

# Temporal field phenomics allows discovery of nature AND nurture, so can we saturate the phenome?

Seth C Murray<sup>1</sup>, Alper Adak<sup>1</sup>, Aaron Desalvio<sup>1</sup>, and Holly Lane<sup>1,2</sup>

<sup>1</sup>Texas A&M University

<sup>2</sup>Washington State University

November 4, 2022



# Temporal field phenomics allows discovery of nature AND nurture, so can we saturate the phenome?

Seth C. Murray<sup>1</sup>, Alper Adak<sup>1</sup>, Aaron DeSalvio<sup>1</sup>, Holly Lane<sup>1,2</sup>

<sup>1</sup>Texas A&M University, College Station, TX, U.S.A.

<sup>2</sup>Washington State University, Pullman, WA, U.S.A

ORCID: 0000-0002-2960-8226

## ABSTRACT

An organism's phenome results from expression of its genome (nature) under certain environment and management effects (nurture) and interactions between these factors, as well as measurement error. For over 30 years, DNA sequencing and genomics tools advanced to where it's now feasible to saturate genomes of segregating individuals, such that polymorphisms at nearly any position can be determined from other known positions. This is due to structure, linkage disequilibrium (LD), or linkage and is a powerful tool for genomic prediction and investigating biological phenomena. In contrast, most phenomics to date focuses on automating previously known "traits" as measurable and interpretable phenotypes; akin to focusing on measuring a single DNA marker rather than measuring an entire saturated genome. Viewing phenomics as a platform for discovery, similar to genomics, opens new methods for capturing phenomena in nature and nurture. Saturating a phenome would mean that an individual's fitness, performance, responses to environment and/or specific phenotypes could be accurately predicted in untested environments. To date, our experience with phenomic prediction for cumulative, complex phenotypes such as grain yield suggests it's possible to predict organismal performance in untested environments, possibly better than genomic methods despite less advanced tools and data. Factors limiting to saturating a phenome are evaluating enough individuals and environments, but more importantly, tools and methods to extract or "sequence" more phenomic features. Successfully saturating phenomes will impact every aspect of science and society, in biological disciplines from germplasm curators, physiologists to breeders, to education, the courtroom and policy.

**Keywords:** phenomics, temporal, phenomic prediction, UAS, UAV, drone

## 1. INTRODUCTION

A grand challenge is an impending threat, often existential, that a community must address. The closest thing to a grand challenge yet discussed in phenomics is connecting an organism's genotype or genome, to their phenotype and their phenome<sup>1-5</sup>. While genomics has had nearly half a century of research, organismal phenomes, going beyond measuring a few targeted phenotypes, remain a relatively new concept<sup>6</sup>. In the plant research community, this links to applications for societal grand challenges, such as sustainable food security under climate change. A moonshot, in contrast, is a large and audacious project. Two notable moonshots in genomics have been to sequence the human genome, which succeeded beyond measure despite some bumps<sup>7,8</sup>, and the Arabidopsis 2010 project<sup>9</sup>, to determine the function of all genes in Arabidopsis by 2010, which proved overly-ambitious and infeasible. Regardless of success, moonshots can serve as guideposts for a community of researchers to work towards, discovering new barriers along the way. One clear moonshot for the phenomics community is to saturate the phenome. Saturating a phenome of an individual or population would mean that no new measurements of an organism can be made that cannot already be predicted from other measures. In other words, the depth of physiological responses of the organism across any possible environment, management, or scenario (within reason) can be reduced to numbers and predicted at any point in time. This is an audacious goal and requires predicting not only

nature (the genome), but nurture (the response of an organism to environment and management conditions) as well as interactions between the two.

