

# TERRA Program Overview

## B. PROGRAM OVERVIEW

### Summary

There is an urgent need to accelerate energy crop development for the production of renewable transportation fuels from biomass. Recent technological advancements have now made it possible to extract massive volumes of genetic, physiological, and environmental data from certain crops, but, even with these resources, the data still cannot be processed into the knowledge needed to predict crop performance in the field. This knowledge is required to improve the breeding development pipeline for energy crops. Building upon precision agriculture innovations and data-intensive computational approaches, ARPA-E believes that it is now possible to accelerate plant breeding, using robust high-throughput precision phenotyping systems, to quantify important agronomic traits in the field throughout the entire lifecycle of an individual plant, and to associate these traits with their genetic and genomic properties. This ARPA-E program, Transportation Energy Resources from Renewable Agriculture (TERRA), is an investment in technologies that increase the precision, accuracy and throughput of energy crop breeding, to enable (a) new predictive algorithms for plant growth, (b) more detailed measurements for plant physiology, and (c) more sophisticated bioinformatics pipelines for gene discovery and trait association. TERRA will enable breeders to evaluate more individual plants, to select appropriate plants for breeding earlier in the growing season, to capture better information about them during their development, and to associate this information with the best genes to propagate. Success will be measured by the prospective ability to predict yield gains early, specifically, to identify which genes can improve carbon capture efficiency in newly cultivated bioenergy crops. Although other crops will be considered, this program intends to focus on energy sorghum as a model system because of its potential for improvement through breeding, its resources for genetic analysis, its geographic adaptability, and its commercial utility.

### Motivation

Fuel used in the U.S. transportation sector has become more diverse in the past several years. While gasoline remains the dominant fuel, the market penetration of diesel, biofuels, and hybrid-electric systems is growing, eroding gasoline's share of the light duty vehicle fuel market. This trend is too slow to reduce energy-related CO<sub>2</sub> emissions: petroleum remains, by far, the largest source of transportation fuel in the world, a significant, but non-renewable, resource. In 2013, 36% of U.S. energy consumption was from petroleum, producing 42% of its energy-related carbon dioxide emissions.<sup>1</sup> Thus, the more rapid development of transportation fuels with decreased, neutral, or negative carbon emissions is required to reduce the amount of foreign oil imports and limit the rate of increase of CO<sub>2</sub> in the atmosphere, addressing two of ARPA-E's core missions.

Economical production of the large amounts of biomass needed to displace petroleum will require significant productivity and efficiency improvements from the agricultural sector, which is also responsible for human and animal nutrition. In 2014, the United Nations warned that world agriculture must increase its output 60% by 2050 (1.6% per year, on average) to support global population growth and economic development.<sup>2</sup> Meanwhile, Hall and Richards report that the annual genetic gain for the main cereal crops best varieties and hybrids falls well below 1.16–1.31% per year and are not able to satisfy projected growing demand.<sup>3</sup> Consequently, the realization of commercially viable agriculture for energy purposes requires unprecedented increases in productivity and resource use efficiency.

A conventional breeding approach takes many years to improve crop varieties. The rate of crop improvement through breeding is strongly correlated to technology, increasing with better and more complete field data, i.e. the precision and accuracy of trait measurements and the throughput of screening. The ability to rapidly identify plants in a breeding

<sup>1</sup> Annual Energy Outlook 2014, United States Department of Energy, Energy Information Agency, <http://www.eia.gov/forecasts/aeo/>

<sup>2</sup> United Nations Food and Agriculture Organization (FAO), World Agriculture Towards 2030/2050, 2014: Europe.

<sup>3</sup> Antonio J Hall and Richard A Richards, Field Crops Research, 2013. 143: p. 18-33.

population with desirable traits will increase the rate of genetic gain of the crop and improve the yield of bioenergy from agriculture.<sup>4</sup>

It is important to define the terminology of breeding: A **phenotype** is an observable or measurable physical trait, such as color, height, size, shape, behavior or chemical composition. Phenotype is determined by the individual's **genotype**, the information encoded in the DNA polymers present in its genome, as well as how that genome interacts with the environment. **Phenotyping** is the measurement of phenotypes, a process that can require substantial effort and may be highly dependent on data interpretation by a plant breeder. The development of advanced plant phenotyping technologies is substantially behind that of genotyping, and thus poses a key bottleneck on the path toward increased bioenergy crop yields. While agriculture has the capability to increase biomass yields and mitigate the effects of anthropogenic greenhouse gas emissions, the slow pace of conventional breeding limits this capability. Advancing and integrating cutting edge technology phenotyping platforms in genomics, computational analytics, proximal sensing, and automation, within the TERRA program, will contribute to breaking through this barrier to bioenergy crop development.

## State of the Art

### Agricultural Phenotyping

Plants use solar energy to convert atmospheric CO<sub>2</sub> into fixed carbon, which is then further used as a source of food, feed, fiber, and fuel. Plant breeding is the process of mating sexually compatible plants to generate agriculturally superior varieties. Traditional breeding methods are slow and inefficient; the challenge is to rapidly identify which genotype-phenotype combinations lead to substantial crop improvement from a large and diverse population.<sup>5</sup>

High throughput phenotyping technologies have been developed, and have the capability to accurately characterize large numbers of plants with much less time and labor than in the past. To date, these technologies have been applied only under tightly controlled laboratory and greenhouse conditions. Much of the technology improvement has focused on fully automated greenhouse systems using plants grown in uniform pots under precise conditions, or laboratory techniques that aim to correlate tissue culture observations with those of plants in the field.<sup>6,7, 8,9,10</sup> While these systems are valuable, they are not easily scalable and the relationship to crop responses in field situations is controversial, because soil volumes, solar radiation, wind speeds and evaporation rates in greenhouses are often much lower than in the field.<sup>11,12,13</sup>

The set of tools that currently exists for use in field environments is small and underdeveloped. These tools range from handheld devices aimed at capturing reflectance and spectral data, to prototype mobile platforms outfitted with crop sensors.<sup>14,15,16,17,18</sup> Today, the primary tool for measuring phenotype in the field is the harvest combine, which accurately collects the most relevant performance metric: terminal yield. However, terminal yield is captured destructively only at the very end of the season, and provides very little feedback or insight on crop development throughout the growing season. Thus, there is an urgent need for robust field-based high-throughput phenotyping systems for quantifying agronomic important traits at field scales throughout the crop lifecycle.<sup>19,20</sup>

<sup>4</sup> Stephen P Moose and Rita H Mumm, *Plant physiology*, 2008. 147(3): p. 969-977.

<sup>5</sup> Stephen P Moose and Rita H Mumm, *Plant physiology*, 2008. 147(3): p. 969-977.

<sup>6</sup> Christian Klukas, Dijun Chen, and Jean-Michel Pape, *Plant physiology*, 2014. 165(2): p. 506-518.

<sup>7</sup> Anja Hartmann, T. Czauderna, R. Hoffmann, N. Stein, and F. Schreiber, *BMC bioinformatics*, 2011. 12(1): p. 148.

<sup>8</sup> Nora Honsdorf, T. John March, Bettina Berger, Mark Tester, and Klaus Pillen, *PloS one*, 2014. 9(5): p. e97047.

<sup>9</sup> Stefan Paulus, Henrik Schumann, Heiner Kuhlmann, and Jens Léon, *Biosystems Engineering*, 2014. 121: p. 1-11.

