

UC Berkeley

UC Berkeley Previously Published Works

Title

Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals?

Permalink

<https://escholarship.org/uc/item/8x4342v3>

Authors

Streich, Jared
Romero, Jonathon
Gazolla, João Gabriel Felipe Machado
et al.

Publication Date

2020-02-01

DOI

10.1016/j.copbio.2020.01.010

Peer reviewed

Can exascale computing and explainable artificial intelligence applied to plant biology deliver on the United Nations sustainable development goals?

Jared Streich¹, Jonathon Romero^{1,2},
João Gabriel Felipe Machado Gazolla¹, David Kainer¹,
Ashley Cliff^{1,2}, Erica Teixeira Prates¹, James B Brown³,
Sacha Khoury⁴, Gerald A Tuskan¹, Michael Garvin¹,
Daniel Jacobson^{1,2,5} and Antoine L Harfouche⁶

ABSTRACT

Human population growth and accelerated climate change necessitate agricultural improvements using designer crop ideotypes (idealized plants that can grow in niche environments). Diverse and highly skilled research groups must integrate efforts to bridge the gaps needed to achieve international goals towards sustainable agriculture. Given the scale of global agricultural needs and the breadth of multiple types of omics data needed to optimize these efforts, explainable artificial intelligence (AI with a decipherable decision-making process that provides a meaningful explanation to humans) and exascale computing (computers that can perform 10^{18} floating-point operations per second, or exaflops) are crucial. Accurate phenotyping and daily-resolution climate production in specific environments at various levels of granularity. We review advances toward tackling technological hurdles to solve multiple United Nations Sustainable Development Goals and discuss a vision to overcome gaps between research and policy.

Addresses

¹ DOE Center for Bioenergy Innovation, Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, United States

² Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, TN 37996-3394, United States

³ Molecular Ecosystems Biology Department, Division of Environmental Genomics and Systems Biology, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States

⁴ Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, United Kingdom

⁵ Department of Psychology, University of Tennessee, Knoxville, TN 37996, United States

⁶ Department for Innovation in Biological, Agro-food and Forest Systems, University of Tuscia, Viterbo, VT 01100, Italy

Corresponding authors: Jacobson, Daniel
(jacobsonda@ornl.gov), Harfouche, Antoine L
(aharfouche@unitus.it)

Introduction

In 2015, the international community committed to the United Nations 17 Sustainable Development Goals (SDGs) to end poverty, protect the planet, ensure global prosperity, and eradicate hunger [1]. Agriculture is globally recognized as critical to achieving the 2030 agenda [2,3], but it is also the greatest resource consumer [4,5], which can have substantial negative impacts on biodiversity, water, and global climate [6,7]. The development of novel sustainable crops with an optimal combination of traits that thrive in specific environments should be focused on food (SDG2) or bioenergy (SDG7) production, while conserving water resources (SDG6) and biodiversity (SDG15) like with multi-criteria ideotypes. Large-scale production of optimized cultivars will require technological advances in order to implement and integrate SDGs [8]; while simultaneously minimizing environmental perturbations [9] and combating climate change (SDG13).

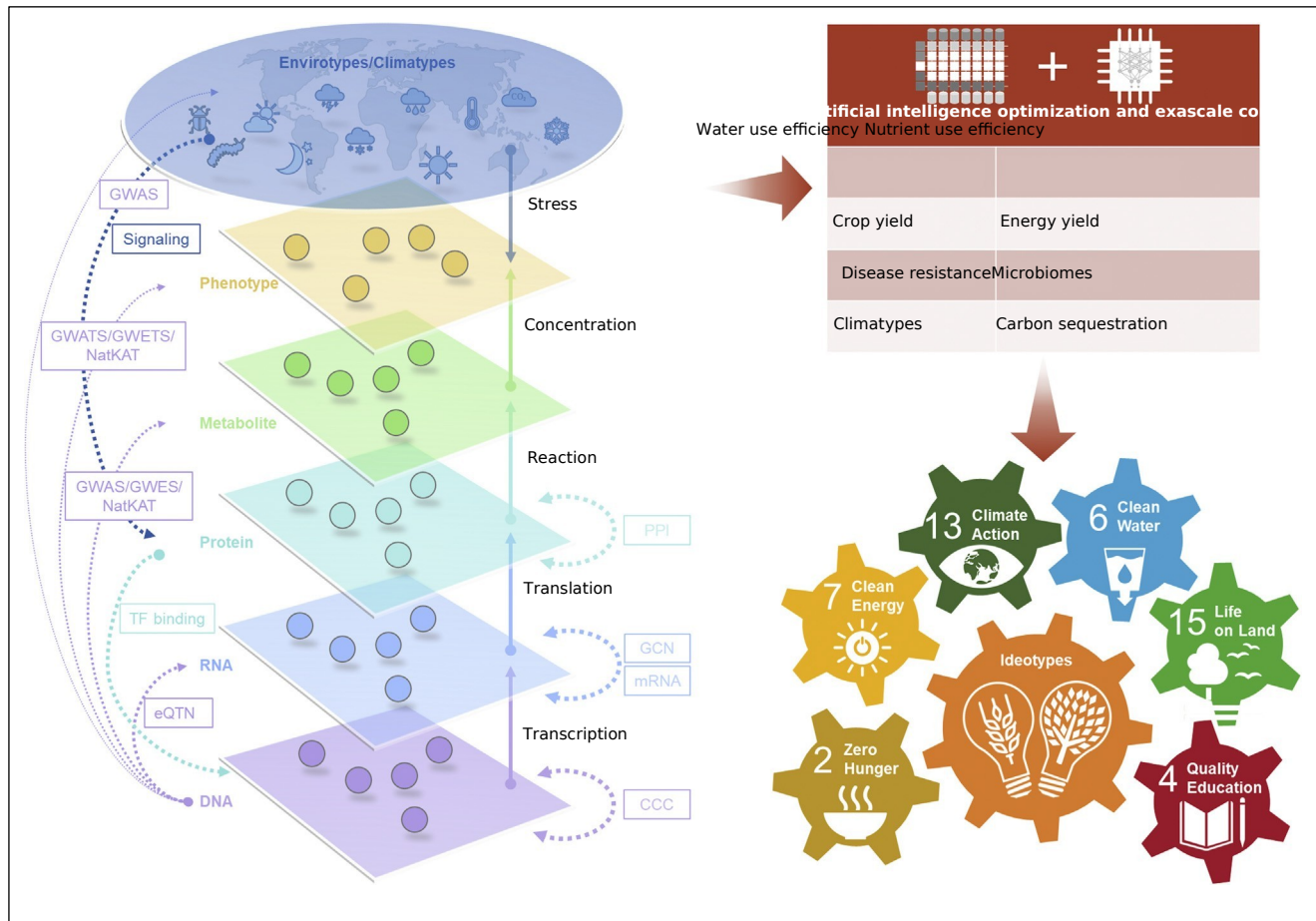
Technological improvement of agricultural production from simple subsistence has been possible through mechanized advances, genomic breeding, and advanced phenotyping. Current efforts routinely integrate multi-omics datasets (genome, epigenome, transcriptome, proteome, metabolome, phytobiome, and phenome) of plant populations to dissect biological mechanisms that underlie

