decisions to be made at an earlier stage, with plants that score highly for certain traits immediately progressed to the next breeding cycle as parents. This can contribute to reducing breeding cycle length by anywhere from 2 to 5 years (Voss-Fels et al., 2019), but also to selection intensity and heritability. Intensity can be increased by selecting consistent numbers of plants within larger populations, but the costs of phenotypic evaluation at scale makes this prohibitive. Evaluation purely based on genotyping of samples rather than phenotyping plants has been suggested as a potential cost-effective solution (Cobb et al., 2019, p. 632). Prediction of breeding value based on genetic data may also be more reliable for selection of certain traits in comparison to visual evaluation, increasing heritability.

3.1.3. Environmental characterisation and multi-environment trials

As indicated in section 2.1, it is of great importance for breeders to ensure that they are selecting for plants whose performance is the result of a significant genetic component (high heritability), rather than those that result from environmental effects. Heritability/selection accuracy can be increased by evaluating and selecting across multiple environments, to correct for environmental influences. For the resulting genetic gain to be meaningful and reliable, however, the range of environments in which selection is undertaken must model the agricultural environments in which farmers will ultimately be growing the crop variety. Exposing breeding materials to a range of different environments that reflect what is known as the Target Population of Environments (TPE)— the varying environmental conditions of the target region for which a crop is being bred, including seasonal as well as geographical variation and predicted future climates (Chenu, 2015)—is therefore of great importance (Collard et al., 2019, pp. 6–7; Atlin et al., 2017, pp. 34–5). Indeed, plant breeder Bernardo has observed that the principal expectation of quantitative genetics in contemporary breeding is "to help identify which candidates have the best genotypic value … in a target population of environments" (2020, pp. 377–8). The TPE concept has a long history in plant breeding, especially for estimating the heritability of traits (e.g. Nyquist, 1991), but much like the breeder's equation the concept has not been widely adopted. More recently, researchers in the CGIAR Research Program on Climate Change, Agriculture and Food Security have stressed the importance of TPE characterisations in adapting agriculture to climate change (Ramirez-Villegas & Heinemann, 2015). In practice, this involves the use of advanced *in silico* methods for precision envirotyping and crop modelling (Ramirez-Villegas et al., 2020). Advocates of genetic gain have further encouraged this as a framework for designing multi-environment field trials (METs), and implementing them early in the breeding process (typically, if METs are undertaken at all it is very late in the process), thus increasing selection pressures for the target environments.

Among all of these methods, decreasing the length of breeding cycles is increasingly favoured as the primary means to accelerate genetic gain.[10] "Given the complexity of the other parameters in the breeder's equation, cycle time is the easiest to understand, cheapest to manipulate, and the most powerful parameter for increasing genetic gain" (Cobb et al., 2019, p. 634). This is in comparison to earlier investments in breeding improvement that have primarily focused on the other three variables, genetic variance, selection intensity and heritability/selection accuracy. Probably the most important rationale for this is cost-efficiency. Cobb et al. (2019, pp. 635–8) observe that while manipulating those three variables tends to be highly effective at first, there are diminishing returns and increasingly large investments are required to achieve them. The way that different variables are prioritised has implications for the impact genetic gain will have on agricultural systems, however. We analyse this in relation to climate-adaptive breeding in section 4.

3.2. New knowledge-control regimes

The above-mentioned methodological shifts within breeding programmes have been accompanied and shaped by a distinctive vision for plant breeding. As breeder Joshua Cobb and colleagues have observed,

> "Advances in statistics, quantitative and population genetics, molecular biology, genomics, phenomics, other -omics, and most recently machine learning and artificial intelligence offer the potential of transforming plant breeding programs toward a data-rich, evidence-based, and team-oriented process and away from the romantic tradition of an individual breeder as an artist" (Cobb et al., 2019, pp. 627–8, references removed).

This is a vision of breeding as a data-driven activity, where selection decisions can be made on the basis of relevant and extensive data flows rather than breeders' (or indeed other stakeholders') subjective evaluations. Achieving this involves "complex processes and will require breeding organizations to adopt a culture of continuous optimization and improvement," assert Cobb et al. (2019, p. 627). Such aims are built into the knowledge-control regimes that regulate plant breeding, especially since breeding modernisation programmes confront institutional and political-economic challenges that lend themselves to digitalised, data-intensive research solutions. Again in the words of Cobb and colleagues,

> "unlike private-sector organizations, public plant breeding programs struggle to leverage and benefit from the collective investment, skills and experience across crops that large transnational breeding companies can command. As a result, public plant breeding programs must form interactive communities of practice that allow them to aggregate demand and stimulate the development of low-cost genotyping, phenotyping, and open-access IT systems for storage, management, analysis, and exchange of data" (Cobb et al., 2019, p. 640).

Hence building capacity for new breeding methods among the diverse breeding institutes that serve the Global South, including CGIAR centres and national agricultural research services, is increasingly reliant on transnational networks and communities of practice that facilitate co-ordination of activity and knowledge exchange, as well as new norms of openness in research and in the development of technical tools (Leonelli, 2022). Data sharing is explicitly encouraged and supported through ever-expanding data infrastructures and related technologies and social venues. In recent years, a range of collaborative platforms and shared facilities have been developed to serve CGIAR and other public breeding programmes. These are particularly focused on developing capacity for data-intensive genomic techniques, such as Genomic Selection (cf. Hickey et al., 2017; Spindel & McCouch, 2016). They include a high-throughput genotyping facility at the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) in India, to serve both

---

[10] Other methods that we do not discuss here include the structured selection of source germplasm to maximise gains in recurrent selection programs (Atlin et al., 2017, pp. 32–33).