<sup>10</sup> Michael Malone, *PhenoDay*. 2011: Wageningen, Netherlands.

<sup>11</sup> J. Cobb, G. DeClerck, A. Greenberg, R. Clark, S. McCouch, *Theoretical Applied Genetics*, 2013. 126(4): p. 867-887.

<sup>12</sup> Robert T Furbank and Mark Tester, 2011. 16(12): p. 635-644.

<sup>13</sup> Jeffrey W White, Pedro Andrade-Sanchez, Michael A Gore, Kevin Bronson, Terry Coffelt, Matthew Conley, Kenneth Feldmann, Andrew French, John Heun, Douglas Hunsaker, *Field Crops Research*, 2012. 133: p. 101-112.

<sup>14</sup> Robert T Furbank and Mark Tester, *Trends in plant science*, 2011. 16(12): p. 635-644.

<sup>15</sup> Jeffrey White, Pedro Andrade-Sanchez, Michael Gore, Kevin Bronson, Terry Coffelt, Matthew Conley, Kenneth Feldmann, Andrew French, John Heun, and Douglas Hunsaker, *Field Crops Research*, 2012. 133: p. 101-112.

<sup>16</sup> Kyle H Holland and James S Schepers, *Precision Agriculture*, 2013. 14(1): p. 71-85.

<sup>17</sup> P. Andrade-Sanchez, John Heun, Michael Gore, Andrew French, E. Carmo-Silva, and M. Salvucci, *Proceedings of the 2012 American Society of Agricultura and Biological Engineers Annual International Meeting*. 2012: Dallas, TX.

<sup>18</sup> Fabio Fiorani and Ulrich Schurr, *Annual review of plant biology*, 2013. 64: p. 267-291.

<sup>19</sup> Robert T Furbank and Mark Tester, *Trends in plant science*, 2011. 16(12): p. 635-644.

<sup>20</sup> Jeffrey White, Pedro Andrade-Sanchez, Michael Gore, Kevin Bronson, Terry Coffelt, Matthew M Conley, Kenneth A Feldmann, Andrew N French, John T Heun, and Douglas J Hunsaker, *Field Crops Research*, 2012. 133: p. 101-112.

## Energy Crops

Numerous crops are under consideration for energy production, and many factors must be taken into account when selecting an energy crop for a particular growing location.<sup>21</sup> *Sorghum bicolor* (a phenotypically-diverse species, currently available in grain, sweet and energy varieties) is a highly productive C4 grass that has been identified as a particularly useful crop for improvement. It is an important cultivated food, feed, and bioenergy crop worldwide, and is the third most widely cultivated cereal crop in the U.S., which is also its number one producer and exporter.<sup>22</sup> Among sorghum types, energy sorghum has exceptional potential as a high biomass bioenergy crop, with yields in certain regions reaching nearly 60 dT ha<sup>-1</sup> (dry metric tons per hectare) under ideal conditions. Energy sorghum adapts well to drought and heat conditions that are inhospitable to food crops, making it an excellent model for studying plant-environment interactions. In addition, sorghum is an annual crop, meaning that it can be bred rapidly and evaluated easily in diverse environments to enable the testing and optimization of automated precision phenotyping, both in the field and in well-controlled environments.<sup>23,24</sup>

## Genomics and Bioinformatics

The field of genomics aims to decipher the information content of the genome. Enabled by the extremely rapid progress of high-throughput gene sequencing technologies, valuable computational analysis at the genomic level using advanced bioinformatics techniques is now possible. The genome interacts with the environment in a complex manner to form a plant's many phenotypes, which develop over the plant's lifecycle. The bioinformatics field of statistical genetics aims to elucidate the dependence of a particular genetic background on a phenotype, given environmental inputs. Analytically, understanding *and predicting* this interaction between genetics (G) and environment (E) that produces a particular phenotype (P), as given by Eq 1, is the central goal of the TERRA program.

$$P = G \times E \quad (\text{Eq. 1})$$

Determination of a genotype need not require complete genome sequencing. Biologists have traditionally used genomic assays such as allele-specific polymerase chain reaction (PCR), which requires the *a priori* identification of characteristic alleles (alternative gene forms) in the genome. The most common of these changes are called single nucleotide polymorphisms (SNPs)<sup>25</sup>; identification of common SNPs in populations has been the subject of much research over the past few decades. As illustrated by Eq 1, assaying these known SNPs in energy crop populations and correlating them with phenotype is one of the foundations of statistical genetics.

Marker-assisted breeding has used selected genetic landmarks such as SNPs for decades to help breeders determine which lines to cross to confer given phenotypes, usually with little knowledge of the underlying physiological basis. Inexpensive DNA sequencing has added depth of information to this field, leading to the sub-discipline of genomic prediction. In genomic prediction studies, models of *all* the SNP data can be created, allowing for a prediction and earlier selection of the phenotype of offspring from knowledge of only the parental genotypes.

Genome Wide Association (GWA) studies are empirical studies that use statistical genetic tools to associate discrete or continuous traits within a population to the specific DNA sequences that underlie that variation. GWA studies can identify regions of the genome (known as quantitative trait loci, or QTLs) that are statistically associated with the variation in phenotype. The genes or regulatory elements within each QTL are inferred to be the mediators of the variation, and the particular genetic differences that comprise those variants.

A commonality between GWA and genomic prediction studies is the goal of predicting phenotypes from genetic information. In the context of agriculture, accurate and early prediction could enable a transformational shift in breeding resources: tens of thousands of crosses could be analyzed *in silico*, and only those with the highest probability of an improved phenotype moved into the breeding pipeline. Because energy crops have received much less attention from

<sup>21</sup> C. Somerville, H. Youngs, C. Taylor, S. Davis, and S. Long, *Science*, 2010. 329(5993): p. 790-792.

<sup>22</sup> Crop Production 2013 Summary, United States Department of Agriculture, <http://usda.mannlib.cornell.edu/usda/current/CropProdSu/CropProdSu-01-10-2014.pdf>

<sup>23</sup> Sara N Olson, Kimberley Ritter, William Rooney, Armen Kemanian, Bruce A McCarl, Yuquan Zhang, Susan Hall, Dan Packer, and John Mullet, *Biofuels, Bioproducts and Biorefining*, 2012. 6(6): p. 640-655.

<sup>24</sup> S. Olson, K. Ritter, J. Medley, T. Wilson, W. Rooney, and J. Mullet, *Biomass and Bioenergy*, 2013. 56: p. 307-316.

<sup>25</sup> Xuehui Huang and Bin Han, , 2014. 65: p. 531-551.

breeders than food and feed crops (and therefore have much more unrealized genetic potential)<sup>26,27,28</sup>, rapid analysis is particularly important. However, realizing predictive genomics in bioenergy crops *depends on* the generation and availability of large amounts of high quality phenotypic data.