desirable traits. Now we propose a next-generation agricultural revolution that hinges on artificial intelligence (AI), which will be essential for designing new crop ideotypes that are optimized for niche environments in a world with a rapidly changing climate while adhering to the SDGs mentioned above (Figure 1).

AI has existed for decades but its popularity has increased with improvements in high performance computing (HPC). Put simply, AI, including machine learning approaches, utilize a variety of qualitative and quantitative data types, iteratively performing a series of calculations/decisions between vector sets followed by decision/ action(s). AI is routinely used in complex, data-heavy

sectors such as banking, business, and power grid management [10,11], and it can be a powerful tool in reducing greenhouse gas emissions and helping society adapt to a changing climate [12]. However, it is sometimes seen as a negative that many algorithms function as ‘black boxes’, that is, the decision making process is not known, and yet in many cases, the reasoning behind a solution is more informative than the solution itself [13]. This agricultural revolution will require a systems-level approach (from individual plant to global scale) using exascale computing (the performance of a computer system reaches the exaflops level, that is, 10^{18} floating point operations per second), especially when ultra-fast mass brute force is required [14], and explainable AI (X-AI, AI

Figure 1



Contribution of explainable artificial intelligence (X-AI) and exascale computing in the design and deployment of new crop ideotypes in relation to six environmentally and socially oriented Sustainable Development Goals (SDGs 2,4, 6, 7, 13 and 15). X-AI models bring together various omics datasets (including the epigenome, genome, transcriptome, proteome, metabolome, and phenome) of populations to understand their underlying biological mechanisms. A 200-petaflop supercomputer, Summit can perform 200 quadrillion (peta-) 64-bit floating point operations per second (-flops) and 3×10^{18} 16-bit flops for vastly accelerated exascale data analytic and machine learning applications. Interpretable AI predictive models are essential to develop crop ideotypes that will thrive in target environments. These ideotypes can be precisely defined based on specific traits or trait combinations, including, but not limited to, water and nutrient use efficiency, high food or net energy yield per hectare, carbon sequestration, optimized microbiome usage, disease resistance, and so on. CCC, custom correlation coefficient; eQTN, expression quantitative trait nucleotide; GCN, gene co-expression network; GWAS, genome-wide association studies; GWES, genome-wide epistasis studies; GWATS, genome-wide association time-series studies; GWETS, genome-wide epistasis time-series studies; NatKAT, nature's knockouts association test; PPIs, protein-protein interactions; TF, transcription

factor.

with decipherable decision making process) methods to integrate and analyze multi-omics data layers (Figure 1). Educating and training early-career scientists and farmers (SDG4) will ultimately fuel the engines of future X-AI and exascale computing applied to plant biology and lead to the implementation of science-driven policies (Figure 1).

Securing sustainable food and bioenergy systems through plant population-scale multi-omics technologies

Multiple omics technologies such as epigenomics, genomics, transcriptomics, proteomics, metabolomics, and phenomics are aimed primarily at the detection of methylation profiles, genes, mRNA, proteins, metabolites, and phenotypes, respectively. As compared to single omics, multi-omics studies offer the opportunity to understand the flow of information that underlies different types of environmental stresses. Crop yield increase has continuously relied on more detailed and accurate information for genotypes, phenotypes, and their local environments. In recent years the acquisition and implementation of multi-omics data reached unprecedented levels. Short-read sequencing made large populations genome level data easily accessible and economical for germline structural variations (variants), expression (RNA-seq), and epigenetics (DNA-methylation) [15], which is now being integrated with microfluidics to interrogate these genomic processes at the single cell level [16,17].

Proteins are of paramount importance as the functional manifestation of genes; they are directly responsible for observable phenotypic traits, and protein variants can be engineered to modulate a desirable trait (e.g. disease resistance or growth). Advances in protein structure prediction methods along the past decade, combined with increased HPC, enable high-quality proteome-wide 3D-modeling [18]. Integrated systems biology models can subsequently predict protein-protein interactions, which can improve protein docking simulations to further provide structural information of protein complexes [19]. We are developing AI based docking protocols to screen 28 million small molecules against proteins associated with phenotypes (e.g. herbicide susceptibility). The 3D-interactome approach becomes the foundation for active compound discovery in other areas as well [20]. Finally, the components of the plant's immediate environment (e.g. microbiome and ionome) and interactions among them greatly influence desired traits and are rapidly becoming readily accessible sources of informative data [21-23]. Exascale computing facilitates multi-omic layer integration to determine features that underlie traits of interest; a

key to that goal is the addition of data-rich phenotypes — phenomics.