animal and plant breeding communities globally, and GOBii, the Genomic Open-Source Breeding Informatics Initiative, based at Cornell University, to provide open source data management, analysis and visualisation tools to public breeding programmes (Nti-Addae et al., 2019; Santantonio et al., 2020, pp. 9–10). Increasingly, these efforts are being coordinated via the Excellence in Breeding platform of the CGIAR, which was founded in 2016 and was initially proposed under the title "Platform for Genetic Gains". The platform, which brings together tools, services, data and access to technology for breeding modernisation, has similar but broader goals to GOBii, which is set to be absorbed within the platform, scaling up the deployment of resources and training for the implementation of genomic breeding methods, especially in Africa and South Asia (Freedman, 2020; Williams, 2019). Moreover, the platform is developing a single breeding information management system, the Enterprise Breeding System, that is intended to replace the diversity of systems (both generic and crop-specific) currently in use across CGIAR and public breeding programmes (Storr, 2019).[11]

These developments in capacity building are accompanied by new forms of funding, regulation and evaluation focused on shifting breeding activity towards the acceleration of genetic gain. Among funding bodies, the BMGF has provided early and persistent financial support for projects dedicated to refocusing breeding on genetic gain, as well as related shifts in infrastructures and modes of collaboration (e.g. Atlin, 2016). The acceleration of genetic gain is being adopted as a key policy item by CGIAR stakeholders across the board, as evidenced in the multi-funder Crops to End Hunger strategy for CGIAR plant breeding modernisation.[12] Beyond funding and policy, perhaps the most formal entrenchment of the indicator consists in the development of new regimes of audit. Take, for example, the Breeding Program Assessment Tool, developed by the University of Queensland and funded by the BMGF, which includes questions on genetic gain.[13] Evaluation via the Breeding Programme Assessment Tool is now obligatory for any plant breeding programme receiving BMGF funding, thereby extending the scope of genetic gain as a key indicator for breeding success. Following wider trends towards the use of altmetrics for performance evaluation in the sciences, it has also been proposed that evaluation of breeders themselves be shifted away from older methods of assessment such as number of journal articles published or varieties released and towards quantification of their practical contribution to realised genetic gain (Cobb et al., 2019, p. 641). Such regimes of evaluation provide additional control, pressures and incentives towards realising the culture of continuous optimisation described by Cobb et al. (2019).

In the remaining sections of the paper, we address some of the conceptual, practical and socio-economic implications that flow from the changes described above. In the next section, we return to our observations from section 2 on the nature of genetic gain as indicator in order to consider how differential investment in and enhancement of the component variables of genetic gain may affect the practical results of increased genetic gain. In particular, we focus on incentives that push for increasing the speed of breeding. This leads to our final section, where we discuss how rapid breeding is accompanied by complementary proposals to increase the rate of varietal uptake and replacement by farmers, through the commercialisation of smallholder seed systems.

4. Disaggregating indicators

While genetic gain is privileged as a "high-level" indicator for agricultural development (Covarrubias-Pazaran, 2020), Cobb et al. note that:

> "it is imperative for senior management and financial stakeholders to be aware of performance against key indicators such as the average age of parental material, selection intensities (i.e., number of parents in the crossing block), selection accuracies (i.e., heritability in multi-location trials), number of lines advanced to each stage, length of the breeding cycle, and percentage of external germplasm used as parental material." (2019, pp. 639–40)

Ostensibly similar figures for genetic gain will in practice have been achieved through improvement of different combinations of these sub-indicators. Understanding the figures for the constituent elements of genetic gain as well as the implications that come from changes to each of these is highly significant, as they hold different meanings for breeders, funders, policymakers and other stakeholders. Indeed, diverse participants in agricultural management and food production associate different forms of value to breeding practice, which range from scientific to economic, political and cultural. The value of faster plant breeding, for example, can encompass the direct economic value of improved cost efficiency, but also political value given the significance of fostering food systems that can rapidly respond to changing economic, environmental or geo-political conditions (as the food shortages created by the 2022 Russian invasion of Ukraine have made abundantly clear; see also Lenaerts et al., 2019).[14]

Establishing a balance between different forms of value is therefore a key task for breeding policy, planning and management, if breeding is to respond to the many different stakeholders invested in agricultural production globally. As has been noted by others working within the CGIAR, however:

> "In today's donor environment, CGIAR breeders are challenged to demonstrate genetic gains on the order of 2% per year, using clearly defined indicators to monitor progress. This can drive them toward the easier goals rather than ones that result from a complex analysis of social, physical and biological challenges, opportunities and goals" (CGIAR Gender and Breeding Initiative, 2018, p. 1).

The point is made in relation to achieving the social goals of agricultural research and development, such as gender equality and support for marginalised or under-resourced groups (Williamson & Leonelli, 2022b). Yet it also holds in regard to the different constituent methods that can be implemented to increase genetic gain. This can be seen in the contemporary emphasis on reducing the length of breeding cycles over other elements.

We can illustrate this by drawing a contrast between the utility of interventions on breeding cycle length and on heritability, respectively, in regard to a specific area: Climate-adaptive plant breeding.[15] Across the plant breeding literature that we have surveyed so far, one of the key benefits attributed to accelerating genetic gain is that it will allow the faster production of new crop varieties, thus allowing breeding systems to keep pace with the rapid change in environmental conditions expected under climate change in the twenty-first century. Reducing cycle length naturally forms a key pillar of this argument, whether through increasing selection gains from cycle to cycle or through increasing the number of cycles realised in a single year. Speed is of the essence, in this vision. Yet speed alone is only one part of the story. For rapid breeding to deliver

---

[11] Outside the CGIAR, the U.S. Department of Agriculture Agricultural Research Service (USDA-ARS) and Cornell University 'Breeding Insight' program (https://www.breedinginsight.org/) seeks to develop similar capacity for increasing genetic gains among smaller plant and animal breeding programs in the USA, initially focussing on blueberry, alfalfa, sweetpotato, grapes and trout. This is nevertheless premised on utilising infrastructures shared with the CGIAR, such as GOBii.

