

# Data-driven decentralised breeding increases genetic gain in a challenging crop production environment

Kauê de Sousa<sup>1,2</sup>, Jacob van Etten<sup>2</sup>, Jesse Poland<sup>3</sup>, Carlo Fadda<sup>2</sup>, Jean-Luc Jannink<sup>4,5</sup>, Yosef Gebrehawaryat Kidane<sup>6,7</sup>, Basazen Fantahun Lakew<sup>6,8</sup>, Dejene Kassahun Mengistu<sup>6,7</sup>, Mario Enrico Pè<sup>7,9</sup>, Svein Øivind Solberg<sup>1</sup>, Matteo Dell’Acqua<sup>7,9[\*]</sup>

<sup>1</sup> Department of Agricultural Sciences, Inland Norway University of Applied Sciences, 2318 Hamar, Norway

<sup>2</sup> Bioversity International, 00054 Maccarese, Rome, Italy

<sup>3</sup> Department of Plant Pathology, Kansas State University, KS 66506 Manhattan, USA

<sup>4</sup> College of Agriculture and Life Sciences, Cornell University, NY 14853 Ithaca, USA

<sup>5</sup> Agricultural Research Service, United States Department of Agriculture, NY 14853 Ithaca, USA

<sup>6</sup> Bioversity International, Ethiopia Office, 1000 Addis Ababa, Ethiopia

<sup>7</sup> Institute of Life Sciences, Scuola Superiore Sant’Anna, 56124 Pisa, Italy

<sup>8</sup> Ethiopian Biodiversity Institute, 1000 Addis Ababa, Ethiopia

<sup>9</sup> Center for Climate Change Studies and Sustainable Actions (3CSA), Pisa, Italy

[\*]Correspondence should be addressed to: m.dellacqua@santannapisa.it

Crop breeding must embrace the broad diversity of smallholder agricultural systems to ensure food security to the hundreds of millions of people living in marginal production environments. This challenge can be addressed by combining genomics, farmers’ knowledge, and environmental analysis into a data-driven decentralised approach (3D-breeding). We tested this idea as a proof-of-concept by comparing a durum wheat (*Triticum durum* Desf.) decentralised trial distributed as incomplete blocks in 1,165 farmer-managed plots across the Ethiopian highlands with a benchmark representing genomic selection applied to conventional breeding. We found that 3D-breeding could double the accuracy of the benchmark. 3D-breeding could identify genotypes with enhanced local adaptation providing consistent yield advantages across seasons and locations. We propose this decentralised approach to leverage the diversity in farmers’ fields and complement conventional plant breeding to enhance local adaptation in challenging crop production environments.

## Introduction

The big data revolution in genomic tools has transformed plant breeding with inexpensive sequencing methods, enabling greatly accelerated variety development<sup>1–3</sup>. At present, plant breeders use data-driven methods, including genomic selection, to increase selection intensity while reducing the time of the breeding cycle and deriving greater genetic gain<sup>4</sup>. Most conventional breeding programs still rely on a centralised scheme aimed at maximizing genetic diversity (G) in the early stages of selection and then identifying superior germplasm on the basis of phenotypic observations made in a limited number of research stations with explicit environmental (E) and management (M) conditions. In this setting, genomic selection may be used to predict the performance of untested new genotypes but is bound to the  $G \times E \times M$  interactions captured by the research stations that are used to train the selection models<sup>5</sup>. This limitation of centralised breeding approaches may result in sub-optimal development and deployment of crop varieties for use by farmers seeking local adaptation in challenging environments<sup>6</sup>.

To respond to local needs impacted by climate change, breeders need to find new ways to accelerate variety development while directly addressing  $G \times E \times M$  interactions to the fullest<sup>3,7,8</sup>. Mobilizing farmers’ traditional knowledge of crop varieties and local adaptation can address this challenge<sup>6,9,10</sup> in a coherent, decentralised breeding program relying on farmer-participatory selection<sup>11–13</sup>. A crowdsourced citizen science approach has demonstrated the feasibility of a data-driven decentralised variety evaluation<sup>14</sup> that enables on-farm variety testing in a digitally supported and cost-efficient way<sup>15</sup>. Predictive accuracy of farmer selection criteria may outperform breeder evaluations even in a context of modern agriculture<sup>16</sup>.

Crowdsourced citizen science further integrates the E and M components into breeding by performing selection directly in target environments and using environmental data to analyse genotypic responses. Thus, the citizen science ap-

proach scales E and M data collection to generate a volume of data that matches the big data dimension of G. Combining genomic selection with citizen science opens the possibility of simultaneously capturing the three dimensions of crop performance, G, E, and M, in a data-driven way. Here, we describe and demonstrate potential benefits of this approach that we call *data-driven decentralised breeding*, or 3D-breeding, for short. Potentially, 3D-breeding could benefit the ~500 million smallholder farmers around the world who often produce in marginal, low-input environments and work with diverse cropping and farming systems and respond to local consumption preferences<sup>17</sup>.

We tested the 3D-breeding approach in the Ethiopian highlands, where many smallholder farmers grow durum wheat (*Triticum durum* Desf.) and select landraces following criteria related to environmental adaptation, food culture, and market demand<sup>18,19</sup>. Rich local wheat diversity has co-evolved with local cultures and landscapes over millennia. Consequently, Ethiopian farmers still often select and cultivate local landraces, which under local conditions tend to outperform modern varieties produced by centralised breeding<sup>20</sup>. In this context, 3D-breeding can leverage local wheat diversity and knowledge, and bring breeding closer to the target environments cutting through the complexity of  $G \times E \times M$ .

We compared 3D-breeding with a centralised breeding approach. In a centralised trial, we collected phenotypic data and farmer evaluations on a panel of 400 fully genotyped durum wheat lines derived from genebank accessions<sup>18</sup> in two managed fields commonly used for varietal selection in the Ethiopian highlands in 2012 and 2013. In a decentralised trial, we distributed a subset of 41 genotypes (Fig. S1) as packaged sets containing incomplete blocks of three genotypes, plus one commercial variety to each farmer, following the *tricot* citizen science approach<sup>15</sup>. A total of 1,165 farmers planted these packages on their farms across three administrative regions of Ethiopia (Fig. S2). We use the data from the centralised and decentralised trials to evaluate

whether 3D-breeding could complement genomics assisted breeding by increasing prediction accuracy in challenging environments.