Modernization of greenhouse and field-based high-throughput phenotyping and microphenotyping across large scales has been essential to capture the diverse traits needed to

complement multi-omic AI predictions [24]. Various image processing modalities can non-invasively phenotype developmental and physiological phenotypes using broad light spectra [25,26]. For example, plant health and stress responses are quantifiable via red-green-blue (RGB) camera or true-color camera that can detect asignal and deliver it into three color bands: red, green, and blue. A genome-wide association study of drought tolerance of 507 rice accessions found 470 variants associated to 51 image-based traits captured through non-destructive RGB phenotyping [27]. Hyperspectral cameras detect hundreds of spectral bands with nm-level resolution between 350 and 2500 nm and can even detail: nitrogen content, physiology, water content, and biochemical traits from environmental stimuli and stress factor; for example, hyperspectral imaging data were used to detect maize plant response to direct insect herbivory as well as plant signaling between infested and non-infested individuals [28]. Thermal infrared (TIR) cameras can detect long-wave infrared radiation that is emitted by plant leaves based on temperature; for example, TIR imaging enabled effective assessment of genotype variability under drought stress conditions via leaf temperature and stomatal conductance in poplar breeding populations [29]. New phenotyping methods have been developed for profiling the root phenotype (i.e. its structure and function), such as 3D-MRI, X-ray, tomography technology, and so on, which have a great potential in breeding for root traits [30]. Additionally, plant photosynthetic performance can be estimated via kinetic chlorophyll fluorescence imaging. Aerial drones can automate phenotype imaging at individual or single-leaf resolution, and they are greenhouse-amenable and field-amenable. All these advances can play an important role in empowering plant biologists with accurate high-dimensional datasets to optimize the operational efficiencies of breeding programs. For example, in soybean, fusing machine learning-based analytics and optimization tools with phenomics data, complex relationships between phenomic traits and seed yield were mapped and their performance predicted [31].

Muti-scale phenomics provides enormous datasets and deep learning/feature extraction can capture specified and cryptic traits [32]. Current image analysis approaches implement segmentation (plant/background separation) and agglomerative whole plant summaries within images [33]. Therefore, substantial amounts of high-resolution, component-specific (e.g. leaves versus stalk) data are often not utilized. Development of AI-driven spatial feature extraction methods will create an unprecedented opportunity to integrate multi-omics results with data collected by automated phenotyping [34] and will underpin efforts to

develop food and bioenergy crop ideotypes with improved climate resilience.

Environment genotype association and clustering the planet's climatotypes

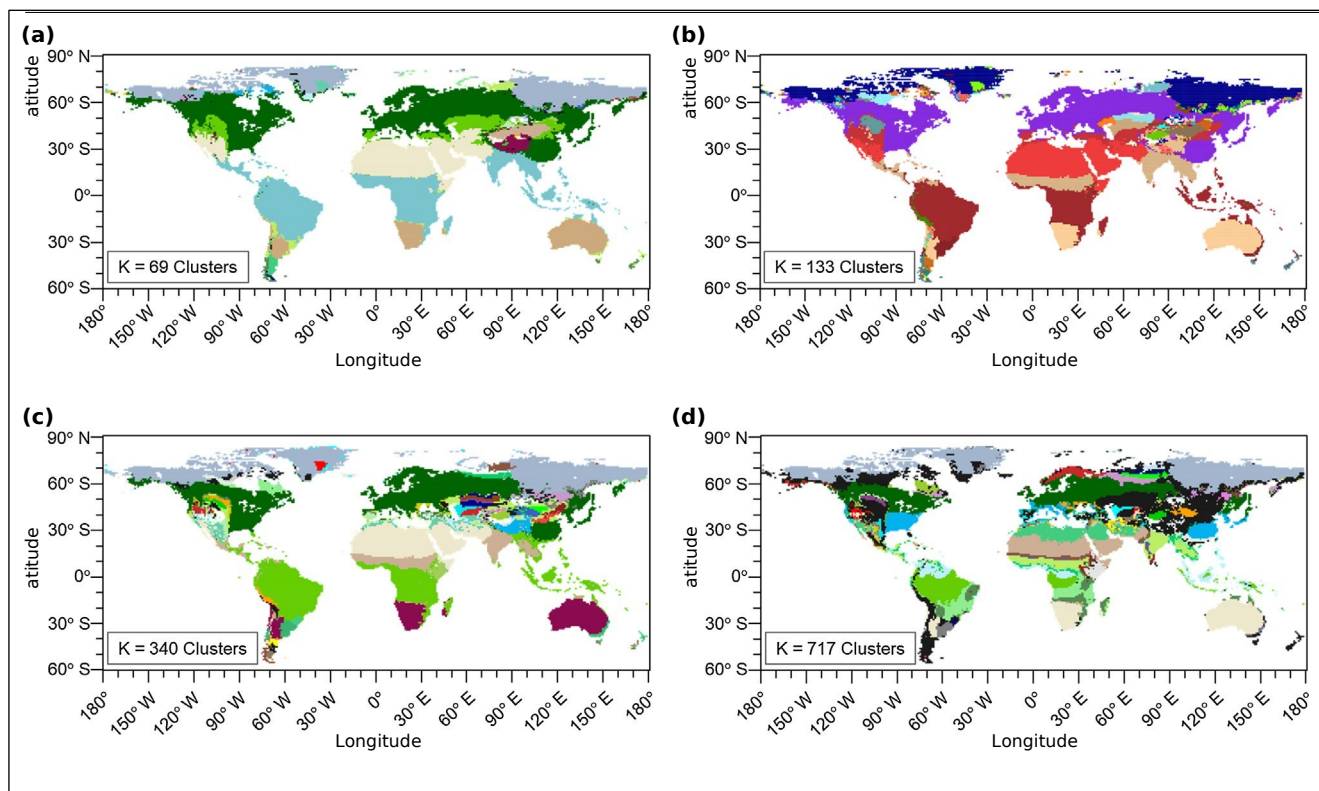
Gene by environment association (G E or GEA) the interaction between genetic variation (heritability) and

the environment to produce a given phenotype is important in crop breeding because the performance of a genotype can vary widely across environments, even over small scales [35–38]. Estimation of G × E is often done by simply rank-ordering genotypes in different environments or by calculating the spread of the expression of a target breeding value across environments [39]. The ability to rapidly generate high-resolution genome-level and phenome-level data under diverse environmental conditions will allow for more accurate and high-resolution G × E estimates to deploy crops that perform optimally across a broader set of environments.