[12] See the strategy document: https://storage.googleapis.com/cgiarorg/2018/11/SC7-B_Breeding-Initiative-1.pdf Similar policy is also being adopted by USAID, for example in their Feed the Future program.

[13] https://plantbreedingassessment.org/.

[14] For similar considerations on breeding indices in animal breeding, see Nuffield Council on Bioethics (2021), pp. 155–160.

[15] This comparison is partly heuristic. In practice, improvements in the two can be pursued jointly via the same methods, as in Genomic Selection. Yet the two elements arguably deliver different contributions to climate adaptation, and this is worth thinking through at some length.

effective climate-responsive solutions, this must be accompanied by more precise breeding for specific environmental conditions, both current and predicted. This requires maintaining high rates of heritability within a closely defined set of environmental parameters, as can be established through the construction of a TPE and implementation of a corresponding breeding program design.

As described in section 3.1, there is increasing investment in multi-environment trials as key components of breeding programs, across multiple crops and institutes, in order to improve heritability rates. Yet identifying and accessing appropriate land and then implementing METs can be very costly for breeding institutes (Collard et al., 2019, p. 7). METs therefore tend to be restricted in number. Moreover, they remain highly controlled environments, and may therefore not be representative of working agricultural environments. As Ceccarelli (2015, p. 90) notes, the differences between research stations and target farming environments can in many cases render calculations of genetic gain "irrelevant". This problem has of course been noted and addressed by breeders. In selection theory, the divergence between trial and farm environments can be accounted for through estimation of a "correlated" response to selection (i.e. genetic gain), a measure of the accuracy with which performance in target farm environments can be predicted in selection environments (Atlin et al., 2001). The effectiveness of using such correlated measures, however, continues to be a matter of controversy (Ceccarelli, 2015, p. 90). In contemporary plant breeding, several funders and policymakers, including the BMGF and the United States Agency for International Development (USAID), have indicated a desire to extend the nascent regimes of evaluation to assess realised rates of genetic gain not simply in field trial environments, but directly on farm.[16] Such on-farm estimation of genetic gain would improve the reliability of crop variety performance for farmers, but poses several practical challenges relating to infrastructure and labour on the one hand (cf. Layden, 2020) as well as design on the other, regarding how many and which farms are selected as evaluation sites. The questions of design may prove especially challenging when trying to balance evaluation in a representative range of farm environments with climatic conditions that may as yet only be predicted for the target region.[17]

Our point here is that faster breeding and increased environmental precision are two distinct components of climate-adaptive breeding that require complicated investment and management, including in both genomics and field trials. Arguably, heritability is the more critical of the two for the production of reliable, climate-adaptive crop varieties and the more difficult to achieve, with speed playing an auxiliary role related to the predicted temporality of climate change and of systems for the dissemination of seed to farmers. Emerging knowledge-control regimes in plant breeding may produce incentives and disincentives that push breeders towards "the easier goals", as the authors of the Gender and Breeding Initiative report put it. It is especially important to maintain a critical eye on such structural issues given a further set of proposals that have been made by key proponents of genetic gain regarding speed and seed systems, as a complement to their primary proposals about breeding organisation and practice. These raise a number of questions about the impact of both sets of reorganisation (to plant breeding and to seed systems) on farmers and other stakeholders in agricultural systems.

5. Genetic gain in the seed system

The efficiency and impact of public plant breeding programmes have often been evaluated in terms of the number of varieties released, a measure that is now widely recognised as a "gross overestimate of the efficiency of a breeding program" (Ceccarelli, 2015, p. 88). There is increasing consensus that any account of the impact of breeding programmes must be focused on the adoption of new varieties by farmers. Alongside the arguments for breeding reorganisation, authors such as Atlin et al. (2017) and Cobb et al. (2019) have put forth complementary arguments that increased rates of genetic gain should be matched by increased rates of varietal adoption and replacement by farmers. This is frequently linked to the need to ensure that agricultural systems are climate-responsive, through widespread dissemination and cultivation of the newest, most adaptive varieties (e.g. Atlin et al., 2017; Cairns & Prasanna, 2018).

This has led to arguments that a wholesale transformation of seed systems in the Global South is necessary, drawing on the model of existing commercial systems:

> "Commercial farmers in temperate regions make data-driven cultivar choices, and replace varieties even for a very small potential yield increase. They have a high degree of confidence in the data provided by seed companies and extension services. Companies thus have a strong incentive to maximize rates of genetic gain, and to disseminate a steady stream of improved products. Matching the effectiveness of such systems in delivering climate change adaptation is a critical challenge for the public sector breeding and seed systems that serve most farmers in the developing world." (Atlin et al., 2017, p. 35)

Indeed, "continuous, rapid varietal turnover, the precondition for plant breeding to contribute effectively to climate change adaptation, is likely only sustainable in commercialized cropping systems where farmers frequently purchase seed" (Atlin et al., 2017, p. 35). Against this assertion, a broad range of risks have been identified with efforts to commercialise smallholder agriculture and seed systems at scale. These include dispossession of farmers from the seed that they rely on (Kloppenburg, 2004, 2010) and the alienation of farmers from practical and social knowledge in regard to varietal selection and cultivation (Flachs, 2019). Despite Atlin et al.'s depiction of farmers making data-driven choices, the scope for informed decision-making is recognised to be highly dependent on formal knowledge-dissemination networks, and on relations with individual, institutional and material actors that may preclude the need for detailed decision-making, for better or for worse. "Farmers need enough trust in both the products and the information provided about them to ensure that they demand improved varieties based on data and recommendations provided by reliable advisers, rather than just on visual demonstration" (Atlin et al., 2017, p. 36). This produces a degree of risk for all parties, however, in which data may not prove reliable when breeding outputs move from research and breeding environments to the less controlled physical environments and different epistemic environments of the farm. The efforts to establish on-farm estimations and evaluations of genetic gain are one attempt to control for such unreliability. Nevertheless, even where such efforts are established, understanding and trust in detailed quantitative data about varietal performance may be lacking on the part of farmers, who may draw on different forms of judgement and expertise (such as that drawn from practical experience) to make their farming decisions (cf. Howlett & Velkar, 2010).[18]

Regardless, the extent to which farmer decision-making is seen as critical to such systems is limited in practice. "In the high-functioning

---

[16] E.g. comments made by Nora Lapitan (US AID Feed the Future programme) during talk entitled 'A Strategic Approach to Modernization of the Breeding Pipeline and Delivery of Genetic Gains', Plant and Animal Genome XXVIII conference, San Diego, 11 Jan 2020.