## Results

### Benchmark: centralised breeding enhanced by genomic selection and farmer participation

We established a benchmark that represents a centralised breeding approach that is a competitive alternative to 3D-breeding. We focused on grain yield (GY) and farmers' overall appreciation of genotypes (OA), which were both recorded in centralised (station) and decentralised (farm) trials. Centralised stations and farmer fields belong to the same agroecological zones of Ethiopia (Fig. S3). To establish the benchmark, we used a genomic selection model trained on  $GY_{STATION}$  and  $OA_{STATION}$  to predict, respectively,  $GY_{FARM}$  and  $OA_{FARM}$  in 1,165 farmer fields located in the same breeding mega-environment (Fig. 1A). The benchmark represents a centralised breeding approach using farmer on-station selection. Its scope and size reflect a regional variety trial, an advanced stage in breeding that focuses on a set of genetic materials and target environments with the aim of selecting the best genotypes for varietal release and recommendation. The stations are commonly used as breeding field trials for Amhara and Tigray regions of Ethiopia, and differ in altitude, temperature, rainfall, and soil<sup>20</sup>. Additional multilocation trials would typically occur in earlier stages of the breeding cycle. On-station involvement of farmers is not common practice but is increasingly conducted in association with breeding<sup>12,16</sup> and makes the benchmark more competitive. 3D-breeding expands on the centralised genotype characterisation by moving the selection to farmer fields (Fig. 1B).

Heritability ( $H^2$ ), the proportion of phenotypic variance explained by genotypic variance, was 0.55 and 0.42 for  $GY_{STATION}$  across locations for 2012 and 2013 respectively (Table S1). Men

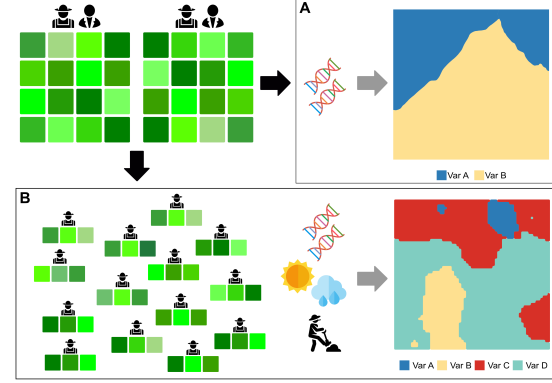


Fig. 1. Centralised breeding (A) derives recommendations from breeders' evaluation and possibly participatory assessments in a limited set of stations, using genomics to accelerate the production of varieties that are eventually recommended with coarse spatial resolution. This system may become more efficient if complemented by 3D-breeding (B), a decentralised approach where the best candidate genotypes are tested by farmers in small, blinded and randomized sets. 3D-breeding produces scalable solutions that can be linked to genomics, farmers' knowledge and environmental data, to enhance the local adaptation of the resulting varieties and tailor their recommendation to the landscape.

and women farmer provided consistent evaluations, although with differences across locations (Fig. S4). In order to capture farmers' traditional knowledge regardless of gender, farmer scores were combined across men and women respondents, the  $H^2$  of  $OA_{STATION}$  was 0.78 across locations (Table S2). We validated the centralised prospect by predicting on-station performance from one season to the next, focusing on a subset of 41 top-performing genotypes in managed trials that were later distributed in decentralised farmer fields. This led to accuracies up to  $\tau = 0.248$  in predicting  $GY_{STATION}$  in the following season (Fig. S5). We found that farmers' evaluations ( $OA_{STATION}$ ) were a better predictor than  $GY_{STATION}$  to capture both  $OA_{STATION}$  and  $GY_{STATION}$ , including when disaggregated by gender (Fig. S6). GY and OA collected in stations showed poor correlations with on-farm performance (Fig. S7).

The benchmark had a low prediction accuracy when using  $GY_{STATION}$  to predict  $GY_{FARM}$  in

individual seasons, with an average of  $\tau = 0.046$ . When using  $OA_{STATION}$  to predict  $OA_{FARM}$ , the average was  $\tau = 0.141$  (Table 1). Accuracy remained low when  $GY_{STATION}$  was used to predict measures of  $GY_{FARM}$  and  $OA_{FARM}$  combined across seasons. However,  $OA_{STATION}$  had consistent positive accuracy in predicting  $GY_{FARM}$  and  $OA_{FARM}$  (Fig. S8). This confirmed that genomic selection can be enhanced by farmers’ traditional knowledge whereas selection based only on GY could result in reduced appreciation by farmers (Fig. S9).

$GY_{STATION}$  provided a more accurate prediction of  $GY_{FARM}$  when restricting the model to cold-tolerant genotypes (Fig. S10). This was likely due to the partial representation of the climatic variation that can be provided by a centralised approach with a handful of stations (Fig. S11), as farms could experience lower temperatures than stations (Fig. S12). Still, centralised predictions of increasingly distant farm environments shown an erratic pattern, showing that variation at the farming sites goes beyond that captured by temperature variation (Fig. S13). Regardless the fact that both stations and farms were located in the same agroecological zone (Fig. S3), the benchmark failed to predict performance under production conditions, showing that the small-scale variation in climate and management may hamper the success of centralised breeding decisions.

### 3D-breeding provides higher prediction accuracy than the benchmark

3D-breeding uses  $GY_{FARM}$  and  $OA_{FARM}$  to generate a model that affords predictive extrapolation across space and time. We determined the accuracy of 3D-breeding with cross-validated Plackett-Luce trees<sup>21</sup> considering seasons as bins. Environmental indices were derived from seasonal time-series climatic conditions observed in each plot. In this case, the model selected the minimum night temperature during vegetative growth and maximum night temperature during reproductive growth as the most critical indices in determining the performance of genotypes (Fig.

S14). Genotypes’ DNA markers were included in the model as an additive matrix in a Bayesian framework. 3D-breeding consistently provided higher accuracy than the benchmark for  $GY_{FARM}$  and  $OA_{FARM}$  with  $\tau = 0.109$  and  $\tau = 0.251$  (Table 1). The prediction accuracy of the 3D-breeding approach was not biased towards specific environmental conditions, suggesting that it could capture the environmental diversity of test sites better than the benchmark (Fig. S15).