Currently, G × E estimation methods lack a true temporal component, thus, we developed genome-wide association time-series studies (GWATS) and periodic GEA (PGEA) to include daily resolution climate data from specific geographic locations for genomic associations (climatypes) (Streich *et al.*, unpublished). Historically, landscapes were classified by local plant morphology, and more recently with climate classes [40]. The current gold standard Köppen-Geiger system uses annual mean temperature and precipitation data with anthropocentric

thresholds that define 31 climate classes. More recent efforts to cluster geography have focused on supervised machine learning algorithms with low resolution data or isolated regions [41]. However, an abundance of global temporal climate data exist for solar radiation, precipitation, temperature, and numerous other variables [42] that could vastly improve climatype definitions. Toward this end, we created global exascale datasets for 12 major elemental layers for soil and 48 light spectra (300 nm–780 nm) across 365 days and calculated similarity indexes using the DUO algorithm [22] on the Summit supercomputer. This generated climate clusters globally at 1 km² resolution (Figure 2). With these data, the GWATS approach depicts spatio-temporal adaptation of candidate alleles, which can identify genotypes (cultivars, varieties) that optimally express a target phenotype at environmental granularities needed for sustainable, high-yield crop breeding. Researchers are currently using AI to find the most important multi-omic interactions and optimize outcomes [43, 44], but the addition of the complex environmental data presented here and new omics layers will result in an enormous feature-interaction space that will require exascale computing and newer AI approaches.

Figure 2



Global climatype clustering. Examples of corresponding results and significant findings applying Markov clustering (MCL) algorithm to a network derived from the DUO similarity metric comparison of the climatype vectors with the use of increasing inflation values to give clusters of increasing granularities (a–d). In order to determine what environments are present around the world at very high resolution 414 640 element tensors representing the environmental and climate conditions of each of the 157 million square kilometers of land on earth were created. By comparing all of these tensors to one another we have created a network that represents the

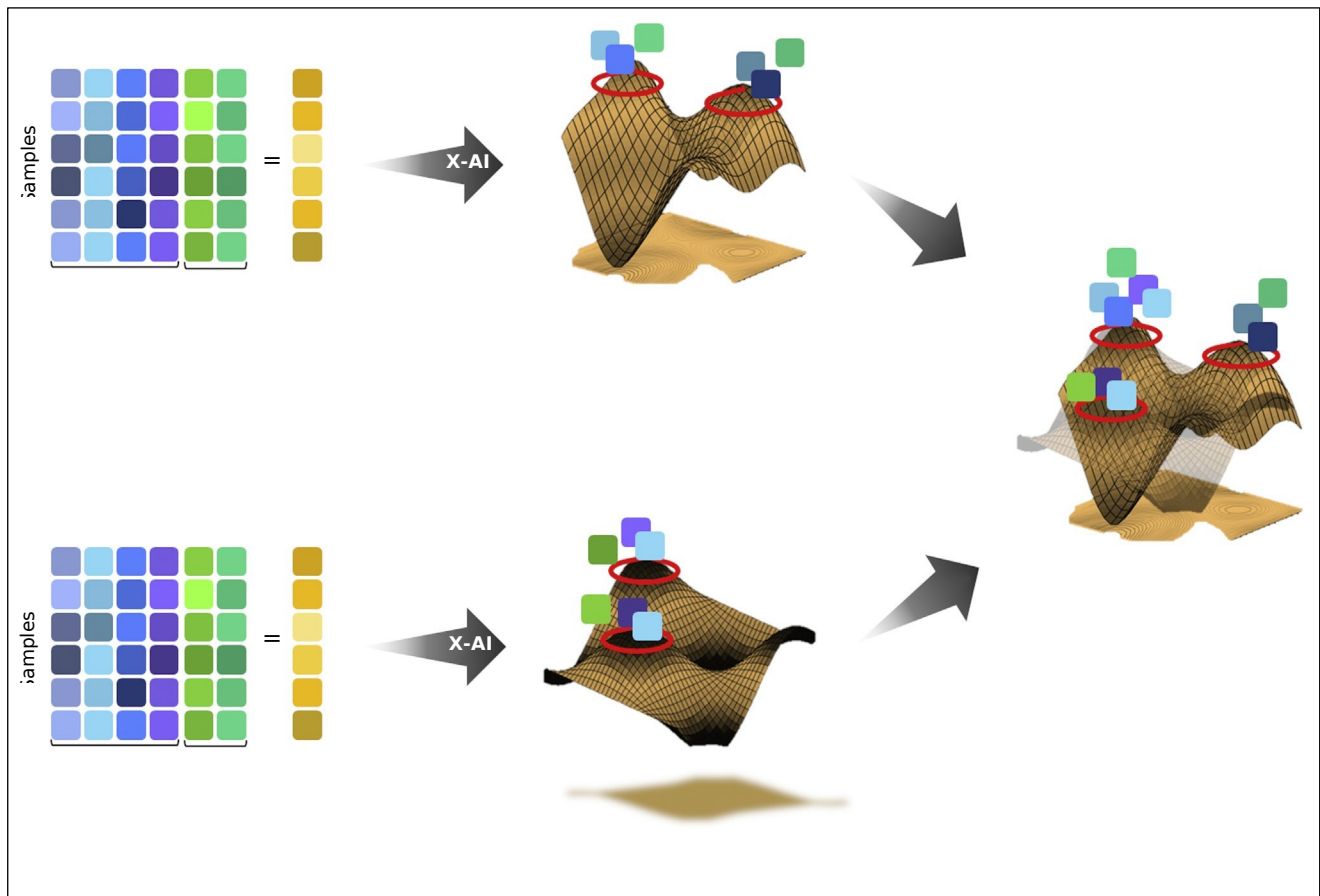
similarity of all points on land to one another.