[17] Participatory breeding methods, in which selection is conducted on farm, present an alternative method to increase on-farm genetic gain. We discuss this at greater length in AUTHORS (Williamson & Leonelli, 2022b; see also Ceccarelli, 2015).

[18] To provide an example from animal breeding, Lonkila (2017) describes how the quantified indexes of performance of dairy cattle bred using Genomic Selection in Finland did not correspond to farmers' ways of evaluating cattle and nor did quantified evaluations maintain their reliability on farm, leading to contestation between farmers and breeders.

seed systems that offer farmers the best protection from climate change, new varieties are pushed into farmers' fields, not pulled" (Atlin et al., 2017, p. 36). As such, responsibility for stimulating varietal turnover is delegated to seed suppliers, regulators and other governmental institutions to leverage indirect control through means such as policy changes, withdrawal of older varieties or even penalties for their use (2017, p. 36; Spielman & Smale, 2017).

Given the risks outlined above, it is necessary to ask whether the objective of increasing varietal turnover serves the ends described. As we have made clear in the previous section, the speed of breeding that underpins speed of turnover is only meaningful as a climate-adaptive measure if it is accompanied by substantive investments in complementary breeding objectives.

6. Conclusion

While discussing the value of indicators of (in her case, socioeconomic) development, Morgan argues that: "while these data certainly contain information indicating characteristics of development, they should not be considered measurements of development" (Morgan, 2020, p. 116). As an indicator of agronomic and breeder performance, genetic gain forms one among a wider set of indicators that form a key part of the knowledge-control regimes of agricultural research and development. These include, to take one example, the indicators comprising Goal 2, Zero Hunger, of the UN Sustainable Development Goals (cf. Morgan & Bach, 2018). They also include indicators more directly associated with the concerns and objectives regarding seed systems described in the previous section. For example, a key method for assessing varietal turnover is through calculating the 'weighted average age of varieties' sown across a given geographical area (see Brennan & Byerlee, 1991). How such indicators are used in practice to govern development interventions, and how they are brought into relation with data on genetic gain (for the two are closely linked in the imperatives for "accelerated" breeding that we have discussed throughout the paper) is a question that requires further research. Indeed, it is a particularly important question given that it is likely to have significant effects on how 'the field' as object of intervention moves from being a set of environmental parameters associated with agronomic performance (*à la* in-field analyses of genetic gain) to a concrete environment inhabited by farmers and other stakeholders whose practices and decisions are being evaluated and governed in regard to crop selection and cultivation.

What is clear is that the cumulative picture provided by these sets of indicators does not, as Morgan indicates, provide a measure of development. This has to do with the very nature of indicators as epistemic objects. As abstract figures calculated from multiple data series they can be composed in different ways, and the particular ways in which they are constructed are critical to understanding their meaning and utility, as we demonstrated in regard to genetic gain and climate adaptation. At the same time, the development and choice of indicators is always expressive of particular sets of concerns. The special importance attributed to breeding speed, and to the commercialisation of seed systems, is linked to a particular vision of agricultural development (Williamson & Leonelli, 2022b), whose long historical roots (Scott, 2017) have found full expression within the so-called Green Revolution (Curry, 2022). Within this vision, biodiversity is often interpreted as a repository of crop varieties whose agronomic value can be tested, improved and eventually sanctioned through research and related commercialisation – a model that Fenzi and Bonneuil characterise as "resourcist", to underscore its emphasis on plants as genetic "resources" from which value should be extracted (Fenzi & Bonneuil, 2016). For example, an additional recommendation alongside the methodological changes described in section 3 has been that breeders' source material should be restricted to elite germplasm, i.e. that from high-performing (and often commercial) varieties (e.g. Atlin et al., 2017, pp. 32–3; Cobb et al., 2019, pp. 629–30). This is warranted in terms of immediate increases to genetic gain, since it makes it possible for breeders to avoid engaging in lengthy backcrossing

processes with non-elite varieties such as farmers' landraces. Nevertheless, calculations of genetic variance, as favourable as they may be, cannot account for the wide range of values of crop biodiversity, including not only for breeding itself (Curry, 2017) but also for health and wellbeing, cultural identity, ecology and sustainability (see Hunter et al., 2017). Moreover, this recommendation introduces further restrictions on breeders in terms of intellectual property regimes. Because most elite varieties derive from commercial breeding programs, and are therefore covered by intellectual property rights, the range of available material for public breeding is in practice limited to those elite varieties released as public goods by CGIAR or other public breeders (Atlin et al., 2017, pp. 32–3). Those released by the CGIAR may also be subject to the organisation's own Intellectual Asset principles (Lopez-Noriega et al., 2019).