Table 1: Performance of the 3D-breeding compared with the benchmark of a centralised genomic selection. 3D-breeding provides higher across-season goodness-of-fit (Kendall  $\tau$ ) than centralised genomic selection on overall appreciation (OA) and grain yield (GY) derived from farmer rankings on decentralised fields.

Approach	OA	GY
<b>Centralised GS</b>		
Season 1 (n=179)	0.134	-0.012
Season 2 (n=651)	0.105	0.076
Season 3 (n=335)	0.183	0.073
	<b>0.141 (<math>\pm 0.03</math>)</b>	<b>0.046 (<math>\pm 0.04</math>)</b>
<b>3D-breeding</b>		
Season 1 (n=179)	0.270	0.160
Season 2 (n=651)	0.276	0.078
Season 3 (n=335)	0.203	0.119
	<b>0.251 (<math>\pm 0.04</math>)</b>	<b>0.109 (<math>\pm 0.04</math>)</b>

Overall appreciation of genotypes in 3D-breeding resulted in higher accuracies than  $GY_{FARM}$  in all farmers’ fields (Fig. S16). Previous studies showed that farmer evaluations are able to capture agronomic performance of genotypes in untested locations<sup>16,19</sup>, as confirmed by the high  $H^2$  observed for  $OA_{STATION}$  (Table S2). Farmers provided OA according to their own experience and preferences, and it presumably depended on a combination of traits of which GY represented only one dimension<sup>20</sup>. By eliciting traditional knowledge of men and women farmers at cropping sites, 3D-breeding successfully predicted varietal performance under local growing conditions (Fig. S8).  $GY_{FARM}$  is objectively and independently measured at each plot and therefore it could not be biased by  $OA_{FARM}$ . It is possible that  $GY_{STATION}$  and  $GY_{FARM}$  failed to capture secondary traits with high heritability (Table S1) that were observed by farmers and

that were correlated to the  $GY_{FARM}$  of genotypes under marginal conditions<sup>19,20</sup>. As  $OA_{FARM}$  is directly related to the probability of variety adoption it is an important complement to GY in driving varietal development for marginal environments.

### 3D-breeding provides consistent recommendations across seasons

Next, we extrapolated the 3D-breeding model predictions to assess the probability that the genotypes selected by 3D-breeding on the basis of OA will outperform current recommendations as per the Wheat Atlas<sup>22</sup>. We found that the best three genotypes in each terminal node of the 3D-breeding model splits had a genetic background markedly separated from that of varieties currently recommended for the region, and consistently higher *worth* (Fig. 2A). Indeed, the model selected genotypes derived from landraces over improved varieties. We estimated the probability that the model recommendation exceeds the current recommendation in terms of  $OA_{FARM}$ . In this assessment, predictions from 3D-breeding outperformed the current varietal recommendations<sup>22</sup> in most of the farmers' fields, with consistently higher probabilities (0.83-0.91), including in marginal areas for which the centralised breeding approach could not provide accurate predictions (Fig. 2B). To provide an agronomic measure, we also predicted the increase in  $GY_{FARM}$  and tested to see if the yield advantage could be maintained by selecting the best three genotypes indicated by 3D-breeding under 15 different growing seasons simulated on target farms. We found that 3D-breeding ensured consistent recommendations over years with expected increases in yield of about 20% (Fig. 2C). Thus, 3D-breeding accurately identified the best performing genotypes to be advanced in breeding efforts targeting local growing conditions, to be developed into suitable new varieties, and to be promoted with environmental-specific recommendations.

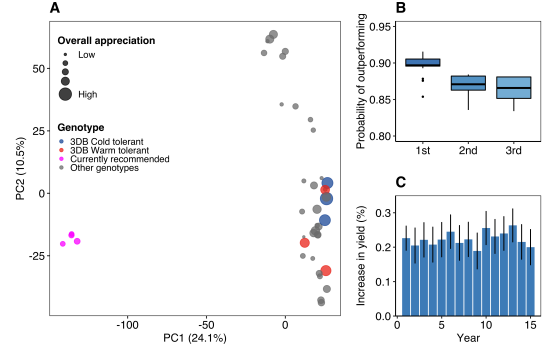


Fig. 2. Selection of durum wheat genotypes based on 3D-breeding. (A) Principal component coordinates of the genetic diversity of tested genotypes. Pink dots represent the varieties currently recommended for the area of study. 3DB Cold tolerant (blue) represents the top 3 genotypes selected by 3D-breeding in cold areas (minimum night temperature  $< 11.5^{\circ}\text{C}$ ). 3DB Warm tolerant (red) represents the top 3 genotypes selected by 3D-breeding in warm areas (minimum night temperature  $> 11.5^{\circ}\text{C}$ ). Size of dots represents the performance of genotypes in farmer fields as overall appreciation (OA). (B) Probability of outperforming improved varieties currently recommended by using genotype selection generated by 3D-breeding with OA. The panel shows the probability of the top 3 genotypes in a given location in outperforming the improved variety recommended for that location. (C) Expected increase in yield across 15 consecutive growing seasons (2001 to 2015) for genotype selection from 3D-breeding.

## Discussion

Genomic selection is a well-known approach to accelerate breeding programs, but current implementations in plant breeding have not yet reached their full potential. The earliest and most successful implementations of genomic selection have arguably occurred in dairy cattle breeding<sup>23</sup>. The accelerated evaluation of bull net merit was key to this<sup>24</sup>, but that success also depended on the fact that breeders had access to phenotyping data from a broad range of environments in the form of milking records, which farmers record for their own management benefit. In conventional crop breeding, all of the phenotyping costs fall on the breeding program and limit the number of target environments that can be represented in the selection process. 3D-breeding seeks to complement and expand the flow of information from a few centralised locations to the whole mega-environment where results from numerous decentralised observations and farmer knowledge may converge to inform breeding decisions.

In centralised breeding, the environmental variation of target environments is factored through experimental control or indirectly as an average response across breeding stations as in our benchmark. This makes extrapolation to real farming conditions challenging.  $G \times E$  affects yield and its components<sup>25,26</sup> and calls for selection models to explicitly account for it<sup>27</sup>. These models, however, are bound to the observations that can be made in resource-intensive breeding trials. Data from crowdsourced citizen science, like 3D-breeding, may further our understanding of the  $G \times E$  interactions that are observed in farmer fields and allow the integration of increasingly accurate seasonal prediction models<sup>28</sup> in breeding and germplasm recommendation pipelines.