In many ways, the phenomics community is already indirectly working towards saturating phenomes. Controlled environment research seeks to create specific conditions to accurately assess organismal responses under varied conditions throughout time<sup>10</sup>. Computer models are developing synthetic plants, *in silico*, to mimic actual plant growth and responses, with goals of prediction, AI training, and understanding gaps in physiological knowledge<sup>11,12</sup>. New software packages can easily create dozens or hundreds of measurements based on imaging tools for field, root, or controlled environments<sup>13-16</sup>. These disparate activities are all integrable into a larger goal of saturating the phenome. Controlled environments reduce environmental, management, and error noise in phenomic measurements and add a level of precision to understanding the phenome. Controlled environments also have the potential capacity to collect more measurements within environments through fully automated systems that are impractical under field conditions<sup>17-18</sup>. However, highly controlled systems cannot screen the number or diversity of environments or the number of individuals needed to saturate phenomes, so this paper focuses primarily on field systems.

### **1.1 Six major approaches to phenomics and high throughput phenotyping**

Goals and use of high-throughput field phenotyping and phenomics across environments, specifically field phenomics with drones or rovers, can be grouped into six major approaches. The first, and most common, is automating existing measurements such as plant height, disease lesions, lodging, and plant population counts<sup>19-23</sup>. While this helps scale traditional biological knowledge to more genotypes, it is like envisioning high performance computing just to complete thousands of peoples taxes. A second approach is making new measurements that were previously infeasible or impossible manually, for example plant growth over time or novel spectral signatures<sup>24,25</sup>. This approach hypothesizes that there are valuable traits that have not previously been measured, inviting new insights into biology that could never be gained without novel phenomic tools. A third is phenomic selection, viewing phenomics as a platform to quantify and characterize biology, requiring a complete shift in philosophy of how to approach these phenomics tools to maximize measurements. This is a more difficult concept, but one that motivates saturating phenomes. A fourth is the use of deep learning (often called artificial intelligence) to find patterns directly from images<sup>26,27</sup>. The fifth and sixth are applications and use cases of the first four with high interest among plant and crop scientists. Fifth is the discovery of new physiology, biology and genetics to help science and scientists advance basic knowledge in these areas, for which new predictive phenotypes are valuable. Sixth is discovering new traits or signatures for intervention that can be deployed on farm. For instance, detecting disease early to deploy pesticides<sup>28,29</sup>, or drought signatures for irrigation before yield is lost<sup>30-32</sup>.

### **2.1 Phenomic selection**

Phenomic selection was first coined, proposed and explored by Rincent et al. (2017)<sup>33</sup> using laboratory near infrared reflectance spectroscopy (NIRS) measures of wheat grain and poplar bark. Incredibly, through cross-validation, they showed that using many reflectance bands from these products could accurately predict yields in a population. This was validated directly or indirectly in multiple other studies<sup>34-36</sup>. Most incredibly, Rincent et al. (2017)<sup>33</sup> demonstrated that phenomic selection approaches could outperform genomic selection, the gold standard in predicting phenotype, and with less cost. In the case of NIRS, thousands of reflectance bands are treated similarly to genomic markers in genomic selection; quasi-independent repeatable measures separating individuals. In field studies however, even with the most intense measurement of known phenotypes (e.g., height, leaf angle, flowering time, etc.) a rich enough dataset cannot be created to repeatably separate and predict each genotype. Furthermore, most traditional phenotypes are endpoint measures, where interactions with environment are integrated over the organisms life. Unoccupied aerial systems (UAS, UAV, drones) have provided new ways to collect massive amounts of data across large numbers of individuals and environments, extracting novel features<sup>22</sup>. Yet unless hyperspectral bands are used, challenges remain in collecting enough features to use in phenomic prediction. Adak et al. (2022)<sup>37</sup> overcame this hurdle by using dense measurements of temporal features, which was multiplicative in the number of phenomic features and allowed earlier near real-time prediction

ability. This showed that temporal phenomic prediction can indeed perform similarly, perhaps better than genomic prediction. However, even with the largest reported UAS dataset, only 896 phenomic features were extracted compared with 11,000 genomic markers. Phenomic saturation demands more features.