Pointing out the limits of genetic gain calculations, and the knowledge control regime in which they are deployed, does not diminish their importance as a working component of plant breeding. As proponents have argued, assessing the rate of genetic gain can be extremely valuable when attempting to quantify and compare the outputs of breeding models and specific programmes, especially in relation to the improvement of quantitative genetic traits and controlling for the effects of genotype-environment interactions. But it is worth critically reflecting on the role such calculations are given in regard to agricultural development and in mitigating the impact of climate change on agriculture. Whether genetic gain should be used as *the* key indicator performance indicator for plant breeding is a question that should be widely debated, with an eye to the importance of developing complementary indicators that take a range of views on what constitutes agricultural development into account (e.g. Ceccarelli, 2015). In other words, we argue that awareness of both the usefulness and the limits of genetic gain as an indicator, as depicted above, needs to translate into an enriched framing of the knowledge control regimes associated with data-intensive agriculture.

It is not within the scope of this paper to provide a comprehensive analysis of what such an enriched framing should involve: such an elaboration requires transdisciplinary cooperation across many stakeholders in the agricultural system, as we have demonstrated through our long-standing collaboration with researchers and institutions from multiple sectors involved in agricultural development (Williamson & Leonelli, 2022a). What we can provide, in closing, are some suggestions for what agricultural knowledge control regimes should include going forward. Alongside the technologies supporting primary breeding indicators like genetic gain, we support the adoption of metrics and infrastructures specifically oriented towards agrobiodiversity and agroecological conservation, the diversification of seed systems, and mitigating the differential impact that climate change is having on soil, crops and communities. Moreover, there needs to be a commitment towards collecting, preserving and disseminating information – such as meta-data and documentation on processes of data standardisation – that will enable stakeholders to disaggregate indicators such as genetic gain and understand their components and the different implications that they pose, thereby increasing the ability to critically engage with performance indicators for rapid breeding. This has concrete implications for current investment in data-intensive agriculture. It calls for resources to: make the underpinning data and models accessible and scrutinizable, which helps to disaggregate indicators and verify their provenance; strengthen the governance of data infrastructures through participative engagement that facilitate farmers' input into the production and use of related resources and resulting indicators; and identify and consider development goals beyond high-yield agriculture, including the valorisation of agrodiversity for crops, soil and relevant farming communities.

CRediT author statement

Hugh F. Williamson: Conceptualization; Investigation; Writing - original draft; Writing - review & editing. **Sabina Leonelli**: Conceptualization; Writing - original draft; Writing - review & editing.

Declaration of competing interest

Acknowledgements

References

Atlin, G. (2016). *Modernizing plant breeding programs to deliver higher rates of genetic gain in the developing world*. ICRISAT seminar, April 29th. Available at: https://youtu.be/r2 Zh64QD8TQ.

Atlin, G. N., Cairns, J. E., & Das, B. (2017). Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Global Food Security, 12*, 31–37. https://doi.org/10.1016/j.gfs.2017.01.008

Atlin, G. N., Cooper, M., & Bjornstad, A. (2001). A comparison of formal and participatory breeding approaches using selection theory. *Euphytica, 122*, 463–475. https://doi.org/10.1023/A:1017557307800

Bell, M. A., Fischer, R. A., Byerlee, D., & Sayre, K. (1995). Genetic and agronomic contributions to yield gains: A case study for wheat. *Field Crops Research, 44*(2–3), 55–65. https://doi.org/10.1016/0378-4290(95)00049-6

Bernardo, R. (2020). Reinventing quantitative genetics for plant breeding: Something old, something new, something borrowed, something BLUE. *Heredity, 125*, 375–385. https://doi.org/10.1038/s41437-020-0312-1

Berry, D. J. (2015). The resisted rise of randomisation in experimental design: British agricultural science, c.1910-1930. *History and Philosophy of the Life Sciences, 37*(3), 242–260. https://doi.org/10.1007/s40656-015-0076-8

Boumans, M., & Leonelli, S. (2020). From Dirty Data to Tidy Facts: Clustering Practices in Plant Phenomics and Business Cycle Analysis. In S. Leonelli, & N. Tempini (Eds.), *Data Journeys in the Sciences*. Cham: Springer.

Brennan, J. P., & Byerlee, D. (1991). The rate of crop varietal replacement on farms: Measures and empirical results for wheat. *Plant Varieties and Seeds, 4*, 99–106.

Byrum, J., Beavis, B., Davis, C., Doonan, G., Doubler, T., Kaster, V., Mowers, R., & Parry, S. (2017). Genetic gain performance metric accelerates agricultural productivity. *INFORMS Journal on Applied Analytics, 47*(5), 369–471. https://doi.org/10.1287/inte.2017.0909

Cairns, J. E., & Prasanna, B. M. (2018). Developing and deploying climate-resilient maize varieties in the developing world. *Current Opinion in Plant Biology, 45*, 226–230. https://doi.org/10.1016/j.pbi.2018.05.004

Cambrosio, A., Keating, P., & Nelson, N. (2014). Régimes Thérapeutiques et Dispositifs de Preuve en Oncologie: L'organisation des Essais Cliniques, des Groupes Coopératifs aux Consortiums de Recherche. *Sciences Sociales et Santé, 32*, 13–42.

Ceccarelli, S. (2015). Efficiency of plant breeding. *Crop Science, 55*, 87–97. https://doi.org/10.2135/cropsci2014.02.0158

CGIAR Gender and Breeding Initiative. (2018). *Gender-responsive product profile development tool*. Workshop Report, Nov 12-13. Available at: https://cgspace.cgiar.org/handle/10568/99094.

Chenu, K. (2015). Characterizing the crop environment – nature, significance and applications. In V. O. Sadras, & D. F. Calderini (Eds.), *Crop physiology: Applications for genetic improvement and agronomy* (2nd ed.). London: Academic Press.