The 3D-breeding approach addresses the low correlation between performance in selection environments and production environments, while taking a step forward to fully data-driven breeding. In this, 3D-breeding is a promising approach that could add to conventional breeding increasing varietal performance in smallholder agriculture, which accounts for more than 80% of all global

farms<sup>17</sup>. Here, the adoption rate of current breeding innovation may be suboptimal due to socioeconomic and environmental factors<sup>20,29–31</sup>. Climate change is pushing these farming systems to the edge of their adaptation capacity with increasing pressure from pest and diseases<sup>32,33</sup>, threats of yield loss<sup>7,34</sup> and increased seasonal climatic variability<sup>35,36</sup>, calling for tailored solutions. 3D-breeding may speed up the turnover of varietal release to address these challenges. As farmers are at the centre of the experimental design, varieties deriving from 3D-breeding are more likely to be adopted and suited to local cultivation<sup>10,37</sup>, increasing the effectiveness of breeding efforts. Indeed, we found that farmers' OA was a better predictor than GY in predicting yield realized both in centralised and decentralised trials (Table 1). Likewise, varieties derived from landraces consistently outranked the performance of improved varieties (Fig. 2A) derived from centralised breeding<sup>18</sup>. Beyond varietal recommendations, 3D-breeding can direct the choice of parents to crosses aiming at the production of recombinant lines to provide higher and more stable yields in local agriculture.

3D-breeding is useful beyond smallholder farming agriculture, and the citizen science approach on which it relies has already been applied to several crops to enhance the selection of climate-adapted varieties<sup>14</sup>. Its general scheme may also be useful in high-input, yield maximizing agriculture to enhance local adaptation and support sustainability and food security, where the usefulness of farmers' evaluations in a genomic setting was already demonstrated<sup>16</sup>. There are a number of open questions in relation to decentralised crop breeding, including how to best motivate new farmers to participate in the evaluation of materials, how much planting material each farmer needs, the logistics of providing farmers with the genetic material, and how to share benefits deriving from the utilisation of farmers' knowledge to produce new varieties.

3D-breeding may be most effective as a complement to a centralised breeding system providing a high-throughput evaluation of correlated traits to support earlier varietal selection to be tested

in farmer fields<sup>38</sup>. Accuracy is just one among the factors controlling genetic gain<sup>39</sup>, thus our findings should be integrated in the broader picture of modern breeding. Multi-trait models may increase prediction accuracy by measuring correlated traits with higher heritability<sup>38,40,41</sup>. These models could be employed in centralised stations and used to narrow down the set of varieties to be distributed to farmers in the 3D-breeding approach aiming to fine-tune local adaptation. Moreover, our findings support the need to further explore the challenge to model farmers' appreciation at the genomic level to improve the effectiveness of genotypes evaluation trials<sup>16</sup>.

The advantages provided by the approach are clear: phenotyping costs would be divided in much smaller packets, supporting the modular expansion of the breeding effort towards new genetic materials or new locations. In return, each generated datapoint would be a better representation of the true farming conditions to which varieties are directed. Previous research found that the involvement of farmers in selection experiments has negligible effects on costs<sup>42</sup>. In 3D breeding the costs are shared by farmers, who would in exchange obtain access to the best materials for their farm. Farmer preference would be collected directly on farms rather than derived from correlated metrics that come from on-station evaluations in centralised breeding. In terms of absolute costs, an implementation of 3D-breeding based on OA would only require seed amplification, seed distribution and telecommunications to obtain feedback from farmers. Genotyping costs are negligible thanks to ever increasing sequencing capabilities<sup>1</sup>. Indeed, a tricot experiment conducted in Nicaragua<sup>14</sup> resulted in a lower cost per datapoint in decentralised evaluations than in stations experiments.

The data-driven focus of 3D-breeding enables embracing the complexity of real-world  $G \times E$  for the benefit of breeding. Such a multidimensional, collaborative approach calls for best practices in data management and sharing<sup>43</sup>. 3D-breeding is based on a documented set of methods, from experimental design<sup>15</sup> to data curation and analysis<sup>21,44</sup>. While our demonstration of

these methods relied on a large dataset, we believe that much larger field sample designs and genomic variant datasets are quite feasible and will provide additional power, as is also much in evidence in livestock genetics. The expansion of the design with the addition of further testing seasons and local management conditions may allow to highlight drivers of local performance of genotypes beyond temperature<sup>45</sup>. The crowd-sourced citizen science approach associated with open-source digital tools makes it possible for breeders and farmers to apply 3D-breeding in new contexts and crops, dependent only on creativity in identifying untested production niches, potentiating a culturally-driven coevolution between farming systems and data-driven breeding to complement traditional breeding.

## Materials and Methods

### Genotypes sampling and DNA extraction

We selected 400 durum wheat (*Triticum durum* Desf.) genotypes from a representative collection of accessions from the Ethiopian Biodiversity Institute. Genomic DNA was extracted from fresh leaves pooled from five seedlings for each of the accessions with the *GenElute*<sup>TM</sup> Plant Genomic DNA Miniprep Kit (Sigma-Aldrich, St Louis, USA) following manufacturer's instructions in the Molecular and Biotechnology Laboratory at Mekelle University, Ethiopia. Genomic DNA was checked for quantity and quality by electrophoresis on 1% agarose gel and Nanodrop<sup>TM</sup> 2000 (Thermo Fisher Scientific Inc., Waltham, USA). Genotyping was performed on the Infinium 90k wheat chip at TraitGenetics GmbH (Gatersleben, Germany). Single nucleotide polymorphisms (SNPs) were called using the tetraploid wheat pipeline in GenomeStudio V11 (Illumina, Inc., San Diego, CA, USA). SNP calls were cleaned for quality by filtering positions and samples with failure rate above 80% and heterozygosity above 50%. Full details on the genotyping are given by Mengistu et al. (2016)<sup>18</sup>. The SNP calls for the genotypes included in this study and the details on the provenance of genotypes tested are

given as part of the full dataset on Dataverse<sup>46</sup>.