## 2. CONCEPT

### 2.2 Phenomic selection and phenomics as a platform

A cultural shift throughout science and society is required to understand and justify phenomics grand challenge and moonshot. In molecular quantitative genomics, radical changes in application and thought came from shifting goals of genetic mapping, trying to find loci controlling specific traits variation, to genomic selection, using all loci across the genome to find “the best” individual<sup>38</sup>. Notably, in both, more genetic markers were always seen as beneficial to characterize more thoroughly what was occurring in the genome. Marker number was initially limited by technology and resources (e.g. RFLPs and SSRs) compared with current financial limitations to screen for SNPs. In maize genetic mapping, the number of DNA markers went from 190 in 2002<sup>39</sup>, to 1329 in 2006<sup>40</sup>, to 12.2 million in 2022<sup>41</sup>. Despite these advances, we seem little closer to the stated Arabidopsis 2010 goal of determining the function of every gene in the genome. It is somewhat paradoxical that capturing as many phenomic features as possible, without *a priori* knowledge of the underlying biology is considered by some a fishing expedition. Scientific communities have yet to fully consider the benefits of such “random” phenomic measures. Humans are a good example of how such random phenomic measures can be extended to discover nature and nurture over time.

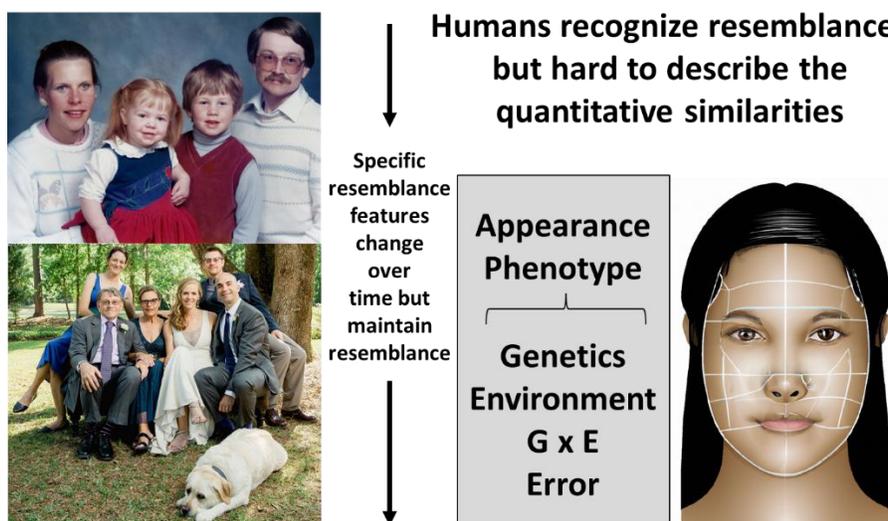


Figure 1. The author’s family showing resemblance over time. Siblings look similar and have facial features suggesting relationships, but exact features and measurements differ due to temporal genetics and divergent interactions with environments. Illustration generated by DALL·E 2 <sup>42</sup>.

### 2.4 Human siblings as an example of phenomic features being products of both nature and nurture

Humans have evolved or been trained to recognize resemblances between relatives. However, when asked what causes resemblance there is usually a loss for words (Fig 1). Obvious traits like hair color, eye color, or other near-Mendelian traits come to mind. However, imaging and computational advances can now accurately measure more subtle and quantitative differences, for example the distance between cheek bones, the wrinkling around the mouth, the shape of the nose<sup>43,44</sup>. Such phenomic approaches could now populate a matrix with thousands of measurements per individual from RGB camera images. These measurements could be used to predict relationships between individuals in the same way kinship matrices use genomic SNPs to find relationships in animals, plants and humans<sup>45-47</sup> and in a similar way that our brains as humans can see these relationships<sup>48</sup>. These connections are important because we believe that closely related individuals tend to perform similarly (although not the same) in many facets such as personality, athletic

ability, or health choices<sup>49,50</sup>. We also know that individuals who grow up in similar environments tend to share phenotypic features, for instance time spent in a war zone will decrease children’s height<sup>51</sup> or that birth months impact human disease probabilities<sup>52</sup>. Therefore, we expect phenomic measures to capture not only genetics but the environment where relatives were raised. It is important to acknowledge the unsettling history of eugenics and ongoing pseudo-scientific racism when talking about measuring humans<sup>53,54</sup>. Yet the human example is useful to connect to readers unfamiliar with plant phenotypes. Plant breeders, who look frequently and closely at their plant progeny, also often recognize resemblances between progeny of their plant crosses that others will not, part of the “breeder’s eye” used in progeny selection.