Chiurugwi, T., Kemp, S., Powell, W., & Hickey, L. T. (2019). Speed breeding orphan crops. *Theoretical and Applied Genetics, 132*, 607–616. https://doi.org/10.1007/s00122-018-3202-7

Cobb, J. N., Juma, R. E., Biswas, P. S., Arbelaez, J. D., Rutkoski, J., Atlin, G., Hagen, T., Quinn, M., & Hwa Ng, E. (2019). Enhancing the rate of genetic gain in public-sector plant breeding programs: Lessons from the breeder's equation. *Theoretical and Applied Genetics, 132*, 627–645. https://doi.org/10.1007/s00122-019-03317-0

Collard, B. C. Y., Beredo, J. C., Lenaerts, B., Mendoza, R., Santelices, R., Lopena, V., Verdeprado, H., Raghavan, C., Gregorio, G. B., Vial, L., Demont, M., Bisawas, P. S., Iftekharuddaula, K. M., Rahman, M. A., Cobb, J. N., & Islam, M. R. (2017). Revisiting

rice breeding methods – evaluating the use of rapid generation advance (RGA) for routine rice breeding. *Plant Production Science, 20*(4), 337–352. https://doi.org/10.1080/1343943X.2017.1391705

Collard, B. C. Y., Gregorio, G. B., Thomson, M. J., Islam, M. R., Vergara, G. V., Laborte, A. G., … Cobb, J. N. (2019). Transforming rice breeding: Redesigning the irrigated breeding pipeline at the International Rice Research Institute (IRRI). *Crop Breeding, Genetics and Genomics, 1*, e190008. https://doi.org/10.20900/cbgg20190008

Covarrubias-Pazaran, G. E. (2020). Genetic gain as a high-level key performance indicator. In *CGIAR Excellence in Breeding Platform, Breeding Scheme Optimization Manual*. https://excellenceinbreeding.org/toolbox/tools/eib-breeding-scheme-optimization-manuals. (Accessed 28 January 2021).

Curry, H. A. (2017). Breeding uniformity and banking diversity: The genescapes of industrial agriculture, 1935-1970. *Global Environment, 10*(1), 83–113. https://doi.org/10.3197/ge.2017.100104

Curry, H. A. (2022). *Endangered maize: Industrial agriculture and the crisis of extinction*. Oakland: University of California Press.

Curry, H.A. and T. Lorek, eds. (n.d.). Research as Development: Historical Perspectives on Agricultural Science and International Aid in the CGIAR Era. Cambridge: Cambridge University Press. In preparation.

Derry, M. E. (2015). *Masterminding nature: The breeding of animals 1750-2010*. Toronto: University of Toronto Press.

Dudley, J. W. (2007). From means to QTL: The Illinois long-term selection experiment as a case study in quantitative genetics. *Crop Science, 47*(S3), S20–S31. https://doi.org/10.2135/cropsci2007.04.0003IPBS

Eathington, S. R., Crosbie, T. M., Edwards, M. D., Reiter, R. S., & Bull, J. K. (2007). Molecular markers in a commercial breeding program. *Crop Science, 47*(S3), S154–S163. https://doi.org/10.2135/cropsci2007.04.0015IPBS

Eberhart, S. A. (1970). Factors affecting efficiencies of breeding methods. *African Soils, 15*, 655–680.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow: Longman.

Fenzi, M., & Bonneuil, C. (2016). From "genetic resources" to "ecosystems services": A century of science and global policies for crop diversity conservation. *Culture, Agriculture, Food and Environment, 38*(2), 72–83.

Flachs, A. (2019). *Cultivating knowledge: Biotechnology, sustainability, and the human cost of cotton capitalism in India*. Tucson: University of Arizona Press.

Freedman, D. H. (2020). *New genomic tools help improve staple crops around the world*. Cornell College of Agriculture and Life Sciences News. April 16 https://cals.cornell.edu/news/new-genomic-tools-help-improve-staple-crops-around-world. (Accessed 27 October 2020).

Gao, C. (2019). The future of CRISPR technologies in agriculture. *Nature Reviews Molecular Cell Biology, 19*, 275. https://doi.org/10.1038/nrm.2018.2

Hallauer, A. R. (2007). History, contribution, and future of quantitative genetics in plant breeding: Lessons from maize. *Crop Science, 47*(S3), S4–S19. https://doi.org/10.2135/cropsci2007.04.0002IPBS

Harfouche, A. L., Jacobson, D. A., Kainer, D., Romero, J. C., Harfouche, A. H., Mugnozza, G. S., Moshelion, M., Tuskan, G. A., Keurentjes, J. J. B., & Altman, A. (2019). Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. *Trends in Biotechnology, 37*(11), 1217–1235. https://doi.org/10.1016/j.tibtech.2019.05.007

Heffner, E. L., Lorenz, A. J., Jannink, J.-L., & Sorrells, M. E. (2010). Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science, 50*, 1681–1690. https://doi.org/10.2135/cropsci2009.11.0662

Helliwell, R., Hartley, S., & Pearce, W. (2019). NGO perspectives on the social and ethical implications of plant genome-editing. *Agriculture and Human Values, 36*, 779–791. https://doi.org/10.1007/s10460-019-09956-9

Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W., & Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants. (2017). Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics, 49*(9), 1297–1303. https://doi.org/10.1038/ng.3920

Hilgartner, S. (2017). *Reordering life: Knowledge and control in the genomics revolution*. Cambridge, MA: MIT Press.

Hill, W. G. (2014). Applications of population genetics to animal breeding, from Wright, Fisher and Lush to genomic prediction. *Genetics, 196*(1), 1–16. https://doi.org/10.1534/genetics.112.147850

Howlett, P., & Velkar, A. (2010). Technology transfer and travelling facts: A perspective from Indian agriculture. In P. Howlett, & M. S. Morgan (Eds.), *How well do facts travel? The dissemination of reliable knowledge*. Cambridge: Cambridge University Press.

Hunt, A.. National breeding programs prepped to measure – and boost – genetic gain. CGIAR Excellence in Breeding Platform blog. (N.D.) https://excellenceinbreeding.org/blog/national-breeding-programs-prepped-measure-%E2%80%93-and-boost-%E2%80%93-genetic-gains. (Accessed 27 January 2021).