## Evaluation of genotypes in centralised trials

Centralised trials were performed in 2012 and 2013 in the districts of Geregera (Amhara) and Hagreselam (Tigray) (Fig. S1). The experimental stations were chosen to represent the highland agroecology of Ethiopia, and are often used as varietal testing sites for local agriculture. The trial was laid out in a replicated alpha lattice design, and field managements were conducted as per local guidelines with manual weeding. Accessions were sown in four rows 2.5 m long, at a seeding rate of  $100 \text{ kg} \cdot \text{ha}^{-1}$ . At sowing,  $100 \text{ kg} \cdot \text{ha}^{-1}$  diammonium phosphate and  $50 \text{ kg} \cdot \text{ha}^{-1}$  urea were applied, with additional  $50 \text{ kg} \cdot \text{ha}^{-1}$  urea at tillering. In 2012, thirty experienced smallholder farmers growing durum wheat (15 men and 15 women) were invited to participate in the trial evaluations at the station plots, held concurrently after flowering stage. The farmers had no previous knowledge of the genotypes included in this study to prevent bias in the evaluations. The participants provided appraisal with Likert<sup>47</sup> scales (1 to 5, worse to best) given to genotypes for overall appreciation (OA)<sup>19,20</sup>. Farmers did not use half-values in order to streamline the evaluation effort. Research technicians measured grain yield (GY) as grams of grain produced per plot, then converted into  $\text{t} \cdot \text{ha}^{-1}$ . Absolute values of GY and OA measured in centralised trials were converted into ordinal rankings.

## Evaluation of genotypes in decentralised trials

A total of 1,165 decentralised plots were established between 2013 and 2015 during three growing seasons across the regions of Amhara (471), Oromia (399) and Tigray (295) (Fig. S1) using a subset of the 41 best genotypes identified through farmer evaluation in centralised trials<sup>20</sup>. The farms were sampled in the same agroecological zones of the centralised fields (Fig. S3). Season 1 (2013) comprised 179 fields, Season 2 (2014) comprised 651 fields, and Season 3 (2015) comprised

335 field. Trials (farmer-managed plots) followed the triadic comparison of technologies (*tricot*) approach<sup>15</sup>. Sets of three local genotypes plus an improved variety (Asassa in Tigray and Amhara, and Hitosa and Ude in Oromia) were allocated randomly to farmers as incomplete blocks, maintaining spatial balance by assigning roughly equal frequencies of the genotypes. Trial size ranged from  $0.4 \text{ m}^2$  to  $1.6 \text{ m}^2$  depending on season and location. Farmers set, managed and evaluated their own experiments indicating the OA of genotypes through ranking the four entries that they received from best to worst, using pre-defined answer forms at the end of the growing season. Differently from the centralised trials, the OA was derived from the relative rankings of genotypes as each farmer evaluated a different set of materials. Research technicians collected GY measures in farmers' plots after harvesting.

## Centralised trait data analysis

All analyses were done in R<sup>48</sup>.  $GY_{STATION}$  and  $OA_{STATION}$  measured in centralised trials were used to derive best linear unbiased prediction (BLUP) values using the R package ASReml<sup>49</sup>, treating locations as a fixed factor and all other factors as random. Full model details are reported in Supplementary Note 1. For the central comparison between benchmark and 3D-breeding, we used measures of  $GY_{STATION}$  combined across seasons and locations. Similarly,  $OA_{STATION}$  in the central comparison represents OA values combined across genders and locations. When relevant,  $GY_{STATION}$  and  $OA_{STATION}$  measures are split by location, season or gender. Agreement between farmer gender groups in evaluating centralised station data was derived from a linear model fit. Spearman correlations between location specific BLUP values and farm performance were also computed.

## Decentralised trait data analysis

For the analysis of the decentralised data we used the Plackett-Luce model<sup>21,50,51</sup>. Plackett-Luce is a rank-based model that estimates the *worth* parameter. These parameters  $\alpha$  are related to the



probability ( $P$ ) that one genotype  $i$  wins against all other  $n$  genotypes in set, and are obtained using the following equation:

Equation [1]

$$P(i \succ \{j, \dots, n\}) = \frac{a_i}{a_1 + \dots + a_n}$$

Each farmer ranking in the input  $OA_{FARM}$  data was converted to a *dense* ranking, which ranked genotypes from 1 (first place) to  $n_r$  (last place). Genotypes not ranked in a given farmer plot had a rank of 0. For  $GY_{FARM}$ , we converted each observation into ordinal rankings, by assigning values from 1 (highest yield) to  $n_r$  (lowest yield), and then converted into a *dense* ranking. Altogether,  $OA_{FARM}$  and  $GY_{FARM}$  represented the probability of winning of any given genotype in any set of testing sites by either farmer choice (OA) or production (GY). In order to detail specific aspects of the dataset, the probability of winning was computed restricting farms to those belonging to any of specific subset (*e.g.* year, environmental distance class, region). Turner et al 2020<sup>21</sup> introduces the Plackett-Luce model and its applications with partial rankings, which was the case in this research. The implementation of Plackett-Luce model to analyse data from decentralised crop variety trials is demonstrated by van Etten et al. 2019<sup>14</sup>.

## Implementation of the genomic selection benchmark

The benchmark representing a centralised breeding system was conducted using genomic selection models and marker-based genetic relationship matrices computed on BLUP data with the package rrBLUP<sup>52</sup>. To measure accuracy of genomic selection predictions, we calculated the Kendall’s tau coefficient ( $\tau$ ), a measure of similarity of rankings<sup>53</sup>, between predicted values and observed values. The use of the  $\tau$  metric, uncommon in breeding, allowed to compare accuracies with the 3D-breeding approach. A Pearson’s correlation, the standard metric for genomic selection accuracy, was also computed but did not show any relevant difference with the Kendall  $\tau$ .