## 2.4 Evaluating phenomic measurement success: how will we know if we’ve saturated the phenome?

In genomics, linkage disequilibrium (LD) is a good measure but also among the most difficult concepts. Complete LD means a polymorphism between two DNA base pairs are perfectly associated, if the variant at one location is known in a sample then so is the other. Each additional locus measured, when not in perfect LD or correlation with others, adds further predictive information. Once additional markers measured are all in complete LD, the genome and population are fully characterized, and no additional variation could be attributable to the genome. In practice, this remains an unachievable ideal unless every base pair is sequenced in every individual. Likewise, in phenomics, a correlation measure similar to LD between all measured phenotypes could be maximized. Currently, if new phenotypes are found with correlations less than 1 (or greater than -1), the phenome is not saturated. From our published and preliminary studies to date, it appears that even with many dozen vegetation indices extracted from a few RGB spectral bands, and many closely spaced timepoints, correlations between phenomic features remain incomplete (Fig. 2). This supports that additional phenotypic variation can still be discovered and exploited.

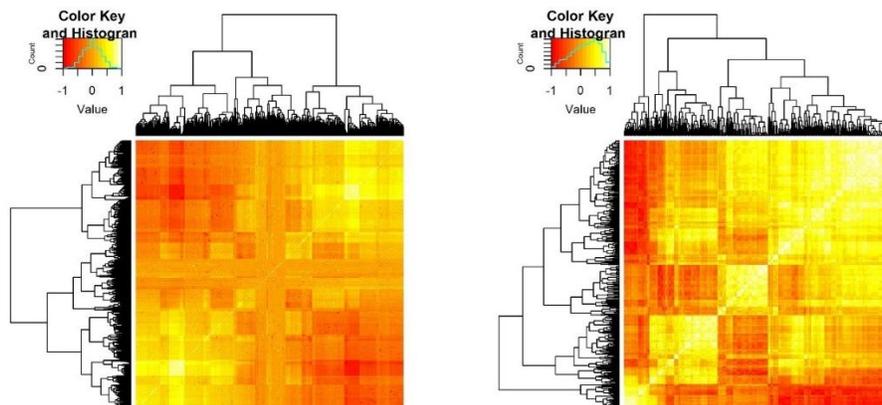


Fig. 2 Genomic correlations (left) and phenomic correlations (right) on 520 recombinant inbred lines. Genomic data had 11,334 SNPs after cleaning<sup>55</sup>. Phenomic data included 896 features (32 RGB derived vegetation indices over 14 flight dates and two management environments, drought and irrigated). Both have measures with high and low correlations (+ or -) and are unsaturated<sup>24</sup>.

## CONCLUSION

“You can’t really know where you are going until you know where you have been.” – Maya Angelou  
 For the phenomics community to mature beyond being seen as simply working to increase the throughput of measurements, or reduced to service of genomics<sup>56,57</sup>, a broader vision including a grand challenge and a moonshot are needed. The area of phenomic selection has already provided tantalizing new insight into how massive amounts of biological phenotypic data can be used to discover unanticipated biological truths. For phenomics discovery to succeed, we need to apply similar approaches to areas such as germplasm conservation, ecology and ultimately improvements in farmers fields. The history and development of the fields of genetics<sup>58</sup>, statistics<sup>59</sup>, and even evolution<sup>60</sup> owe a great deal to agricultural improvement research interests, and phenomics could follow a similar trajectory. The bottleneck to date, is high volume temporal measurements made on a population’s individuals across many environments.

## DATA AVAILABILITY STATEMENT

No novel data was used in this concept or production of this paper.

## ACKNOWLEDGMENTS

The authors wish to thank our mentors, predecessors and the many colleagues, staff and students that contributed to the scientific evolution to get us to this exciting time in the evolution of phenomics, as well as the entire phenomics community pushing the frontiers of knowledge in this area. Financial support for this research has been provided by: USDA–NIFA–AFRI Award Nos. 2020-68013-32371, and 2021-67013-33915, USDA–NIFA Hatch funds, Texas A&M AgriLife Research, the Texas Corn Producers Board, and the Iowa Corn Promotion Board, the Eugene Butler Endowed Chair in Biotechnology.