Hunter, D., Guarino, L., Spillane, C., & McKeown, P. C. (Eds.). (2017). *Routledge handbook of agricultural biodiversity*. London: Routledge.

Kloppenburg, J. R. (2004). *First the seed: The political economy of plant biotechnology, 1492-2000* (2nd ed.). Madison: University of Wisconsin Press.

Kloppenburg, J. R. (2010). Impeding dispossession, enabling repossession: Biological open source and the recovery of seed sovereignty. *Journal of Agrarian Change, 10*(3), 367–388.

Landecker, H. (2013). Post-industrial metabolism: Fat knowledge. *Public Culture, 25*(3), 495–522.

Layden, K. (2020). *Basic principles of genetic gain and feasibility of on-farm estimation* (Masters thesis). Iowa State University.

Lenaerts, B., Collard, B. C. Y., & Demont, M. (2019). Review: Improving global food security through accelerated plant breeding. *Plant Science, 287*, 110207. https://doi.org/10.1016/j.plantsci.2019.110207

Leonelli, S. (2016). In *Data-Centric Biology: A Philosophical Study*. Chicago: University of Chicago Press.

Leonelli, S. (2022). How Data Cross Borders: Globalizing Plant Knowledge through Transnational Data Management and Its Epistemic Economy. In J. Krige (Ed.), *Knowledge Flows in a Global Age: A Transnational Approach*. Chicago: University of Chicago Press.

Li, H., Rasheed, A., Hickey, L. T., & He, Z. (2018). Fast-forwarding genetic gain. *Trends in Plant Science, 23*(3), 184–186. https://doi.org/10.1016/j.tplants.2018.01.007

Lonkila, A. (2017). Making Invisible cattle: Commodifying genomic knowledge in dairy cattle breeding. *Trace: Finnish Journal for Human-Animal Studies, 3*, 28–52. https://doi.org/10.23984/fjhas.53184

Lopez-Noriega, I., Halewood, M., Abberton, M., Amri, A., Angarawai, I. I., Anglin, N., , … Payne, T., et al. (2019). CGIAR operations under the plant treaty framework. *Crop Science, 59*(3), 819–832. https://doi.org/10.2135/cropsci2018.08.0526

Lowe, J. W. E., & Bruce, A. (2019). Genetics without genes? The centrality of genetic markers in livestock genetics and genomics. *History and Philosophy of the Life Sciences, 41*, 50. https://doi.org/10.1007/s40656-019-0290-x

Maat, H. (2008). Statistics and field experiments in agriculture: The emerging discipline of inferential statistics. In J. G. S. J. Maarseveen (Ed.), *The statistical mind in modern society: The Netherlands 1850-1940*. Amsterdam: Askant.

McGoey, L. (2016). *No Such Thing as a Free Gift: The Gates Foundation and the Price of Philanthropy*. London: Verso.

Montenegro de Wit, M. (2020). Democratizing CRISPR? Stories, practices and politics of science and governance on the agricultural gene editing frontier. *Elementa: Science of the Anthropocene, 8*, 9. https://doi.org/10.1525/elementa.405

Morgan, M. S. (2020). The datum in context: Measuring frameworks, data series and the journeys of individual datums. In S. Leonelli, & N. Tempini (Eds.), *Data Journeys in the Sciences*. Cham: Springer. https://doi.org/10.1007/978-3-030-37177-7_6.

Morgan, M. S., & Bach, M. (2018). Measuring development—from the UN's perspective. *History of Political Economy, 50*(S1), 193–210. https://doi.org/10.1215/00182702-7033932

Nti-Addae, Y., Matthews, D., Ulat, V. J., Syed, R., Sempéré, G., Pétel, A., Renner, J., Larmande, P., Guignon, V., Jones, E., & Robbins, K. (2019). Benchmarking database systems for Genomic Selection implementation. *Database, 2019*, baz096. https://doi.org/10.1093/database/baz096

Nuffield Council on Bioethics. (2021). *Genome editing and farmed animal breeding: Social and ethical issues*. London: Nuffield Council on Bioethics. Report https://www.nuffieldbioethics.org/publications/genome-editing-and-farmed-animals.

Nyquist, W. E. (1991). Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences, 10*(3), 235–322. https://doi.org/10.1080/07352689109382313

Parolini, G. (2015a). In pursuit of a science of agriculture: The role of statistics in field experiments. *History and Philosophy of the Life Sciences, 37*(3), 261–281. https://doi.org/10.1007/s40656-015-0075-9

Parolini, G. (2015b). The emergence of modern statistics in agricultural science: Analysis of variance, experimental design, and the reshaping of research at Rothamsted Experimental Station, 1919-1933. *Journal of the History of Biology, 48*, 301–335. https://doi.org/10.1007/s10739-014-9394-z

Pestre, D. (2003). Regimes of knowledge production in society: Towards a more political and social reading. *Minerva, 41*, 245–261. https://doi.org/10.1023/A:1025553311412

Pirscher, F., Bartkowski, B., Theesfeld, I., & Timaeus, J. (2018). Nature-identical outcomes, artificial processes: Governance of CRISPR/Cas genome editing as an ethical challenge. In H. S. James (Ed.), *Ethical Tensions from new technology: The case of agricultural biotechnology*. Wallingford: CABI.

Provine, W. B. (2001). *The origins of theoretical population genetics*. Chicago: University of Chicago Press.

Ramirez-Villegas, J., & Heinemann, A. B. (2015). *Environmental characterisation to guide breeding decisions in a changing climate*. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS). Working Paper No. 144.