### Sensing and Automation

Sensors and automation have already made significant contributions toward lowering production costs in agriculture, reducing manual labor and raising the quality of production.<sup>29,30</sup> Over the past decade, automated precision farming technologies that employ vision systems, laser sensors, and satellite positioning instruments have emerged. The most widely adopted technologies include autonomous equipment navigation, yield mapping, and variable rate technologies for planting and application of fertilizers and crop protection chemicals.<sup>31</sup> The investments in precision agriculture have enabled better specificity in the field as well. As opposed to classical techniques of measuring or manipulating the field that focus on entire rows or plots, high-precision systems incorporate GPS and other positional accuracy aids to enable plant level specificity of 2 cm or better.<sup>32</sup>

Current methods for plant phenotyping using sensor measurements span many modalities and include imaging across various spectral channels, fluorescence measurements, topographical analysis using LiDAR or stereo cameras, and environmental sensing.<sup>33,34,35,36,37,38</sup> Numerous phenotypes are important for agricultural breeder selections, but accurate measurements of aboveground biomass throughout the growing season (from which growth rates can be derived) are of critical importance. Such estimates are possible using greenhouse systems that measure individual plants from several angles (*R* values of 0.98 have been demonstrated).<sup>39</sup> Furthermore, some of these collection methods are faster than others. For example, hyperspectral and fluorescence measurements take much longer than camera, thermal, or reflectance measurements, but provide a wealth of additional data.

Spectroscopic measurements taken in the field can be correlated with quantitative phenotypes of interest using indices such as leaf area index (LAI), vegetation fraction, and canopy nitrogen. These indices are typically measured by taking reflectance spectroscopy measurements of two or more specific spectral bands. Examples include the Normalized Difference Vegetation Index (NDVI) and Visible Atmospheric Resistant Index (VARI) indices. Such methods have spawned a small community of researchers discovering new spectral indices that correlate with various phenotypic quantities of interest. Combining indices or adding other modalities such as thermal imaging can further enhance phenotypic accuracy. This approach, however, has several shortcomings: (1) the correlation between the proposed index and a quantity of interest is often weak, (2) these indices often do not account for changes in the environment, weather, or across testing domains, and do not necessarily translate among different crops/species, (3) the indices often have limited observational regimes, making prediction above a certain value (of LAI, for example) impossible, and (4) the development of these indices is largely empirical. Researchers propose a new index (for example, selecting two spectral channels and taking their ratio) and then determine its correlation with some phenotypic quantity of interest. Not surprisingly, this often leads to weak correlations that lack robustness.

As researchers have deployed the first set of field-based automated sensing platforms for phenotyping, critical tradeoffs have emerged between collection time, phenotypes of interest, and data volume or data precision. At this point breeders

<sup>26</sup> Philip G Pardey, Julian M Alston, and Connie Chan-Kang, *Agricultural Economics*, 2013. 44(s1): p. 103-113.

<sup>27</sup> OECD, *Improving Agricultural Knowledge and Innovation Systems: OECD Conference Proceedings*, 2012, <http://dx.doi.org/10.1787/9789264167445-en>

<sup>28</sup> Jenifer Piesse and C Thirtle, *Agricultural R&D, technology and productivity*. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2010. 365(1554): p. 3035-3047.

<sup>29</sup> Yael Edan, Shufeng Han, and Naoshi Kondo, 2009, Springer. p. 1095-1128.

<sup>30</sup> Tony Grift, Qin Zhang, Naoshi Kondo, and KC Ting, *Journal of Biomechatronics Engineering*, 2008. 1(1): p. 37-54.

<sup>31</sup> David Schimmelpfennig and Robert Ebel, 2011: US Department of Agriculture, Economic Research Service.

<sup>32</sup> Jeffrey W White, Pedro Andrade-Sanchez, Michael A Gore, Kevin F Bronson, Terry A Coffelt, Matthew M Conley, Kenneth Feldmann, Andrew French, John Heun, Douglas Hunsaker, *Field Crops Research*, 2012. 133: p. 101-112.

<sup>33</sup> Michael Malone, *High-throughput phenotyping – taking crop biotechnology to the next level*, in *PhenoDay*. 2011: Wageningen, Netherlands.

<sup>34</sup> Joshua N Cobb, Genevieve DeClerck, Anthony Greenberg, Randy Clark, and Susan McCouch, *Theoretical and Applied Genetics*, 2013. 126(4): p. 867-887.

<sup>35</sup> Robert T Furbank and Mark Tester, *Trends in plant science*, 2011. 16(12): p. 635-644.

<sup>36</sup> Jeffrey White, Pedro Andrade-Sanchez, Michael Gore, Kevin Bronson, Terry Coffelt, Matthew Conley, Kenneth Feldmann, Andrew French, John Heun, Douglas Hunsaker, *Field Crops Research*, 2012. 133: p. 101-112.

<sup>37</sup> Kyle H Holland and James S Schepers, *Precision Agriculture*, 2013. 14(1): p. 71-85.

<sup>38</sup> P. Andrade-Sanchez, John T Heun, Michael Gore, Andrew French, E. Carmo-Silva, and M. Salvucci, *Proceedings of the 2012 American Society of Agricultura and Biological Engineers Annual International Meeting*. 2012: Dallas, TX.

<sup>39</sup> Nora Honsdorf, Timothy John March, Bettina Berger, Mark Tester, Klaus Pillen, *PLoS one*, 2014. 9(5): p. e97047.

still have to choose between employing a slow system ( $\leq 2 \text{ km h}^{-1}$ ) to achieve plant-level specificity, or a fast system such as an aerial survey that can cover an entire field within seconds but focuses more on plot-level specificity.<sup>40</sup> Neither of these approaches offers breeders access to data on individual plants with the capacity for frequent (daily or better) updates across a large area (tens to hundreds of acres). And without knowing which phenotypes are the most essential, selecting the appropriate sensor suite and corresponding collection rate is exceedingly challenging.

### C. PROGRAM OBJECTIVES

The overall objective of the TERRA program is to develop tools that enable an increase in the rate and extent of genetic improvement of the yield of bioenergy crops grown in the field. If successful, the program will enhance land use efficiency, reduce competition between bioenergy and food crops, improve environmental sustainability, and provide a more stable supply of biomass for transportation fuels and bio-refineries.

As shown in Equation 2, the rate of genetic gain or crop improvement per cycle of breeding  $R_t$  is related to the intensity of genetic selection,  $i$ , the extent of phenotypic variation controlled by genetics within the materials being screened (phenotypic variance,  $\sigma_A$ ), the selection accuracy ( $r$ ) impacted by the heritability of the trait(s) under selection, and the length of the breeding cycle,  $y$ . Automated precision phenotyping for renewable energy has the potential to increase the intensity of genetic selection by providing more detailed phenotypic information (multiple times during a growing season, with spatial information and more information content) and by allowing more genetic material with a greater range of phenotypes to be screened per breeding cycle, thus accelerating the rate of genetic gain.

Therefore, the TERRA program will emphasize development of innovative phenotyping systems that increase the precision, accuracy and throughput of breeding by developing approaches that can predict terminal phenotypes earlier in the growth cycle.