## REFERENCES

- [1] Jez, J.M., Topp, C.N., Matthews, M.L. and Marshall-Colón, A., “Multiscale plant modeling: from genome to phenome and beyond.” *Emerging topics in life sciences*, 5(2), 231-237. (2021).
- [2] Henkhaus, N., Bartlett, M., Gang, D., Grumet, R., Jordon-Thaden, I., Lorence, A., Lyons, E., Miller, S., Murray, S., Nelson, A. and Specht, C., “Plant science decadal vision 2020–2030: Reimagining the potential of plants for a healthy and sustainable future,” *Plant direct*, 4(8), e00252 (2020).
- [3] Rexroad, C., Vallet, J., Matukumalli, L.K., Reecy, J., Bickhart, D., Blackburn, H., Boggess, M., Cheng, H., Clutter, A., Cockett, N. and Ernst, C., “Genome to phenome: improving animal health, production, and well-being—a new USDA blueprint for animal genome research 2018–2027,” *Frontiers in genetics*, 10, 327 (2019).
- [4] Shakoor, N., Northrup, D., Murray, S., & Mockler, T. C. “Big data driven agriculture: big data analytics in plant breeding, genomics, and the use of remote sensing technologies to advance crop productivity,” *The Plant Phenome Journal*, 2(1), 1-8. (2019).
- [5] Tuggle, C.K., Clarke, J., Dekkers, J., Ertl, D., Lawrence-Dill, C.J., Lyons, E., Murdoch, B.M., Scott, N.M. and Schnable, P.S., “The Agricultural Genome to Phenome Initiative (AG2PI): creating a shared vision across crop and livestock research communities,” *Genome biology*, 23(1), 1-11 (2022).
- [6] Carroll, A.A., Clarke, J., Fahlgren, N., Gehan, M.A., Lawrence-Dill, C.J. and Lorence, A., “NAPPN: Who we are, where we are going, and Why You Should Join Us!,” *The Plant Phenome Journal*, 2(1), 1-4 (2019).
- [7] Lander, E. S. “Initial impact of the sequencing of the human genome,” *Nature*, 470(7333), 187-197 (2011).
- [8] Church, D.M., “A next-generation human genome sequence,” *Science*, 376(6588), 34-35 (2022).
- [9] NSF. 2010 Project. “To determine the function of all genes in *Arabidopsis thaliana* by the year 2010” <<https://www.nsf.gov/pubs/2004/nsf04502/nsf04502.htm>> (21 October 2010).
- [10] Wee, C. W., & Dinneny, J. R., “Tools for high-spatial and temporal-resolution analysis of environmental responses in plants,” *Biotechnology letters*, 32(10), 1361-1371 (2010).
- [11] Marshall-Colon, A., Long, S.P., Allen, D.K., Allen, G., Beard, D.A., Benes, B., Von Caemmerer, S., Christensen, A.J., Cox, D.J., Hart, J.C. and Hirst, P.M., “Crops in silico: generating virtual crops using an integrative and multi-scale modeling platform,” *Frontiers in plant science*, 8, 786 (2017).
- [12] Ubbens, J., Cieslak, M., Prusinkiewicz, P. and Stavness, I., “The use of plant models in deep learning: an application to leaf counting in rosette plants,” *Plant methods*, 14(1), 1-10 (2018).
- [13] Das, A., Schneider, H., BurrIDGE, J., Ascanio, A.K.M., Wojciechowski, T., Topp, C.N., Lynch, J.P., Weitz, J.S. and Bucksch, A., “Digital imaging of root traits (DIRT): a high-throughput computing and collaboration platform for field-based root phenomics,” *Plant methods*, 11(1), 1-12 (2015).
- [14] Gehan, M.A., Fahlgren, N., Abbasi, A., Berry, J.C., Callen, S.T., Chavez, L., Doust, A.N., Feldman, M.J., Gilbert, K.B., Hodge, J.G. and Hoyer, J.S., “PlantCV v2: Image analysis software for high-throughput plant phenotyping,” *PeerJ*, 5, e4088 (2017).