Explainable AI-assisted crop ideotype design for food security and bioenergy sustainability Multi-trait genomic selection is relatively a recent development and the use of Bayesian methods have been the most accurate, yet slow; recent AI methods have proven to be just as accurate with an increased speed giving them an edge over Bayesian models [45]. We draw on examples from machine-learning algorithms for scientific insights to show that newly developed X-AI algorithms [46] such as iterative Random Forest (iRF) can help identify phenotype-conveying variants using additive datasets, given genotype specific traits (Figure 3). X-AI can identify epistatic (the phenotypic effect of one gene can vary depending on the modifier gene) and pleiotropic (a single gene affects multiple phenotypic traits) relationships in dense variant sets while accounting for non-additive effects. The output accounts for all possible genome variants within the scope of targeted phenotypes such as yield or disease resistance, as well as climatotypes (Figure 3). We believe X-AI can discover

phenotypically superior high-yielding breeding pair combinations with other beneficial traits for consumption, energy production, and geotargeted locations even on marginal land that is of poor quality and likely to be drought-prone, thus meeting both sustainability and agricultural needs.

Rethinking science-driven policy to enable ideotype deployment at scale
 Here we describe a novel three thirds approach for effective science-policy engagement to scale new crop ideotype design and sustainable deployment (Figure 4). The three third influences - entitled 'aligning research with policy and practice', 'encouraging evidence-based policy making', and 'promoting science communication, public engagement and outreach' are interconnected, thus it is important to understand how these factors fit together, influence one another, and as such, none can be viewed in isolation.

Figure 3



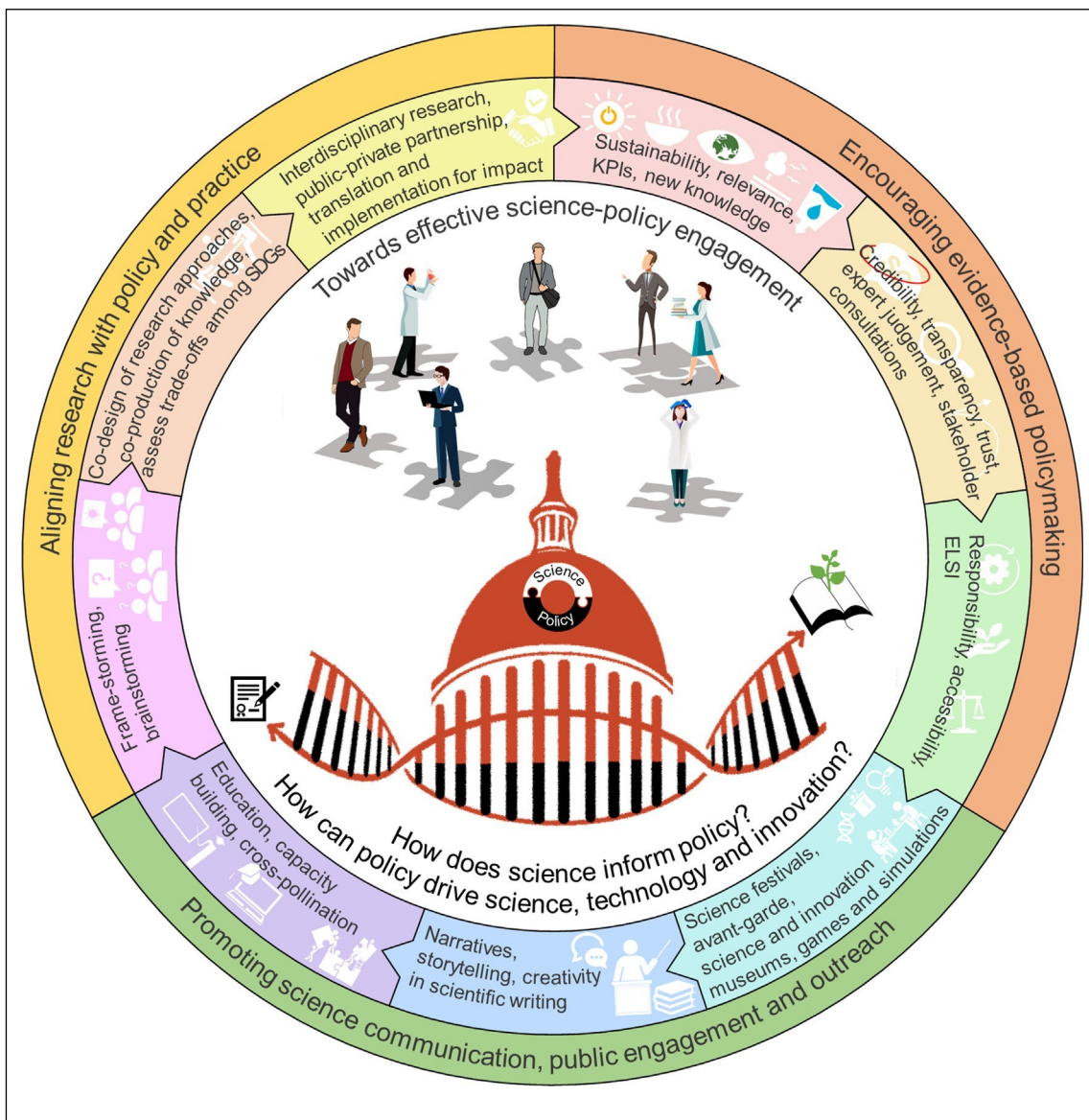
An illustration of how explainable artificial intelligence (X-AI)-based genomic selection can help with multi-criteria ideotype design. X-AI does not require traditional mathematics to combine datasets and this remains true for genomic selection with multi-criteria ideotype design. X-AI analyses can make smart interpretable decisions to link genotype to phenotype and genotype to climatotype relationships