As shown in Figure 2, ARPA-E seeks to establish multidisciplinary teams to leverage advancements in sensor technologies, computational analytics and low-cost nucleotide sequencing. It is the objective of the program to establish the key intermediate phenotypes related to yield, that can be collected with enough accuracy to predict the growth of an individual plant or population of plants of a particular genetic makeup, and to do so across a high volume (multiple thousands of breeder plots) at the field level.

$$R_t = \frac{ir\sigma_A}{y}$$

$R_t$  (genetic gain over time)

$i$  (selection intensity)

$r$  (selection accuracy)

$\sigma_A$  (genetic variance)

$y$  (years/cycle)

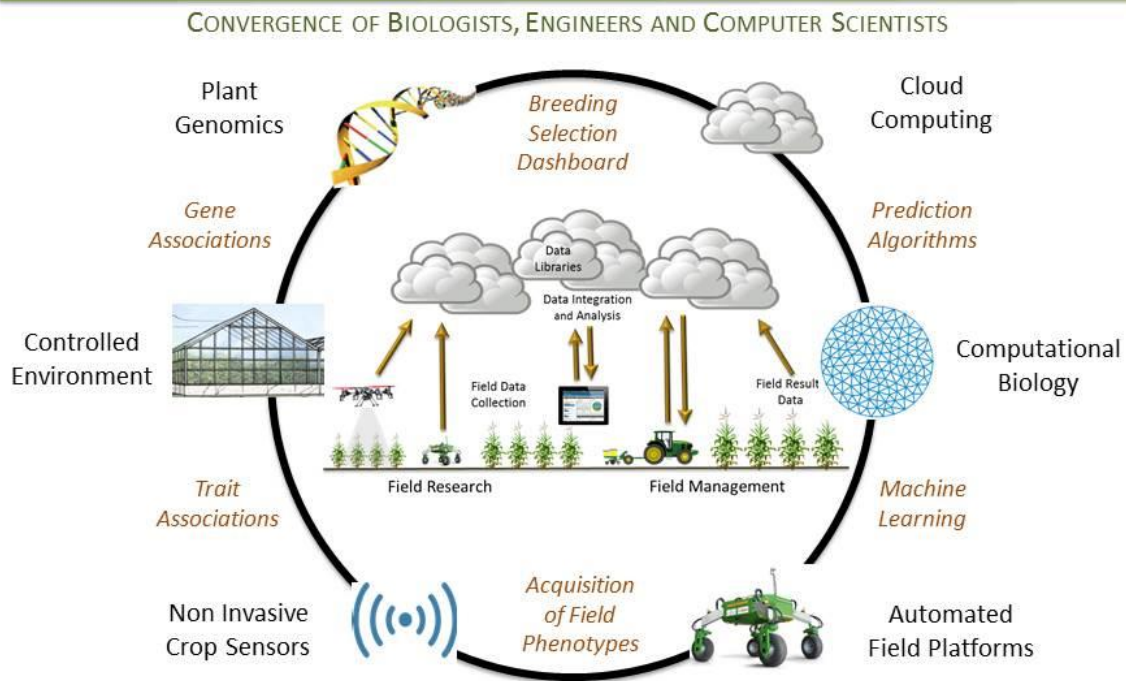
J. Poland, KSU - ARPAAE Workshop 2014

Equation 2: Breeders'

Yield

<sup>40</sup> Jeffrey W White, Pedro Andrade-Sanchez, Michael A Gore, Kevin F Bronson, Terry A Coffelt, Matthew M Conley, Kenneth Feldmann, Andrew French, John Heun, Douglas Hunsaker, Field Crops Research, 2012. 133: p. 101-112.

## ARPA-E Program Vision Integrated Technologies Provides Platform for Innovation



**Figure 2: ARPA-E Advanced Phenotyping Vision<sup>41</sup>**

<sup>41</sup> ARPA-E Plant Phenotyping Workshop. 2014: Chicago, IL.

## D. TECHNICAL CATEGORIES OF INTEREST

Precision phenotyping of bioenergy crops under natural field conditions is a complex systems challenge that requires the integration of multiple scientific and engineering disciplines.

**ARPA-E is primarily interested in applications that propose complete systems solutions** that combine genetics, automation, sensors, and computation into integrated phenotyping platforms. Each platform should be capable of quantifying, modeling, and accurately predicting plant performance in the field and include provisions for data quality control, standardization, and digital communication of data and algorithms to the community. However, applications are also sought for partial system solutions that comprise the key components of the complete integrated phenotyping system: high throughput automated hardware and sensing technology; computational solutions for phenotype selection and prediction; and genomic and bioinformatic genotype by phenotype trait associations. Finally, applications are also sought for programmatic reference data generation and hosting.

### Category 1: Complete Integrated Phenotyping Systems Solutions

ARPA-E seeks applications describing complete solutions that span crop breeding, field automation, plant sensing, genomics, computational analytics and bioinformatics, all of which are critical to the success of the TERRA program. This program should be conducted under both field and controlled environmental conditions, where exploration of genotype-phenotype associations in controlled environments can be used to help predict trait associations in the field.

Applicants should propose automated host platforms and sensor suites, along with the initial phenotypes that will be measured. Applicants should have a plan for selecting to a minimum set of phenotypes necessary to drive yield gain. Based on that selection, the phenotyping platform must be able to evolve further into a simpler, faster, and less costly system. Applicants are expected to demonstrate how their proposed system will meet the in-field throughput (see Section I.E of the FOA) and cost metrics outlined in Section I. G of the FOA.

Category 1 integrated phenotyping systems solutions applicants must include ALL of the elements enumerated below in component Categories 2, 3 and 4.

### Category 2: High Throughput Automated Hardware and Sensing Technologies

ARPA-E seeks applications that develop high throughput automated sensing technologies to acquire phenotypic data in the field across diverse breeding populations, with a preference for energy sorghum. Teams should generate high-fidelity image - and spectrum-based phenotype data and whole-plant assessments throughout the plant's growth cycle for sorghum accessions/lines analyzed in field. This category requires high-precision data on growth rates, biomass accumulation, and the physiological state of each genotype to infer variation in the carbon-system traits of interest. TERRA teams may propose more than one platform solution in order to provide full season and/or complete crop phenotyping capabilities.

Teams are asked to propose the phenotypes they expect to be most predictive of end of season biomass growth/yield. Examples of types of phenotype data, as well as the sensory and platform collection means required to gather the data, are presented in Table 1.

**Table 1:** Examples of candidate representative phenotypes, sensors, and platforms sought from teams submitting in Category 2 (not exhaustive)

<u><b>PHENOTYPES</b></u>	<u><b>SENSORY SYSTEMS</b></u>	<u><b>PLATFORMS</b></u>
<ul style="list-style-type: none"> <li>▪ BIOMASS YIELD</li> <li>▪ CARBON (ENERGY) YIELD PER UNIT TIME, SYNTHETIC INPUT, AND MOISTURE (GENETIC EFFICIENCY)</li> <li>▪ CARBON PARTITIONING TO SINKS</li> <li>▪ GROWTH RATE, HEIGHT, LODGING</li> <li>▪ MATURITY, PHOTOPERIOD SENSITIVITY</li> <li>▪ COMPOSITIONAL TRAITS (NONSTRUCTURAL AND STRUCTURAL CARBOHYDRATES AND OTHER CARBON AND MINERAL COMPOUNDS)</li> <li>▪ PHOTOSYNTHETIC EFFICIENCY</li> <li>▪ DROUGHT TOLERANCE AND WATER USE EFFICIENCY</li> <li>▪ COLD TOLERANCE AND REGROWTH POTENTIAL</li> </ul>	<ul style="list-style-type: none"> <li>▪ ACTIVE REFLECTANCE</li> <li>▪ SPECTRAL INDICES (E.G. NDVI)</li> <li>▪ VISUAL IMAGING FOR MORPHOMETRIC ANALYSIS (STEREO CAMERAS)</li> <li>▪ THERMAL IMAGING</li> <li>▪ HYPERSPECTRAL / MULTISPECTRAL IMAGING</li> <li>▪ FLUORESCENCE MEASUREMENT</li> <li>▪ ENVIRONMENTAL MONITORS (TEMPERATURE, RAINFALL, HUMIDITY, INSOLATION, CO<sub>2</sub>, ETC.)</li> <li>▪ SPATIAL IMAGING (LIDAR, ULTRASONIC)</li> </ul>	<ul style="list-style-type: none"> <li>▪ UNMANNED OR OPTIONALLY PILOTED GROUND VEHICLES, WHEELED OR OTHERWISE</li> <li>▪ GANTRY AND/OR CRANE SYSTEMS</li> <li>▪ CABLE-BASED SYSTEMS</li> <li>▪ SELECT UNMANNED AERIAL VEHICLES (UAV), WITH APPROPRIATE GOVERNMENT APPROVALS</li> </ul>

Applicants should have a plan for selecting to a minimum set of phenotypes necessary to drive yield gain. Based on that selection, the phenotyping platform must be able to evolve further into a simpler and faster system. Applicants are expected to demonstrate how their proposed technologies will meet the in-field throughput and cost metrics outlined in Section I. G of the FOA.