- [15] Matias, F. I., Green, A., Lachowiec, J. A., LeBauer, D., & Feldman, M., “Bison-Fly: An open-source UAV pipeline for plant breeding data collection,” *The Plant Phenome Journal*, 5(1), e20048 (2022).
- [16] Seethepalli, A., Dhakal, K., Griffiths, M., Guo, H., Freschet, G.T. and York, L.M., “RhizoVision Explorer: open-source software for root image analysis and measurement standardization,” *AoB plants*, 13(6), plab056 (2021).
- [17] Langstroff, A., Heuermann, M.C., Stahl, A. and Junker, A., “Opportunities and limits of controlled-environment plant phenotyping for climate response traits,” *Theoretical and Applied Genetics*, 1-16 (2021).
- [18] Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O. and Schurr, U., “Phenotyping: new windows into the plant for breeders,” *Annual review of plant biology*, 71 689-712 (2020).
- [19] Chen, R., Chu, T., Landivar, J.A., Yang, C. and Maeda, M.M., “Monitoring cotton (*Gossypium hirsutum* L.) germination using ultrahigh-resolution UAS images,” *Precision Agriculture*, 19(1), 161-177 (2018).
- [20] Chu, T., Starek, M.J., Brewer, M.J., Murray, S.C. and Pruter, L.S., “Assessing lodging severity over an experimental maize (*Zea mays* L.) field using UAS images,” *Remote Sensing*, 9(9), 923 (2017).
- [21] Neupane, K. and Baysal-Gurel, F., “Automatic identification and monitoring of plant diseases using unmanned aerial vehicles: A review,” *Remote Sensing*, 13(19), 3841 (2021).
- [22] Shi, Y., Thomasson, J.A., Murray, S.C., Pugh, N.A., Rooney, W.L., Shafian, S., Rajan, N., Rouze, G., Morgan, C.L., Neely, H.L. and Rana, A., “Unmanned aerial vehicles for high-throughput phenotyping and agronomic research,” *PloS one*, 11(7), e0159781 (2016).
- [23] Wu, H., Wiesner-Hanks, T., Stewart, E.L., DeChant, C., Kaczmar, N., Gore, M.A., Nelson, R.J. and Lipson, H., “Autonomous detection of plant disease symptoms directly from aerial imagery,” *The plant phenome journal*, 2(1), pp.1-9 (2019).
- [24] Anderson, S.L., Murray, S.C., Chen, Y., Malambo, L., Chang, A., Popescu, S., Cope, D. and Jung, J., “Unoccupied aerial system enabled functional modeling of maize height reveals dynamic expression of loci,” *Plant Direct*, 4(5), e00223. (2020).
- [25] Pauli, D., Andrade-Sanchez, P., Carmo-Silva, A.E., Gazave, E., French, A.N., Heun, J., Hunsaker, D.J., Lipka, A.E., Setter, T.L., Strand, R.J. and Thorp, K.R., “Field-based high-throughput plant phenotyping reveals the temporal patterns of quantitative trait loci associated with stress-responsive traits in cotton,” *G3: Genes, Genomes, Genetics*, 6(4), 865-879 (2016).
- [26] Ghosal, S., Blystone, D., Singh, A.K., Ganapathysubramanian, B., Singh, A. and Sarkar, S., “An explainable deep machine vision framework for plant stress phenotyping,” *Proceedings of the National Academy of Sciences*, 115(18), 4613-4618 (2018).
- [27] Ubbens, J.R. and Stavness, I., “Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks,” *Frontiers in plant science*, 8, 1190 (2017).
- [28] DeSalvio, A.J., Adak, A., Murray, S.C., Wilde, S.C. and Isakeit, T., “Phenomic data-facilitated rust and senescence prediction in maize using machine learning algorithms,” *Scientific reports*, 12(1), 1-14 (2022).
- [29] Wu, H., Wiesner-Hanks, T., Stewart, E.L., DeChant, C., Kaczmar, N., Gore, M.A., Nelson, R.J. and Lipson, H., “Autonomous detection of plant disease symptoms directly from aerial imagery,” *The plant phenome journal*, 2(1), 1-9 (2019).
- [30] Brewer, K., Clulow, A., Sibanda, M., Gokool, S., Odindi, J., Mutanga, O., Naiken, V., Chimonyo, V.G. and Mabhaudhi, T., “Estimation of maize foliar temperature and stomatal conductance as indicators of water stress based on optical and thermal imagery acquired using an Unmanned Aerial Vehicle (UAV) platform,” *Drones*, 6(7), 169 (2022).
- [31] Ezenne, G.I., Jupp, L., Mantel, S.K. and Tanner, J.L., “Current and potential capabilities of UAS for crop water productivity in precision agriculture,” *Agricultural Water Management*, 218, 158-164 (2019).
- [32] Zhang, L., Niu, Y., Zhang, H., Han, W., Li, G., Tang, J. and Peng, X., “Maize canopy temperature extracted from UAV thermal and RGB imagery and its application in water stress monitoring,” *Frontiers in plant science*, 10, 1270 (2019).