The following genomic selection scenarios were considered: (i) using  $GY_{STATION}$  and  $OA_{STATION}$  to predict  $GY_{STATION}$  measured in the same locations in the following season; (ii) using  $GY_{STATION}$  and  $OA_{STATION}$  to predict  $GY_{FARM}$  and  $OA_{FARM}$ . In the first scenario, the training population was made of all genotypes measured in stations in 2012 and the test population was the subset of 41 genotypes included in the decentralised trials and measured in stations in 2013. In the second scenario, the training population resulted from the combined measures of  $GY_{STATION}$  and  $OA_{STATION}$  across seasons and locations and the test population was the subset of 41 genotypes measured for  $GY_{FARM}$  and  $OA_{FARM}$  independently for each season. In a siding analysis stations were used to predict increasingly different farms based on quantiles of environmental distances according to the distances derived from climatic data. Note that in both scenarios the use of a training sample overlapping the test sample was meant to allow a fair comparison with the 3D-breeding that also uses the entire dataset to train the model.

Mirroring the approach used in the 3D-breeding, the accuracy of genomic selection in the second scenario was derived from a cross-validation approach averaging Kendall  $\tau$  specific for Season 1, Season 2, and Season 3 using the square root of the sample size as weights<sup>54</sup>.

Alternative genomic selection scenarios were also performed but showed consistent accuracy values. Most notably: (i) without overlap between training and test samples, (ii) restricting training and test samples to the subset distributed to decentralised fields, (iii) using rankings derived from BLUP values to train the GY and OA model. In none of the cases above, the benchmark provided accuracies with a noticeable difference from the second scenario reported above and used for the central comparison with 3D-breeding.

## 3D-breeding implementation

The model representing the 3D-breeding approach was built with the data generated by

the citizen science decentralised trials using the Plackett-Luce model. We used two variants of the model, one using  $OA_{FARM}$  and the other using  $GY_{FARM}$  to check which of these metrics had the higher model precision, as shown in Table 1. DNA data from SNPs was added into the model as a prior using an additive matrix. Climatic variables were linked to the rankings using the Model-Based Recursive Partitioning approach<sup>55</sup>. Daily temperature and precipitation data were obtained from the NASA LaRC POWER Project (<https://power.larc.nasa.gov/>), using the R package `nasapower`<sup>56</sup>. A set of climatic covariates were extracted for the vegetative, reproductive and grain filling phases and the whole growth period (from planting date to harvesting as measured on-site) in each observation point using the R package `climatrends`<sup>57</sup>. This resulted in 110 covariates.

To create a model that provides generalizable predictions across seasons with few covariates, we used blocked cross-validation (with seasons as blocks) combined with a forward selection<sup>58</sup>. We used the deviance values of each validation season to calculate an Akaike weight, which is the probability that a given variable combination represents the best model<sup>59</sup>. We performed forward selection, using this combined Akaike weight as our selection criterion. The PLT models had a cut-off value of  $\alpha = 0.01$  and a minimal group size of 20 percent of the total dataset partitioning selection. The variables selected under this procedure were the maximum night temperature ( $^{\circ}\text{C}$ ) during reproductive growth and the minimum night temperature ( $^{\circ}\text{C}$ ) during the vegetative growth. To compare the accuracy of the model representing 3D-breeding with the benchmark, we calculated the Kendall  $\tau$  between observed rankings and predicted coefficients.

### Generalisation of the 3D-breeding

We then evaluated if the model obtained with the variable selection procedure retained predictive power across seasons. We simulated untested future seasonal climate with representative seasonal scenarios of past climate conditions by extracting

the last 15 years of daily climate data derived from NASA POWER (2001-2015). We determined three windows for sowing dates in each growing season as the midpoints of equiprobable quantile intervals estimated from the observed planting dates in the data set. We predicted genotype performance for 15 seasons  $\times$  3 sowing dates (45 seasonal scenarios) for 1,200 random points generated across an alpha hull area within the range of the trials' coordinates. We averaged genotype probability of winning across these scenarios for each planting date interval, excluding the seasons used as testing data.

We calculated the reliability (Fig. 2B), the probability of outperforming a check variety<sup>60</sup>. We used the *worth* parameters from Plackett-Luce to determine the values of positive-valued parameters  $\alpha_i$  associated with each genotype  $i$ , by comparing the *worth* from the check variety (Asassa, Hitosa and Ude, currently recommended for the mega-environment<sup>22</sup>) with the *worth* of the selected genotypes from 3D-breeding. These parameters ( $\alpha$ ) are related to the probability ( $P$ ) that genotype  $i$  wins against all other  $n$  genotypes in set following the Luce's Choice Axiom, which states that the probability that one item beats another is independent from the presence or absence of any other items in the set. Reliability was calculated following Equation 2:

Equation [2]

$$P(i \succ \{j, \dots, n\}) = \frac{\alpha_i}{\alpha_1 + \dots + \alpha_n} = \frac{\alpha_i}{1} = \alpha_i$$

### Environmental characterisation of test sites and genotypes

The agroecological zonation of Ethiopia was obtained by the Ethiopian Institute of Agricultural Research (EIAR)<sup>61</sup>. GPS coordinates of centralised stations and decentralised farmer fields were used to retrieve climatic data from NASA POWER. Temperature indices for covariates used in the PL model were retrieved for the growing seasons object of the study in the time span from sowing date and flowering dates as measured on-

site. Climatic variables considered were the maximum night temperature ( $^{\circ}\text{C}$ ) during reproductive growth and the minimum night temperature ( $^{\circ}\text{C}$ ) during the vegetative growth, which showed to be the most relevant for the sampled data. A principal component analysis (PCA) was used to summarise and depict variation at test sites. Climatic distance of test sites was derived from a multidimensional scaling (MDS) of the multivariate climate dataset. For each of the two stations, climatic distance was computed with all farm sites. Wheat genotypes were split in cold adapted and warm adapted according to the altitude of their original sampling site with a one-tailed, unequal-variance t-test.

## Supporting software

Organising the datasets relied on R packages `data.table`<sup>62</sup>, `caret`<sup>63</sup>, `gosset`<sup>64</sup>, `janitor`<sup>65</sup>, `magrittr`<sup>66</sup> and `tidyverse`<sup>67</sup>. Climatic variables were obtained using the packages `climatrends`<sup>57</sup> and `nasapower`<sup>56</sup>. Statistical analysis was performed using packages `PlackettLuce`<sup>21</sup>, `gosset`<sup>64</sup> and `qvcalc`<sup>68</sup>. Spatial visualisation was performed with the packages `dismo`<sup>69</sup>, `raster`<sup>70</sup>, `sf`<sup>71</sup> and `smoothr`<sup>72</sup>. Charts were produced using packages `corrplot`<sup>73</sup>, `ggplot2`<sup>74</sup> and `patchwork`<sup>75</sup>.