showing the additive, epistatic, and pleiotropic combinations of specific alleles that lead to optimal crop phenotypes and climatype adaptation. SNPs, single nucleotide polymorphisms.

The first third is critical to achieve SDGs and will require the integrated efforts of researchers, policy-makers, breeders, landowners, and relevant stakeholders. This consortium can facilitate interdisciplinary research, co-design future projects and data analyses, encourage co-production of knowledge, evaluate sustainability trade-offs, and assess interactions among SDGs. Implementing innovations for agriculture-driven SDGs requires constructive participation by public and private sectors, non-governmental organizations, and philanthropic institutions, while meeting each organization's individual priorities.

The desired goal of the second third is to integrate scientific evidence in policy/decision-making processes for SDG success and future global food security. We believe enhancing credibility, responsibility, and transparency through science-policy engagement is a priority to overcome knowledge gaps and allow scientific advances to proceed ethically and with greater public confidence. Public trust is a crucial component in science-policy dynamics that positively impacts future science, policy, and society. The potential of scientific discoveries will not be fully realized until ethical, legal and social issues (ELSI) [47] are addressed by effective science

Figure 4



Strengthening the interface between science and policy for food and bioenergy sustainable development. A novel 'three thirds' approach for effective science-policy engagement is needed in order to scale new crop ideotype design and sustainable deployment. The three are interconnected and it is therefore important to understand how they fit together and influence one another, rather than viewing them in isolation.

policy and communication. Scientists, policymakers and other stakeholders must redouble their efforts to extract more from ongoing research and innovation in bioenergy and food sustainability by measuring what matters (e.g. key performance indicators [KPIs]) and forecasting their impacts.

The final third is to increase public knowledge and participation in science-related policymaking that will favorably impact the value of science to society. The best start for science students and early-career scientists wanting to increase their advocacy beyond the bench is getting involved in policy through education and capacity building. Sharing knowledge is imperative for the cross-pollination of ideas and practices that influence students; science communication is an integral part of doing research. Storytelling and narratives [48,49], as well as creativity in scientific writing [50,51], can help connect science to non-experts and present engaging discoveries to the public. Museums of science and innovation, science festivals, games, simulations, and avant-garde outreach [52] can also boost public understanding that will ultimately increase public receptiveness to science.

Conclusions

In the face of a changing climate it will be necessary to design food and bioenergy crop ideotypes that thrive in future environments at microscale levels. An AI-driven paradigm shift is significantly transforming plant omics and breeding research. At present, large amounts of multi-omics, imaging, ecophysiology, and field-based data are becoming available for large-scale population-level studies, revealing subtle differences in plants' genetic-based adaptation potential and allowing design interventions. Advanced AI approaches can be used to model climatype patterns and clustering across the last 50 years against the backdrop of carbon dioxide levels in an effort to predict future patterns. This will be used to predict future climatype patterns that can be targets for new ideotypes to optimize agricultural productivity and sustainability. Rapid progress in these areas will be indispensable for future innovation in plant biotechnology and AI-driven crop ideotypes design. It is critical that researchers and policymakers integrate and promote their efforts to deliver SDG solutions [8] while efficiently and clearly communicating these to society.

Funding

Completion of this manuscript was supported, in part, by funding from the EU 7th Framework Programme - Development of improved perennial non-food biomass and bioproduct crops for water stressed environments WATBIO, grant

no 311929; the Center for Bioenergy Innovation, a U.S. Department of Energy DOE Bioenergy Research Center, the Plant Microbe Interface Scientific Focus Area and the Feedstock Genomics program, all three supported by the Biological and

Environmental Research in the DOE Office of Science; and the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. Funding was also provided by the DOE, Laboratory Directed Research and Development funding ORNL AI Initiative ProjectID 9613 at the Oak Ridge National Laboratory.

Surveying the environmental footprint of urban food consumption. *J Ind Ecol* 2017, 21:151-165.

Conflict of interest statement

Nothing declared.

CRedit authorship contribution

statement Jared Streich: Writing - original draft, Methodology. Jonathan Romero: Writing - original draft. João Gabriel Felipe Machado Gazolla: Writing - original draft. David Kainer: Writing - original draft. Ashley Cliff: Writing - original draft. Erica Teixeira Prates: Writing - original draft. James B Brown: Methodology, Software, Writing - original draft. Sacha Khoury: Writing - original draft, Writing - review & editing. Gerald A Tuskan: Conceptualization, Writing - original draft. Michael Garvin: Writing - original draft. Daniel Jacobson: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition, Resources, Methodology, Software. Antoine L Harfouche: Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- .. of outstanding interest