Further, it is understood that some nascent technologies may not be at a development stage to address all of the attributes described above. Consequently, ARPA-E will support the development of novel enabling technologies that could transformationally and significantly contribute to progress towards the objectives in this category.

ARPA-E is particularly interested in key enabling technologies that include, for example:

- Novel sensors (spectrometers, electrochemical sensors, etc.) that provide significantly improved performance and/or lower cost than existing sensor technologies.
- Technologies that enable measurements of individual plants beneath the closed leaf canopy of the crop.
- Technologies that enable non-destructive below-ground field characterization, e.g. root architecture or mass.
- Sensor approaches for soil profile characteristics, soil microbiome analyses.

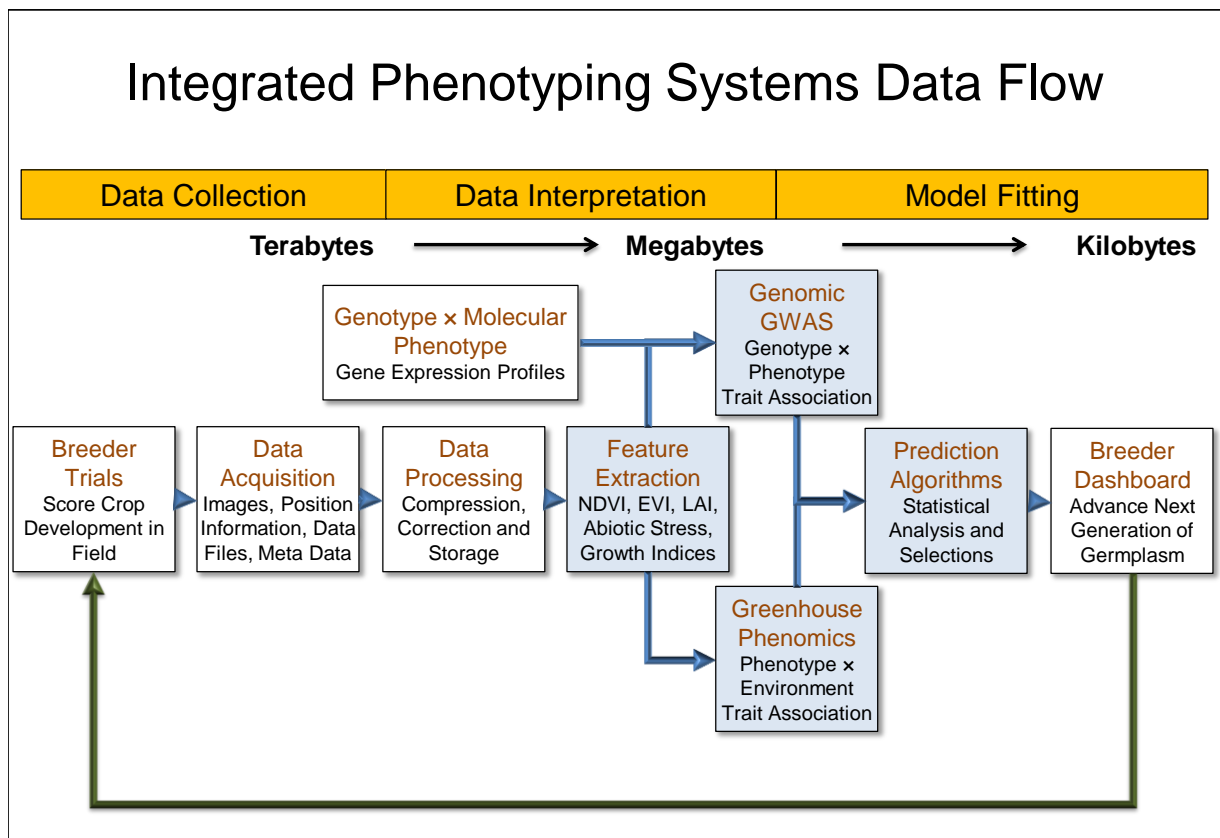
### Category 3: Computational Solutions for Selection and Prediction

ARPA-E seeks applications to develop algorithm-based computational solutions that enable the discovery of the most important and predictive attributes of the crop phenotypic data. Phenotyping systems should encompass automated data collection, data reduction, data interpretation and model fitting (Figure 3) to include plant identification (segmentation of plant from background), feature detection (e.g. height, area, color), and feature analysis (e.g. growth rate).



Teams must develop a robust informatics pipeline that addresses quantitative approaches to image and data processing that must include, at a minimum:

- Calibration: Manage inconsistent orientations and alignments of multiple images, image effects due to sensor noise, and image variance due to the variety of conditions found in the field (e.g., light intensities, air temperatures, and humidity levels). Of particular importance is proper registration and georeferencing of images, in addition to turn-key solutions for converting hyperspectral scans to “data-cubes” of accurate reflectance values.
- Segmentation: Effective methods and algorithms for separating plant foreground from field background. Included in this should be more specialized computer vision algorithms to extract canopy boundaries, for example.
- Feature extraction: Computer vision techniques for identifying inputs to machine learning classification or regression algorithms for phenotyping (below).



**Figure 3:** Integrated Phenotyping Systems Data Flow<sup>42</sup>

As shown in Figure 3, after data acquisition and processing, performers must apply machine learning regression and classification algorithms that operate on multimodal data-streams to extract key features for phenotyping. In addition to the imaging data described above, performers should consider adding other multimodal data streams as inputs. These streams may include text input (such as field reports) or qualitative information such as breeder visual scores (e.g. lodging), or prior beliefs and observations. Applicants should include information concerning the structure of their models, particularly regarding the algorithms employed. For example, phenotyping algorithms might include any number of machine learning or biological models, such as:

<sup>42</sup> ARPA-E Plant Phenotyping Workshop. 2014: Chicago, IL. <http://arpa-e.energy.gov/?q=arpa-e-events/plant-phenotyping-workshop>

- Static Machine Learning Models: Least squares regression, logistic regression, support vector machines, sparse regression, ridge regression, random forests, etc.
- Dynamic Machine Learning Models: Markov models, Gaussian processes, etc.
- Biophysical Models: Functional growth analysis, canopy models, etc.