## Data availability

Full data and code are available through `Dataverse`<sup>46</sup>.

## References

- Poland, J. Breeding-assisted genomics. *Current Opinion in Plant Biology* **24**, 119–124 (2015).
- Hickey, J. M., Chiurugwi, T., Mackay, I., Powell, W. & Implementing Genomic Selection in CGIAR Breeding Programs Workshop Participants. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. *Nature Genetics* **49**, 1297–1303 (2017).
- Eshed, Y. & Lippman, Z. B. Revolutions in agriculture chart a course for targeted breeding of old and new crops. *Science* **366**, eaax0025 (2019).
- Hickey, L. T. *et al.* Breeding crops to feed 10 billion. *Nature Biotechnology* **1** (2019). doi:10.1038/s41587-019-0152-9
- Heffner, E. L., Sorrells, M. E. & Jannink, J. L. Genomic selection for crop improvement. *Crop Science* **49**, 1–12 (2009).
- Ceccarelli, S. & Grando, S. Decentralized-participatory plant breeding: An example of demand driven research. *Euphytica* **155**, 349–360 (2007).
- Challinor, A. J., Koehler, A.-K., Ramirez-Villegas, J., Whitfield, S. & Das, B. Current warming will reduce yields unless maize breeding and seed systems adapt immediately. *Nature Climate Change* **6**, 954–958 (2016).
- Tester, M. & Langridge, P. Breeding technologies to increase crop production in a changing world. *Science* **327**, 818–822 (2010).
- Witcombe, J. R., Joshi, A., Joshi, K. D. & Sthapit, B. R. Farmer participatory crop improvement. I. Varietal selection and breeding methods and their impact on biodiversity. *Experimental Agriculture* **32**, 445–460 (1996).
- Rhoades, R. E. & Booth, R. H. Farmer-back-to-farmer: A model for generating acceptable agricultural technology. *Agricultural Administration* **11**, 127–137 (1982).
- Ceccarelli, S. Efficiency of Plant Breeding. *Crop Science* **55**, 87 (2015).
- Ceccarelli, S. & Grando, S. Participatory plant breeding: Who did it, who does it and where? *Experimental Agriculture* **2019**, 1–11 (2019).
- van Eeuwijk, F. A., Cooper, M., DeLacy, I. H., Ceccarelli, S. & Grando, S. Some vocabulary and grammar for the analysis of multi-environment trials, as applied to the analysis of FPB and PPB trials. *Euphytica* **122**, 477–490 (2001).
- van Etten, J. *et al.* Crop variety management for climate adaptation supported by citizen

- science. *Proceedings of the National Academy of Sciences* **116**, 4194–4199 (2019).
15. van Etten, J. *et al.* First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot). *Experimental Agriculture* **55**, 275–296 (2019).
  16. Annicchiarico, P., Russi, L., Romani, M., Pecetti, L. & Nazzicari, N. Farmer-participatory vs. conventional market-oriented breeding of inbred crops using phenotypic and genome-enabled approaches: A pea case study. *Field Crops Research* **232**, 30–39 (2019).
  17. Lowder, S. K., Skoet, J. & Raney, T. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development* **87**, 16–29 (2016).
  18. Mengistu, D. K. *et al.* High-density molecular characterization and association mapping in Ethiopian durum wheat landraces reveals high diversity and potential for wheat breeding. *Plant Biotechnology Journal* **14**, 1800–1812 (2016).
  19. Kidane, Y. G. *et al.* Genome Wide Association Study to Identify the Genetic Base of Smallholder Farmer Preferences of Durum Wheat Traits. *Frontiers in Plant Science* **8**, 1230 (2017).
  20. Mancini, C. *et al.* Joining smallholder farmers’ traditional knowledge with metric traits to select better varieties of Ethiopian wheat. *Scientific Reports* **7**, 9120 (2017).
  21. Turner, H. L., van Etten, J., Firth, D. & Kosmidis, I. Modelling rankings in R: the Plackett-Luce package. *Computational Statistics* (2020). doi:10.1007/s00180-020-00959-3
  22. CIMMYT. Wheat Atlas. Available at: <http://beta.wheatatlas.org/>.
  23. Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**, 1819–1829 (2001).
  24. Hayes, B. J. & Daetwyler, H. D. 1000 Bull Genomes Project to Map Simple and Complex Genetic Traits in Cattle: Applications and Outcomes. *Annual Review of Animal Biosciences* **7**, 89–102 (2019).
  25. Quintero, A., Molero, G., Reynolds, M. P. & Calderini, D. F. Trade-off between grain weight and grain number in wheat depends on G×E interaction: A case study of an elite CIMMYT panel (CIMCOG). *European Journal of Agronomy* **92**, 17–29 (2018).
  26. Ceccarelli, S., Erskine, W., Hamblin, J. & Grando, S. Genotype by environment interaction and international breeding programmes. *Experimental Agriculture* **30**, 177–187 (1994).
  27. Eeuwijk, F. A. van *et al.* Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Science* **282**, 23–39 (2019).
  28. Bauer, P., Thorpe, A. & Brunet, G. The quiet revolution of numerical weather prediction. *Nature* **525**, 47–55 (2015).
  29. Dixon, J. *et al.* Adoption and economic impact of improved wheat varieties in the developing world. **144**, 489–502 (2006).
  30. Jallea, T. Participatory evaluation of the performance of some improved bread wheat (*Triticum aestivum*) varieties in the Jijiga plains of eastern Ethiopia. *Experimental Agriculture* **40**, 89–97 (2004).
  31. Tesfaye, S., Bedada, B. & Mesay, Y. Impact of improved wheat technology adoption on productivity and income in Ethiopia. *African Crop Science Journal* **24**, 127–135 (2016).
  32. Deutsch, C. A. *et al.* Increase in crop losses to insect pests in a warming climate. *Science* **361**, 916–919 (2018).
  33. Carvajal-Yepes, M. *et al.* A global surveillance system for crop diseases. *Science* **364**, 1237–1239 (2019).
  34. Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C. & Foley, J. A. Recent patterns of crop yield growth and stagnation. *Nature Communications* **3**, (2012).