1. UN General Assembly: *Transforming our world: The 2030 agenda for sustainable development A/RES/70/1*. 2015.
2. DeClerck FAJ, Jones SK, Attwood S, Bossio D, Girvetz E, Chaplin-Kramer B, Enfors E, Fremier AK, Gordon LJ, Kizito F et al.: Agricultural ecosystems and their services: the vanguard of sustainability? *Curr Opin Environ Sustain* 2016, 23:92-99.
3. Rasmussen LV, Coolsaet B, Martin A, Mertz O, Pascual U, Corbera E, Dawson N, Fisher JA, Franks P, Ryan CM: Social-ecological outcomes of agricultural intensification. *Nat Sustain* 2018, 1:275-282.
4. Tukker A, Bulavskaya T, Giljum S, de Koning A, Lutter S, Simas M, Stadler K, Wood R: Environmental and resource footprints in a global context: Europe's structural deficit in resource endowments. *Global Environ Change* 2016, 40:171-181.
5. Velazquez JA, Negrutiu I: Agriculture and global physicochemical deregulation: planetary boundaries that challenge planetary health. *Lancet Planet Health* 2019, 3:e10- e11.
6. Tanentzap AJ, Lamb A, Walker S, Farmer A: Resolving conflicts between agriculture and the natural environment. *PLoS Biol* 2015, 13:e1002242.
7. Goldstein B, Birkved M, Fernandez J, Hauschild M:

8. Scown MW, Winkler KJ, Nicholas KA: Aligning research with policy and practice for sustainable agricultural land systems in Europe. *Proc Natl Acad Sci U S A* 2019, 116:4911-4916

The authors describe how they have assessed differences among perspectives in research, policy, and practice to balance and implement sustainability goals for sectors and regions. This method can be a starting point for designing future research to more comprehensively analyze trade-offs and identify opportunities for achieving the UN SDGs.

9. Fujimori S, Hasegawa T, Rogelj J, Su X, Havlik P, Krey V, Takahashi K, Riahi K: Inclusive climate change mitigation and food security policy under 1.5 C climate goal. *Environ Res Lett* 2018, 13:074033.
10. Davenport TH: From analytics to artificial intelligence. *J Bus Analytics* 2018, 1:73-80.
11. Tang Y, State Key Laboratory of Power Grid Safety and Energy Conservation, China Electric Power Research Institute, Huang Y, Wang H, Wang C, Guo Q, Yao W, State Key Laboratory of Power Grid Safety and Energy Conservation, China Electric Power Research Institute et al.: Framework for artificial intelligence analysis in large-scale power grids based on digital simulation. *CSEE J Power Energy* 2018, 4:459-468.
12. Rolnick D, Donti PL, Kaack LH, Kochanski K, Lacoste A, Sankaran K, Ross AS, Milojevic-Dupont N, Jaques N, Waldman-Brown A et al.: Tackling climate change with machine learning. *arXiv* 2019. 190605433 [csCY].
13. Samek W, Müller K-R: Towards explainable artificial intelligence. In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Lecture Notes in Computer Science*. Edited by Samek W, Montavon G, Vedaldi A, Hansen L, Müller KR. Springer; 2019:5-22

The authors present recent developments and applications of explainable

artificial intelligence methods for visualizing, explaining and interpreting deep learning models and recommend a wider use of explainable learning algorithms in practice.

14. Hey T, Trefethen A: The fourth paradigm 10 years on. *Informatik Spektrum* 2020, 42:441-447 <http://dx.doi.org/10.1007/s00287-019-01215-9>.
15. Rasheed A, Hao Y, Xia X, Khan A, Xu Y, Varshney RK, He Z: Crop breeding chips and genotyping platforms: progress, challenges, and perspectives. *Mol Plant* 2017, 10:1047-1064.
16. Ryu KH, Huang L, Kang HM, Schiefelbein J: Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol* 2019, 179:1444-1456

The authors present the first use of a droplet-based microfluidics platform

for high-throughput scRNA-seq to obtain single-cell transcriptomes from protoplasts of more than 10 000 Arabidopsis root cells. This approach provides novel opportunities for mapping gene expression in plants at single-cell resolution.

17. Leonavicius K, Nainys J, Kuciauskas D, Mazutis L: Multi-omics at single-cell resolution: comparison of experimental and data fusion approaches. *Curr Opin Biotechnol* 2019, 55:159-166.
18. Lam SD, Das S, Sillitoe I, Orengo C: An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. *Acta Crystallogr D Struct Biol* 2017, 73:628-640.
19. Dong S, Lau V, Song R, Ierullo M, Esteban E, Wu Y, Sivieng T, Nahal H, Gaudinier A, Pasha A et al.: Proteome-wide, structure-based prediction of protein-protein interactions/ new molecular interactions viewer. *Plant Physiol* 2019, 179:1893-1907.
20. Piazza I, Kochanowski K, Cappelletti V, Fuhrer T, Noor E, Sauer U, Picotti P: A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* 2018, 172:358-372.e23.
21. Hu L, Robert CAM, Cadot S, Zhang X, Ye M, Li B, Manzo D, Chervet N, Steinger T, van der Heijden MGA et al.: Root exudate metabolites drive plant-soil feedbacks on

progressive drought and cyclic drought. *Phytobiomes J* 2018, 2:249-260.