A crucial component is expected to be cloud computation and parallel analysis (possibly in the field) of the very large data sets necessary to train the algorithms for more accurate phenotypic prediction. Data will be derived from a series of sensors (thermal, hyperspectral, moisture, CO<sub>2</sub>, fluorescence, etc.) that collect a high-volume “data-cube” of information. The data size for integrated systems has been estimated to be up to terabytes per 24 hour run. Finding combinations of algorithms and hardware platforms that apply to this “data-at-scale” will be a research challenge that must be addressed.

As the program matures, applicants will be expected to identify the dimensions of the data that are most predictive for phenotypic quantities of interest (this would be an example of new index discovery by principled data reduction from the full volume of collected information). This will require dimensional reduction and feature selection algorithms. It is expected that the relevant data will not have a simple linear structure, making a straightforward Principle Component Analysis (PCA) inappropriate. Applicants are expected to develop more refined, non-linear dimensional reduction algorithms (such as manifold learning, etc.) to achieve the goals of this program.

Further, it is understood that some nascent technologies may not be at a development stage to address all the attributes described above. Consequently, ARPA-E will support the development of novel technologies that could transformationally and significantly contribute to progress towards the objectives in this category.

ARPA-E is particularly interested in key enabling technologies that include, for example, data-processing algorithms and predictive models that have unique capabilities for evaluating plant performance.

#### Category 4: Genetics, Genomics and Bioinformatics

ARPA-E seeks applications to develop high-resolution genetic maps to provide the foundation for elucidating genotype-phenotype associations. The applicants should endeavor to correlate genetic markers with physiological phenotypes, using applicable tools such as genotyping-by-sequencing (GBS), whole genome re-sequencing, and other -omic (in particular transcriptomic) profiling.

The application must integrate these high-throughput sequencing derived genotypes with phenotypic data collected in the field and under controlled-environments (greenhouse and laboratory) to drive statistical genetics (QTL, eQTL, GWAS, genetic prediction, etc.) analyses focusing on traits associated with carbon productivity, for example: photosynthesis, accumulation and partitioning, plant idiomorph (height, leaf angle, and tillering), plant composition, and terminal yield (grain, sugar, or biomass).

TERRA Category 4 teams are expected to generate a number of significant resources, such as:

- Genomic characterization and high-quality molecular markers suitable for whole-genome predictions and genome selection, identified as single nucleotide polymorphisms (SNPs) through genotyping-by-sequencing (GBS) and analysis of genomic re-sequencing data obtained from sorghum individuals.
- Identification of putative functional SNPs to enable increased resolution in marker-assisted breeding and provide further mechanistic insight into such biomass traits.
- Statistical genetics resources and outcomes (or analyses) including GWAS, QTL, eQTL elucidation of epistatic interactions and epigenetics.
- Controlled environment gene expression data from a diverse collection of genotypes from the juvenile stage to harvest; enabling elucidation of eQTLs and correlations of gene expression, juvenile traits, and terminal yields.

- Detailed phenotypic data determined through controlled environment high-throughput phenomics platforms and field based platforms developed during the program throughout the entire growing season.

Further, it is understood that some nascent technologies may not be at a development stage to address all the attributes described above. Consequently, ARPA-E will support the development of novel technologies that could transformationally and significantly contribute to progress towards the objectives in this category.

ARPA-E is particularly interested in key enabling technologies that include, for example, algorithms for improving predicting genes, regulatory elements, or genetic loci associated with biomass or other yield phenotypes.

### Category 5: Programmatic Reference Data Generation and Hosting

The successful applicant for Category 5 will generate and host for TERRA teams reference phenotyping data through an energy sorghum field test plot monitored by a state-of-the-art field phenotyping system provided by ARPA-E.

To facilitate the development of accurate, predictive phenotyping algorithms and support all program teams, ARPA-E will collaborate with the Category 5 awardee to make available to other awardees a high-fidelity dataset collected on sorghum field plots throughout the program. This will be accomplished by way of an automated platform that will deliver raw sensory data to all teams starting in 2015. This dataset can be used in advance of and in conjunction with any data collected by the teams using their own field platforms.

Data from a wide array of sensors on the phenotyping system will deliver high-resolution imagery and other sensory products. The data will be transferred continuously to an on-line, awardee-accessible system within 48 hours of collection. This system will offer a rich amount of spectral data, and will offer per-pixel resolution and positional accuracy suitable for plant-level (sorghum) specificity. The high degree of sensor instrumentation on this system is expected to provide insight into the use of a diverse array of sensors, and this insight can inform the design of the low-cost field sensing platforms to be developed by TERRA teams.

Metadata will be appended to all sensory products for the purpose of timestamping and georeferencing all collected information, with standards clearly documented for the performer teams. With this metadata and a fairly continuous stream of environmental information (wind speed, temperature, etc.), the environmental context of each data product can be determined. The data generated by the reference platform will be available to all TERRA teams and all necessary interface specifications for accessing and working with the data will be provided.

The reference team, with the substantial involvement of the ARPA-E Program Director, will install, operate and maintain a reference field phenotyping system provided by ARPA-E.

ARPA-E is selecting and will provide to the reference team a state-of-the-art field phenotyping system (hereafter referred to as Government Furnished Equipment, GFE). The GFE will have the features provided in Table 2.

**Table 2.** Features of the GFE field phenotyping system.

PERFORMANCE CHARACTERISTIC	SPECIFICATION
FORM FACTOR	<ul style="list-style-type: none"> <li>▪ GANTRY SYSTEM, SENSOR BAY HEIGHT RANGE 0-6 METERS</li> </ul>
RESOLUTION AND ACCURACY	<ul style="list-style-type: none"> <li>▪ PAYLOAD BAY POSITIONAL ACCURACY OF 1-5 CM OR BETTER IN X, Y AXES, AND 0.5 CM OR BETTER IN Z-AXIS</li> <li>▪ IMAGERY (HYPERSPCTAL, IR, RGB) RESOLUTION OF 1 MM OR BETTER</li> </ul>

<p><b>SENSOR SUITE</b></p>	<ul style="list-style-type: none"> <li>▪ HYPERSPECTRAL IMAGER (400-2500 NM RANGE OBJECTIVE, ~700-2000 NM THRESHOLD)</li> <li>▪ THERMAL INFRARED (IR)</li> <li>▪ DEDICATED NDVI SENSOR</li> <li>▪ LIDAR IMAGER</li> <li>▪ STEREO RGB CAMERAS</li> <li>▪ ACTIVE REFLECTANCE SENSOR</li> <li>▪ FLUORESCENCE MEASUREMENT SYSTEM</li> <li>▪ ENVIRONMENTAL SENSORS – RAINFALL, TEMP, HUMIDITY, LIGHT INTENSITY, CO<sub>2</sub></li> <li>▪ POSSIBLE SIDE-LOOK OR SLANT-ANGLE RGB AND LIDAR SENSORS DISTINCT FROM PRIMARY LOOK ANGLE</li> </ul>
<p><b>AREA AND COVERAGE RATES</b></p>	<ul style="list-style-type: none"> <li>▪ MINIMUM 1 ACRE FOR FIELD TEST PLOT, WITH ADDITIONAL SURROUNDING AREA</li> <li>▪ 20 METER USABLE WIDTH FOR ASSAYING PLANTS</li> <li>▪ PRIMARY SENSORS: COLLECTED 2-4x DAILY, 2x/WEEK</li> <li>▪ ENVIRONMENTAL SENSORS: COLLECTED EVERY 15 MINUTES</li> </ul>

The reference team is expected to have: space and personnel available to install the GFE in the 2015 growing season; for the gantry form factor, concrete foundations poured in the field; a dedicated field site with at least 20 meters of usable width and at least 200 meters long; a power supply (10s of kW expected) to the field phenotyping system; and all relevant construction and safety permits for operation of this system. Subject to the specific provisions of the ARPA-E award, the GFE may be a permanent installation at the reference team’s site, with the system remaining with the reference team at the end of the project.