- [33] Rincent, R., Charpentier, J.P., Faivre-Rampant, P., Paux, E., Le Gouis, J., Bastien, C. and Segura, V., “Phenomic selection is a low-cost and high-throughput method based on indirect predictions: proof of concept on wheat and poplar,” *G3: Genes, Genomes, Genetics*, 8(12), 3961-3972 (2018).
- [34] Lane, H.M., Murray, S.C., Montesinos López, O.A., Montesinos López, A., Crossa, J., Rooney, D.K., Barrero-Farfan, I.D., De La Fuente, G.N. and Morgan, C.L., “Phenomic selection and prediction of maize grain yield from near-infrared reflectance spectroscopy of kernels,” *The Plant Phenome Journal*, 3(1), e20002 (2020).
- [35] Krause, M.R., González-Pérez, L., Crossa, J., Pérez-Rodríguez, P., Montesinos-López, O., Singh, R.P., Dreisigacker, S., Poland, J., Rutkoski, J., Sorrells, M. and Gore, M.A., “Hyperspectral reflectance-derived relationship matrices for genomic prediction of grain yield in wheat,” *G3: Genes, Genomes, Genetics*, 9(4), 1231-1247 (2019).
- [36] Robert, P., Brault, C., Rincent, R. and Segura, V., “Phenomic selection: A new and efficient alternative to genomic selection,” In [Complex Trait Prediction] Humana, New York, 397-420 (2022).
- [37] Adak, A., Murray, S.C. and Anderson, S.L., “Temporal phenomic predictions from unoccupied aerial systems can outperform genomic predictions,” *BioRxiv*, G3 accepted (2021).
- [38] Bernardo, R. and Yu, J., “Prospects for genomewide selection for quantitative traits in maize,” *Crop Science*, 47(3), 1082-1090 (2007).
- [39] Lee, M., Sharopova, N., Beavis, W.D., Grant, D., Katt, M., Blair, D. and Hallauer, A., “Expanding the genetic map of maize with the intermated B73× Mo17 (IBM) population,” *Plant molecular biology*, 48(5), 453-461 (2002).
- [40] Fu, Y., Wen, T.J., Ronin, Y.I., Chen, H.D., Guo, L., Mester, D.I., Yang, Y., Lee, M., Korol, A.B., Ashlock, D.A. and Schnable, P.S., “Genetic dissection of intermated recombinant inbred lines using a new genetic map of maize,” *Genetics*, 174(3), 1671-1683 (2006).
- [41] Wu, D., Li, X., Tanaka, R., Wood, J.C., Tibbs-Cortes, L.E., Magallanes-Lundback, M., Bornowski, N., Hamilton, J.P., Vaillancourt, B., Diepenbrock, C.H. and Li, X., “Combining GWAS and TWAS to identify candidate causal genes for tocochromanol levels in maize grain,” *bioRxiv* (2022).
- [42] DALL·E 2 using prompt “Photorealistic image centered one blank human female face attention map showing genetic control of facial morphology regions without text,” <<https://openai.com/dall-e-2/>> (23 October 2022).
- [43] Liu, F., Van Der Lijn, F., Schurmann, C., Zhu, G., Chakravarty, M.M., Hysi, P.G., Wollstein, A., Lao, O., De Bruijne, M., Ikram, M.A. and Van Der Lugt, A., “A genome-wide association study identifies five loci influencing facial morphology in Europeans,” *PloS Genetics* e1002932 (2012).
- [44] Richmond, S., Howe, L.J., Lewis, S., Stergiakouli, E. and Zhurov, A., “Facial genetics: a brief overview,” *Frontiers in genetics*, 9, 462 (2018).
- [45] Goudet, J., Kay, T. and Weir, B.S., “How to estimate kinship,” *Molecular ecology*, 27(20), 4121-4135 (2018).
- [46] Speed, D. and Balding, D.J., “Relatedness in the post-genomic era: is it still useful?,” *Nature Reviews Genetics*, 16(1), 33-44 (2015).
- [47] Yu, J., Zhang, Z., Zhu, C., Tabanao, D.A., Pressoir, G., Tuinstra, M.R., Kresovich, S., Todhunter, R.J. and Buckler, E.S., “Simulation appraisal of the adequacy of number of background markers for relationship estimation in association mapping,” *The Plant Genome*, 2(1) (2009).
- [48] Alvergne, A., Huchard, E., Caillaud, D., Charpentier, M.J., Setchell, J.M., Ruppli, C., Féjan, D., Martinez, L., Cowlshaw, G. and Raymond, M., “Human ability to recognize kin visually within primates,” *International Journal of Primatology*, 30(1), 199-210 (2009).
- [49] Georgiades, E., Klissouras, V., Baulch, J., Wang, G. and Pitsiladis, Y., “Why nature prevails over nurture in the making of the elite athlete,” *BMC genomics*, 18(8), 59-66 (2017).
- [50] Malouff, J.M., Rooke, S.E. and Schutte, N.S., “The heritability of human behavior: Results of aggregating meta-analyses,” *Current Psychology*, 27(3), 153-161 (2008).
- [51] Currie, J. and Vogl, T., “Early-life health and adult circumstance in developing countries,” *Annu. Rev. Econ.*, 5(1), 1-36 (2013).