Ramirez-Villegas, J., Milan, A. M., Alexandrov, N., Asseng, S., Challinor, A. J., Crossa, J., , … Sukumaran, S., et al. (2020). CGIAR modelling approaches for resource-constrained scenarios: I. Accelerating crop breeding for a changing climate. *Crop Science, 60*(2), 547–567. https://doi.org/10.1002/csc2.20048

Richardson, S., & Stevens, H. (Eds.). (2015). *Postgenomics: Perspectives on biology after the genome*. Durham: Duke University Press.

Rottenburg, R., Merry, S. E., Park, S.-J., & Mugler, J. (Eds.). (2015). *The world of indicators: The making of governmental knowledge through quantification*. Cambridge: Cambridge University Press.

Rutkoski, J. E. (2019a). A practical guide to genetic gain. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 157). Academic Press.

Rutkoski, J. E. (2019b). Estimation of realized rates of genetic gain and indicators for breeding program assessment. *Crop Science, 59*, 981–993. https://doi.org/10.2135/cropsci2018.09.0537

Santantonio, N., Atanda, S. A., Beyene, Y., Varshney, R. K., Olsen, M., Jones, E., Roorkiwal, M., Gowda, M., Bharadwaj, C., Gaur, P. M., Zhang, X., Dreher, K., Ayala-Hernandez, C., Crossa, J., Perez-Rodriguez, P., Rathore, A., Gao, S. Y., McCouch, S., & Robbins, K. R. (2020). Strategies for effective use of genomic information in crop breeding programs serving Africa and South Asia. *Frontiers in Plant Science, 11*, 353. https://doi.org/10.3389/fpls.2020.00353

Sarkar, S. (1998). *Genetics and reductionism*. Cambridge: Cambridge University Press.

Scott, J. C. (2017). *Against the grain: A deep history of the earliest states*. New Haven and London: Yale University Press.

Spielman, D. J., & Smale, M. (2017). *Policy options to accelerate variety change among smallholder farmers in South Asia and Africa South of the Sahara*. IFPRI Discussion Paper 01666.

Spindel, J. E., & McCouch, S. R. (2016). When more is better: How data sharing would accelerate genomic selection of crop plants. *New Phytologist, 212*, 814–826. https://doi.org/10.1111/nph.14174

Storr, S. (2019). *First steps taken to unify breeding software*. CIMMYT News. November 21 https://www.cimmyt.org/news/first-steps-taken-to-unify-breeding-software/. (Accessed 27 October 2020).

Theunissen, B. (2020). *Beauty or statistics: Practice and science in Dutch livestock breeding, 1900-2000*. Toronto: University of Toronto Press.

Varshney, R. J., Abhishek, B., Roorkiwal, M., Barmukh, R., Cowling, W. A., Chitikineni, A., Lam, H.-M., Hickey, L. T., Croser, J. S., Bayer, P. E., Edwards, D., Crossa, J., Weckwerth, W., Millar, H., Kumar, A., Bevan, M. W., & Siddigue, K. H. M. (2021). Fast-forward breeding for a food-secure world. *Trends in Genetics, 37*(12), 1124–1136. https://doi.org/10.1016/j.tig.2021.08.002

Voss-Fels, K. P., Herzog, E., Dreisigacker, S., Sukumaran, S., Watson, A., Frisch, M., Hayes, B., & Hickey, L. T. (2019). "SpeedGS" to accelerate genetic gain in spring wheat. In T. Miedaner, & V. Korzun (Eds.), *Applications of genetic and genomic research in cereals*. Elsevier. https://doi.org/10.1016/B978-0-08-102163-7.00014-4.

Waldmann, P. (2001). Additive and non-additive genetic architecture of two different-sized populations of *Scabiosa canescens*. *Heredity, 86*, 648–657. https://doi.org/10.1046/j.1365-2540.2001.00873.x

Wallace, J. G., Rodgers-Melnick, E., & Buckler, E. S. (2018). On the road to breeding 4.0: Unraveling the good, the bad, and the boring of crop quantitative genetics. *Annual Review of Genetics, 52*, 26.1–26.24. https://doi.org/10.1146/annurev-genet-120116-024846

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M.-D., , … Edwards, D., et al. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants, 4*, 23–29. https://doi.org/10.1038/s41477-017-0083-8

Williams, J. (2019). *Excellence in breeding: Developing breeding programs in Africa and South Asia*. Global Engage. July 29 http://www.global-engage.com/agricultural-bio technology/excellence-in-breeding-developing-breeding-programs-in-africa-and-so uth-asia/. (Accessed 27 October 2020).

Williamson, H. F., Brettschneider, J., Caccamo, M., Davey, R. P., Goble, C., Kersey, P. J., … Leonelli, S. (2021). Data management challenges for artificial intelligence in plant and agricultural research [version 1; peer review: 1 approved with reservations]. *F1000, 10*, Article 324. https://doi.org/10.12688/f1000research.52204.1

Williamson, H. F., & Leonelli, S. (Eds.). (2022a). *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. Cham: Springer. In press.

Williamson, H. F., & Leonelli, S. (2022b). Cultivating Responsible Plant Breeding Strategies: Conceptual and Normative Commitments in Data-Intensive Agriculture. In H. F. Williamson, & S. Leonelli (Eds.), *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. Cham: Springer. In press.

Wolfe, M. D., Del Carpio, D. P., Alabi, O., Ezenwaka, L. C., Ikeogu, U. N., Kayondo, I. S., , … Jannink, J.-L., et al. (2017). Prospects for genomic selection in cassava breeding. *The Plant Genome, 10*(3). https://doi.org/10.3835/plantgenome2017.03.0015

Xu, Y., Liu, X., Fu, J., Wanh, H., Huang, C., Prasanna, B., Olsen, M. S., Wang, G., & Zhang, A. (2020). Enhancing genetic gain through genomic selection: From livestock to plants. *Plant Communications, 1*(1), 100005. https://doi.org/10.1016/j.xplc.2019.100005