The reference team award will provide for the operation of the GFE for the duration of the project, including obtaining any necessary repairs and modifications from the GFE provider under a separate budget. As part of the GFE installation, training will be provided by the manufacturer. In collaboration with ARPA-E, the reference team will host at least one field day each year to demonstrate the GFE to the other TERRA performers.

The reference team will provide the TERRA teams access to the raw data and computational workflows.

ARPA-E has concluded that progress in phenotyping will be accelerated if raw data and computational workflows can be shared as widely and quickly as possible among the TERRA teams. For this reason the reference team is expected to make available the raw and processed data obtained using the GFE, along with all of the metadata and environmental data relevant to the interpretation and use of the measurements. This data will be provided in a format established by a data and computational standards committee (see below). ARPA-E anticipates approximately 1 Terabyte of data per day to be generated by the GFE, such that over a growing season hundreds of Terabytes may be generated, and over the course of the program on the order of several Petabytes may be generated. Given the large data volumes that will be produced, the data will be stored on a server to which TERRA teams can obtain a terminal window in order to access and analyze the data, rather than needing to move it to a local computer storage system.

In addition to providing access to the raw data, the reference team is expected to establish a data analysis pipeline similar to the one shown in Figure 3, and make its use (though not necessarily the source code of each component) available to performers in the TERRA program.

It is possible that no one website or software product will be uniquely poised to address the large data and computational demands of sensor streams, the complicated analytical tools of bioinformatics, and maintain a collaborative user environment. Applicants are therefore encouraged to partner with experts in particular areas relating to components of the data storage and analysis pipeline.

Finally, the GFE system provided by ARPA-E will include software not only for gantry and sensor controls and data handling, but also for phenotype extraction. The reference team will be expected to provide results from the GFE system

to provide a clear benchmark to TERRA performers in phenotype extraction from a variety of types of images. This should take the form of documenting the phenotypes extracted from a set of raw images generated by the GFE sensors. At this time, it is the intent of ARPA-E for the Category 5 awardee to release all data to the public at an appropriate time determined by ARPA-E that is consistent with DOE regulations.

The reference team will convene a data and computational standards committee.

The standards committee – selected in collaboration with the ARPA-E TERRA Program Director - will create the foundation for the straightforward sharing and processing of data among the numerous participants of diverse backgrounds involved in the TERRA program. The standards committee will be composed of experts in data fusion (the combination of multiple streams of data of different formats, time increments, context, etc), image analysis and feature extraction, bioinformatics, and computational pipelines, and have representatives from each of the other TERRA performers.

In collaboration with the data standards committee the reference team should provide a dataset focusing on derived phenotypes, genotypes, and measured environmental data resolved to individual plants or plots. It is most important to provide derived parameters that require extensive computation to analyze. For example, leaf count or LAI can be represented as single numbers though they may be based on numerous images. For genotypes, it is most relevant to share SNP data sets rather than whole genome sequences for comparison between individuals. The purpose of this data table is to lower computational barriers for computer scientists, statisticians, and biologists seeking to find determinants of final phenotypes and create predictive models.

The reference team must have access to a controlled chamber / greenhouse phenotyping system.

A controlled chamber / greenhouse system will be required to enable direct comparisons and correlations between lines grown under controlled environment and field conditions. In addition, the greenhouse will allow year-round operation. Given the growth potential of energy sorghum (heights exceeding 5m by maturity) and the corresponding root system needs, ARPA-E expects the greenhouse studies to focus primarily on the early stages of plant growth.

## E. TECHNICAL PERFORMANCE TARGETS

Advanced plant phenotyping platforms should increase the utility (information content), time resolution (seasonal, daily), and amount (# phenotypes × # populations) of crop phenotyping data in order to accelerate energy crop breeding. The TERRA program’s primary objective is to enable breeders and geneticists to identify crop phenotypes/traits of interest, genetic architecture of traits, and alleles in breeding pools that can be used to accelerate genetic gain relative to current breeding approaches.

### Category 1: Complete Integrated Phenotyping Systems Solutions

#### Primary Technical Targets

ID	DESCRIPTION	TARGET
1.1	COMPLETE INTEGRATED PHENOTYPING SYSTEMS SOLUTIONS TARGETS INCLUDE ALL OF THE TARGETS GIVEN BELOW FOR COMPONENT CATEGORIES 2, 3, AND 4.	
1.2	TOTAL SYSTEM COST	<\$20K/COVERED HECTARE, THREE YEAR PAYBACK

Explanation of technical targets:

1.2: Total system cost (amortized capital cost + operating cost) is less than \$20k per covered hectare for a three-year payback period. For additional information on this cost metric see Section I. G of the FOA.

## Category 2: High Throughput Automated Hardware and Sensing Technologies

### Primary Technical Targets

ID	DESCRIPTION	TARGET
2.1	PHENOTYPIC SELECTIVITY	TOP 10%
2.2	TECHNICAL REPEATABILITY	$R^2 > 0.99$
2.3	PLATFORM SCALABILITY	>50 HECTARES

### Secondary Technical Targets

ID	DESCRIPTION	TARGET
2.4	PHENOTYPE DATA CAPTURE RATE	2 TIMES/DAY AND 2 TIMES/WEEK
2.5	ENVIRONMENTAL TOLERANCE	OPERATIONAL RANGE 32-110 °F, 30 MPH WIND GUSTS, DUST/RAIN PROTECTED TO IEC 60529 RATING OF IP54
2.6	IMAGE RESOLUTION	SUITABLE FOR PLANT-LEVEL SPECIFICITY, PER PIXEL RESOLUTION 1 CM <sup>2</sup> OR LESS

Explanation of technical targets:

2.1: The platform(s) should have sufficient resolution, accuracy, and precision to be able to identify the top 10% of phenotypes/traits from the analyzed populations in the relevant range of growing environments.

2.2: For phenotypic measurements taken one right after another obtain values with  $R^2 > 0.99$ .

2.3: The platform(s) needs to operate over an area of 50 hectares.

2.4: The platform(s) should ideally be able to measure an area of >50 hectares multiple times per day, as well as multiple times per week, in order to obtain phenotypic data at much higher temporal frequency than is presently collected.

2.5: The platform(s) need to be able to operate outdoors in a range of environmental conditions.

2.6: Sensors should be able to resolve the phenotypic information of individual plants, requiring per-pixel resolution of no more than 1 cm<sup>2</sup>, and preferably less.

## Category 3: Computational Solutions for Selection and Prediction

### Primary Technical Targets

ID	DESCRIPTION	TARGET
3.1	ABOVE GROUND BIOMASS PREDICTION ACCURACY	$R^2 > 0.97$
3.2	OTHER PHENOTYPE PREDICTIONS	SIGNIFICANT IMPROVEMENT VS EXISTING STATE OF THE ART (SEE EXPLANATION BELOW)
3.3	DETERMINE THE PHENOTYPES THAT CORRELATE WITH THE VARIATION IN	FIND PHENOTYPES THAT ACCOUNT FOR 95% OF THE VARIATION IN TERMINAL BIOMASS

	TERMINAL BIOMASS YIELD	YIELD
--	------------------------	-------

### Secondary Technical Targets

ID	DESCRIPTION	TARGET
3.4	DATA PROCESSING SPEED/TURNAROUND	1 WEEK'S PHENOTYPING DATA WITHIN 48 HRS

Explanation of technical targets:

3.1: Above ground biomass should be predicted from sensor data and then compared with ground truth measurements. Above ground biomass predictions from emergence to full maturity are required.