- [52] Boland, M.R., Shahn, Z., Madigan, D., Hripesak, G. and Tatonetti, N.P., “Birth month affects lifetime disease risk: a phenome-wide method,” *Journal of the American Medical Informatics Association*, 22(5), 1042-1053 (2015).
- [53] Marks, J., “Historiography of eugenics,” *American Journal of Human Genetics*, 52(3), 650 (1993).
- [54] Panofsky, A., Dasgupta, K. and Iturriaga, N., “How White nationalists mobilize genetics: From genetic ancestry and human biodiversity to counterscience and metapolitics,” *American Journal of Physical Anthropology*, 175(2), 387-398. (2021).
- [55] Chen, Y., “High-density linkage map construction, mapping of agronomic traits in tropical maize (*Zea Mays* L.) and validating SNPs controlling maize grain yield and plant height in southern hybrid testcrosses” (Texas A&M Doctoral dissertation) 2016.
- [56] Furbank, R.T. and Tester, M., “Phenomics—technologies to relieve the phenotyping bottleneck,” *Trends in plant science*, 16(12), 635-644 (2011).
- [57] Houle, D., Govindaraju, D.R. and Omholt, S., “Phenomics: the next challenge,” *Nature reviews genetics*, 11(12), 855-866 (2010).
- [58] Allen, G. E., “Mendel and modern genetics: the legacy for today,” *Endeavour*, 27(2), 63-68 (2003).
- [59] Box, J.F., “Guinness, Gosset, Fisher, and small samples,” *Statistical science*, 45-52 (1987).
- [60] East, E.M., “The relation of certain biological principles to plant breeding,” *Connecticut Agricultural Experiment Station*, No. 158, (1907).

## **REVIEWS**

Anonymous reviews will be published

## **RESPONSE TO REVIEWERS**

Response to reviews will be published