35. Cai, W. *et al.* Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change* **5**, 1–6 (2014).
36. Holmgren, M., Hirota, M., van Nes, E. H. & Scheffer, M. Effects of interannual climate variability on tropical tree cover. *Nature Climate Change* **3**, 755–758 (2013).
37. Ceccarelli, S. *Plant Breeding with Farmers: A technical manual*. 125 (International Center for Agricultural Research in the Dry Areas (ICARDA), 2012). doi:<https://hdl.handle.net/20.500.11766/7745>
38. Lado, B. *et al.* Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theoretical and Applied Genetics* **131**, 2719–2731 (2018).
39. Heffner, E. L., Lorenz, A. J., Jannink, J.-L. & Sorrells, M. E. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* **50**, 1681–1690 (2010).
40. Runcie, D. & Cheng, H. Pitfalls and remedies for cross validation with multi-trait genomic prediction methods. *G3: Genes, Genomes, Genetics* **9**, 3727–3741 (2019).
41. Ibba, M. I. *et al.* Genome-based prediction of multiple wheat quality traits in multiple years. *The Plant Genome* (2020). doi:10.1002/tpg2.20034
42. Mangione, D., Senni, S., Puccioni, M., Grando, S. & Ceccarelli, S. The cost of participatory barley breeding. *Euphytica* **150**, 289–306 (2006).
43. Leonelli, S., Davey, R. P., Arnaud, E., Parry, G. & Bastow, R. Data management and best practice for plant science. *Nature Plants* **3**, 1–4 (2017).
44. van Etten, J. *et al.* ClimMob: Software to support experimental citizen science in agriculture. version 3.1. (2020). Available at: <https://climmob.net>.
45. Kehel, Z., Crossa, J. & Reynolds, M. Identifying Climate Patterns during the Crop-Growing Cycle from 30 Years of CIMMYT Elite Spring Wheat International Yield Trials. in *Applied mathematics and omics to assess crop genetic resources for climate change adaptive traits* (eds. Bari, A., Damania, A. B., Mackay, M. & Dayanandan, S.) 151–174 (CRC Press, 2016).
46. de Sousa, K. *et al.* Replication data for: "Data-driven decentralized breeding increases genetic gain in a challenging crop production environment. (2020). doi:10.7910/DVN/OEZGVP
47. Likert, R. A technique for the measurement of attitudes. *Archives of Psychology* **22**, 55 (1932).
48. R Core Team. R: A language and environment for statistical computing. version 4.0.2. (2020).
49. Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. *ASReml User Guide Release 4.1*. (Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, 2015).
50. Plackett, R. L. The Analysis of Permutations. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **24**, 193–202 (1975).
51. Luce, R. D. *Individual Choice Behavior*. 153 (Courier Corporation, 1959).
52. Endelman, J. B. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *The Plant Genome Journal* **4**, 250 (2011).
53. Kendall, M. G. A new measure of ranking correlation. *Biometrika* **30**, 81–93 (1938).
54. Whitlock, M. C. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology* **18**, 1368–1373 (2005).
55. Zeileis, A., Hothorn, T. & Hornik, K. Model-based recursive partitioning. *Journal of Computational and Graphical Statistics* **17**, 492–514 (2008).
56. Sparks, A. H. nasapower: A NASA POWER Global Meteorology, Surface Solar Energy and

- Climatology Data Client for R. *Journal of Open Source Software* **3**, 1035 (2018).
57. de Sousa, K., van Etten, J. & Solberg, S. Ø. climatrends: Climate Variability Indices for Ecological Modelling. R package version 0.1.6. (2020). Available at: <https://CRAN.R-project.org/package=climatrends>.
  58. Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environmental Modelling and Software* **101**, 1–9 (2018).
  59. Wagenmakers, E. J. & Farrell, S. AIC model selection using Akaike weights. *Psychonomic Bulletin and Review* **11**, 192–196 (2004).
  60. Eskridge, K. M. & Mumm, R. F. Choosing plant cultivars based on the probability of outperforming a check. *Theoretical and Applied Genetics* **84–84**, 494–500 (1992).
  61. Ministry of Agriculture of Ethiopia. *Agro-ecological Zonations of Ethiopia*. (2020).
  62. Dowle, M. & Srinivasan, A. data.table: Extension of ‘data.frame’. R package version 1.12.8. (2019). Available at: <https://CRAN.R-project.org/package=data.table>.
  63. Kuhn, M. caret: Classification and Regression Training. R package version 6.0-85. (2020). Available at: <https://CRAN.R-project.org/package=caret>.
  64. de Sousa, K., van Etten, J., Dumble, S., Greliche, N. & Steinke, J. gosset: Modelling Metadata and Crowdsourced Citizen Science. R package version 0.2.1. Available at: <https://agrobioinfoservices.github.io/gosset/>.
  65. Firke, S. janitor: Simple Tools for Examining and Cleaning Dirty Data. R package version 1.2.1. (2020). Available at: <https://CRAN.R-project.org/package=janitor>.
  66. Bache, S. M. & Wickham, H. magrittr: A Forward-Pipe Operator for R. R package version 1.5. (2014). Available at: <https://CRAN.R-project.org/package=magrittr>.
  67. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
  68. Firth, D. qvcalc: Quasi Variances for Factor Effects in Statistical Models. R package version 1.0.1. (2019). Available at: <https://CRAN.R-project.org/package=qvcalc>.
  69. Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. dismo: Species Distribution Modeling. R package version 1.1-4. (2017). Available at: <https://CRAN.R-project.org/package=dismo>.
  70. Hijmans, R. J. *et al.* raster: Geographic Data Analysis and Modeling. R package version 2.5-8. (2015). Available at: <https://cran.r-project.org/package=raster>.
  71. Pebesma, E. Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* **10**, 439–446 (2018).
  72. Strimas-Mackey, M. smoothr: Smooth and Tidy Spatial Features. R package version 0.1.2. (2020). Available at: <https://CRAN.R-project.org/package=smoothr>.
  73. Wei, T. & Simko, V. *R package "corrplot": Visualization of a correlation matrix*. (2017).
  74. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
  75. Pedersen, T. L. patchwork: The Composer of Plots. R package version 1.0.0. (2019). Available at: <https://CRAN.R-project.org/package=patchwork>.