3.2: While ARPA-E is targeting a high value of  $R^2$  for above ground biomass prediction, numerous other phenotypes (see Table 1 above for examples) are expected to correlate with terminal biomass yield, not all of which can be determined with such high prediction accuracy. Ground truth measurements for other phenotypes are also required.

3.3: In addition to the extraction of phenotypes from sensor data, ARPA-E also expects applicants in Category 3 to determine *which* phenotypes correlate most highly with the variation in terminal biomass yield. Given the fact that phenotypes will be collected frequently and are in the context of environmental conditions, finding the truly important phenotypes will require data-intensive algorithms. The determination of the phenotypes that are most important for producing high-yielding plants will ultimately allow for the design of low-cost sensor and computation platforms that include components targeting only the most important phenotypes. In addition, ARPA-E is particularly interested in the correlation of phenotypes that can be identified early in the season with terminal biomass yield in order to reduce the length of the breeding cycle.

3.4: Algorithms should be parallelizable and capable of being run on a standard Hadoop cluster to produce an analysis of one week's data within 48 hrs.

## Category 4: Genetics, Genomics and Bioinformatics

### Primary Technical Targets

ID	DESCRIPTION	TARGET
4.1	GENETIC BASIS FOR BIOMASS ACCUMULATION	EXPLAIN AT LEAST 70% OF THE GENETIC COMPONENT OF THE PHENOTYPIC VARIANCE FOR BIOMASS ACCUMULATION

### Secondary Technical Targets

ID	DESCRIPTION	TARGET
4.2	GENOTYPE TO PHENOTYPE ASSOCIATIONS	SCREEN THOUSANDS OF ACCESSIONS AND/OR BREEDER LINES AND IDENTIFY PUTATIVE FUNCTIONAL SNPs AND MOLECULAR MARKERS LINKED TO PHENOTYPIC TRAITS OF INTEREST, $p < 5 \times 10^{-6}$

Explanation of technical targets:

4.1: Elucidate the genetic architecture and heritability of trait/phenotypes that impact biomass yield ( $P=G \times E$ ). Determine the portion of variation in field biomass accumulation that is explained by genetics. Explain at least 70% of the genetic component of the phenotypic variance for biomass accumulation in a population designed for technology validation, based on field phenotypic data collected from more than one environment.

4.2: Identify molecular markers and associate traits that will significantly improve the efficiency of breeder efforts to increase biomass genetic gain. Identify field-validated markers corresponding to functional variation in specific genes/traits that modify biomass yield.

## Category 5: Programmatic Reference Data Generation with GFE and Data Hosting

### Primary Technical Targets

ID	DESCRIPTION	TARGET
5.1	FIELD SITE	>1 ACRE SUITABLE FOR THE INSTALLATION AND OPERATION OF THE GFE SYSTEM
5.2	GENERATE REFERENCE PHENOTYPIC DATASETS	APPLICABLE PERFORMANCE METRICS DESCRIBED IN CATEGORY 2, EXCLUDING 2.3
5.3	GENERATE REFERENCE COMPUTATION DATASETS	SAME PERFORMANCE METRICS DESCRIBED IN CATEGORY 3
5.4	GENERATE REFERENCE GENOMIC DATASETS	SAME PERFORMANCE METRICS DESCRIBED IN CATEGORY 4
5.5	CREATE DATA STANDARDS AND SOURCE COMPUTATION PIPELINE AVAILABLE TO TERRA PERFORMERS.	HOST DATA PLATFORM CAPABLE OF PROCESSING AND STORING PETABYTES OF PHENOTYPIC DATA

Explanation of technical targets:

5.1: The field site made available by an applicant to Category 5 should be >1 acre of dedicated breeding area suitable for permanent installation of a gantry system, with at least 20 x 200 meters suitable for growing sorghum and appropriate buffer areas. The site should have irrigation, a power supply, a uniform grade, high quality soil, and appropriate site security.

5.2: The GFE will have the hardware specifications to meet the requirements of Category 2 (except for 2.3); a recipient of a Category 5 award should therefore focus on carrying out the tasks described by the Category 2 metrics.

5.5: See the discussion under Category 5 in Section I.D of the FOA for more information on data standards and the data hosting and computation platform.

## F. APPLICATIONS SPECIFICALLY NOT OF INTEREST

The following types of applications will be deemed nonresponsive and will not be reviewed or considered (see Section III.C.2 of the FOA):

- Applications that fall outside the technical parameters specified in Section I.E of the FOA
- Applications that were already submitted to pending ARPA-E FOAs.
- Applications that are not scientifically distinct from applications submitted to pending ARPA-E FOAs.
- Applications for basic research aimed at discovery and fundamental knowledge generation.
- Applications for large-scale demonstration projects of existing technologies.
- Applications for proposed technologies that represent incremental improvements to existing technologies.



- Applications for proposed technologies that are not based on sound scientific principles (e.g., violates a law of thermodynamics).
- Applications for proposed technologies that are not transformational, as described in Section I.A of the FOA and as illustrated in Figure 1 in Section I.A of the FOA.
- Applications for proposed technologies that do not have the potential to become disruptive in nature, as described in Section I.A of the FOA. Technologies must be scalable such that they could be disruptive with sufficient technical progress (see Figure 1 in Section I.A of the FOA).
- Applications that are not scientifically distinct from existing funded activities supported elsewhere, including within the Department of Energy.
- Applications in Categories 1-4 that propose to work with maize, or other crops that don't have the following characteristics:
  - A demonstrated yield of >25 dry metric tons biomass/hectare/year on at least a one hectare plot in the continental United States,
  - Existing infrastructure and grower expertise,
  - At least one published and annotated genome sequence.
- Applications in Category 5 that propose to work with a crop other than energy sorghum.

## G. ECONOMIC ANALYSIS

In order to achieve these ambitious goals, the value provided by phenotyping platforms in terms of advanced traits must be greater than their capital and operating cost at a reasonable payback period. A standard seed industry trait valuation model was utilized to set the cost and throughput metrics for this FOA. The value of deploying an in-field phenotyping system was calculated by assuming an additional 2% year over year yield gain due to the use of a high-throughput phenotyping system, \$50/ton energy crop price, set field input costs and a 5 year new seed adoption rate by the farmer. This value was split so that one-third is assigned to the seed company and two-thirds to the farmer. The "total system" includes robotics, sensors, software, data storage, technical and operating labor and the total allowable capital and operating cost was calculated with a 3 year payback period to the breeder. This calculation does not take into account the value of additional traits that are likely to be identified utilizing these phenotyping platforms; it only takes into account the value from increased yield.

Based on this preliminary economic analysis, the economic metric for this FOA is set at a total system cost of less than \$20,000 per hectare covered for field phenotyping systems intended to be used by breeders and bioinformaticians on advanced field research stations. For example, a 50-hectare breeding station system cost should be less than \$1M which, again, is for a three-year payback period.