23. Huang X-Y, Salt DE: Plant ionomics: from elemental profiling to environmental adaptation. *Mol Plant* 2016, 9:787-797.
24. Tardieu F, Cabrera-Bosquet L, Pridmore T, Bennett M: Plant phenomics, from sensors to knowledge. *Curr Biol* 2017, 27: R770-R783.
25. Holman F, Riche A, Michalski A, Castle M, Wooster M, Hawkesford M: High throughput field phenotyping of wheat plant height and growth rate in field plot trials using UAV based remote sensing. *Remote Sens* 2016, 8:1031.
26. Walter A, Liebisch F, Hund A: Plant phenotyping: from bean weighing to image analysis. *Plant Methods* 2015, 11:14.
27. Guo Z, Yang W, Chang Y, Ma X, Tu H, Xiong F, Jiang N, Feng H, Huang C, Yang P et al.: Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. *Mol Plant* 2018, 11:789-805.
28. do Prado Ribeiro L, Klock ALS, Filho JAW, Tramontin MA, Trapp MA, Mithofer A, Nansen C: Hyperspectral imaging to characterize plant-plant communication in response to insect herbivory. *Plant Methods* 2018, 14:54.
29. Ludovisi R, Tauro F, Salvati R, Khoury S, Mugnozza Scarascia G, Harfouche A: UAV-based thermal imaging for high-throughput field phenotyping of black poplar response to drought. *Front Plant Sci* 2017, 8:1681.
30. Atkinson JA, Pound MP, Bennett MJ, Wells DM: Uncovering the hidden half of plants using new advances in root phenotyping. *Curr Opin Biotechnol* 2019, 55:1-8
- The authors describe how innovations in imaging and sensor technologies are advancing root phenomics. Coupling these innovations with artificial intelligence-driven data analytics has a major impact to drive the selection of the next-generation of crops to deliver real world solutions for ongoing global food security efforts.
31. Parmley K, Nagasubramanian K, Sarkar S, Ganapathysubramanian B, Singh AK: Development of optimized phenomic predictors for efficient plant breeding decisions using phenomic-assisted selection in soybean. *Plant Phenomics* 2019, 2019:5809404.
32. Sharma N, Jain V, Mishra A: An analysis of convolutional neural networks for image classification. *Procedia Comput Sci* 2018, 132:377-384.
- growth and defense by shaping the rhizosphere microbiota. *Nat Commun* 2018, 9:2738.
22. Garcia BJ, Labbe JL, Jones P, Abraham PE, Hodge I, Climer S, Jawdy S, Gunter L, Tuskan GA, Yang X et al.: Phytobiome and transcriptional adaptation of *Populus deltoides* to acute

33. Williams J, Schonlieb C-B, Swinfield T, Lee J, Cai X, Qie L, Coomes DA: 3D segmentation of trees through a flexible multiclass graph cut algorithm. *IEEE Trans Geosci Remote Sens* 2020, 58:754-776.
34. Bolger AM, Poorter H, Dumschott K, Bolger ME, Arend D, Osorio S, Gundlach H, Mayer KFX, Lange M, Scholz U *et al.*: Computational aspects underlying genome to phenome analysis in plants. *Plant J* 2019, 97:182-198.
35. Juliana P, Poland J, Huerta-Espino J, Shrestha S, Crossa J, Crespo-Herrera L, Toledo FH, Govindan V, Mondal S, Kumar U *et al.*: Improving grain yield, stress resilience and quality of bread wheat using large-scale genomics. *Nat Genet* 2019, 51:1530-1539.
36. Wallace JG, Rodgers-Melnick E, Buckler ES: On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. *Annu Rev Genet* 2018, 52:421-444.
37. Harfouche A, Meilan R, Kirst M, Morgante M, Boerjan W, Sabatti M, Mugnozza GS: Accelerating the domestication of forest trees in a changing world. *Trends Plant Sci* 2012, 17:64-72.
38. Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burguénío J, González-Camacho JM, Pérez-Elizalde S, Beyene Y *et al.*: Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci* 2017, 22:961-975.
39. Li Y, Suontama M, Burdon RD, Dungey HS: Genotype by environment interactions in forest tree breeding: review of

methodology and perspectives on research and application.
Tree Genet Genomes 2017, 13:60.

40. Belda M, Holtanova E, Halenka T, Kalvova J: Climate classification revisited: from Köppen to Trewartha. *Clim Res* 2014, 59:1-13.
41. Kumar J, Mills RT, Hoffman FM, Hargrove WW: Parallel k-means clustering for quantitative ecoregion delineation using large data sets. *Procedia Comput Sci* 2011, 4:1602-1611.
42. Fick SE, Hijmans RJ: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 2017, 37:4302-4315.
43. van Eeuwijk FA, Bustos-Korts D, Millet EJ, Boer MP, Kruijer W, Thompson A, Malosetti M, Iwata H, Quiroz R, Kuppe C *et al.*: Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding. *Plant Sci* 2019, 282:23-39.
44. Xu Y, Xu C, Xu S: Prediction and association mapping of agronomic traits in maize using multiple omic data. *Heredity* 2017, 119:174-184.
45. Montesinos-Lopez OA, Montesinos-Lopez A, Crossa J, Gianola D, Hernandez-Suarez CM, Martin-Vallejo J: Multi-trait, multi-environment deep learning modeling for genomic-enabled prediction of plant traits. *G3* 2018, 8:3829-3840.
46. Basu S, Kumbier K, Brown JB, Yu B: Iterative random forests to discover predictive and stable high-order interactions. *Proc Natl Acad Sci U S A* 2018, 115:1943-1948.
47. Greenbaum D: Expanding ELSI to all areas of innovative science and technology. *Nat Biotechnol* 2015, 33:425-426.
48. Davidson B: Storytelling and evidence-based policy: lessons from the grey literature. *Palgrave Commun* 2017, 3:17093.
49. Martinez-Conde S, Macknik SL: Opinion: finding the plot in science storytelling in hopes of enhancing science communication. *Proc Natl Acad Sci U S A* 2017, 114:8127-8129.
50. Freeling B, Doubleday ZA, Connell SD: Opinion: how can we boost the impact of publications? Try better writing. *Proc Natl Acad Sci U S A* 2019, 116:341-343.
51. Oliver K, Cairney P: The dos and don'ts of influencing policy: a systematic review of advice to academics. *Palgrave Commun* 2019, 5:1-11.

The authors present genotype-to-phenotype models for the integration of genomic and new phenotyping information, as well as a detailed evaluation of the utility of phenotyping information in the context of breeding programs.

The authors provide important tangible experiences and recommendations of how academics can effectively engage with policymakers and create impact on policy.

52. Dance A: Science and culture: avant-garde outreach, with science rigor. *Proc Natl Acad Sci U S A* 2016, 113:11982-11983.