Emerging Avenues for Utilization of Exotic Germplasm

Culing Wang  
*Henan University of Science and Technology*

Songlin Hu  
*Iowa State University*

Candice Gardner  
*United States Department of Agriculture*

Thomas Lubberstedt  
*Iowa State University, thomasl@iastate.edu*

Follow this and additional works at: [http://lib.dr.iastate.edu/agron_pubs](http://lib.dr.iastate.edu/agron_pubs)

Part of the [Agronomy and Crop Sciences Commons](http://lib.dr.iastate.edu/agron_pubs/298).

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/agron_pubs/298](http://lib.dr.iastate.edu/agron_pubs/298). For information on how to cite this item, please visit [http://lib.dr.iastate.edu/howtocite.html](http://lib.dr.iastate.edu/howtocite.html).

This Article is brought to you for free and open access by the Agronomy at Iowa State University Digital Repository. It has been accepted for inclusion in Agronomy Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Emerging Avenues for Utilization of Exotic Germplasm

Abstract
Breeders have been successful in increasing crop performance by exploiting genetic diversity over time. However, the reported annual yield increases are not sufficient in view of rapid human population growth and global environmental changes. Exotic germplasm possesses high levels of genetic diversity for valuable traits. However, only a small fraction of naturally occurring genetic diversity is utilized. Moreover, the yield gap between elite and exotic germplasm widens, which increases the effort needed to use exotic germplasm and to identify beneficial alleles and for their introgression. The advent of high-throughput genotyping and phenotyping technologies together with emerging biotechnologies provide new opportunities to explore exotic genetic variation. This review will summarize potential challenges for utilization of exotic germplasm and provide solutions.

Disciplines
Agronomy and Crop Sciences

Comments

Rights
Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.
Breeders have been successful in increasing crop performance by exploiting genetic diversity over time. However, the reported annual yield increases are not sufficient in view of rapid human population growth and global environmental changes. Exotic germplasm possesses high levels of genetic diversity for valuable traits. However, only a small fraction of naturally occurring genetic diversity is utilized. Moreover, the yield gap between elite and exotic germplasm widens, which increases the effort needed to use exotic germplasm and to identify beneficial alleles and for their introgression. The advent of high-throughput genotyping and phenotyping technologies together with emerging biotechnologies provide new opportunities to explore exotic genetic variation. This review will summarize potential challenges for utilization of exotic germplasm and provide solutions.

The Value of Exotic Germplasm in a Changing World
Modern agriculture increased food production in a remarkable way, but one in nine persons remains undernourished and unable to consume sufficient protein and energy [1]. Breeding activities have achieved a linear increase at an average rate of 32 million metric tons per year [2]. However, because the global population is projected to exceed 9 billion by 2050, an average annual increase of 44 million metric tons per year is needed [3,4]. Moreover, altered consumption habits and the production of raw materials for biofuels and other products may require a further increase in productivity, with the added challenge of global environmental changes. Climate change increases the frequency of unpredictable extreme events such as extreme temperatures, unexpected timing, and quantity of rainfall (drought or flooding), and the incidence and movement of new pests, plant pathogens, and invasive species [5]. Development and adoption of stress-resilient crop cultivars is imperative to ensure sustainability of agricultural production and global food security [6]. Breeders successfully improved germplasm performance with continuously increased yield potential and adaptation to mechanization and new agriculture management practices, and these have contributed to dramatically increased agricultural productivity since the middle of the 20th century [7]. Some beneficial genes from exotic germplasm such as the SUB1 gene in rice, the TY-1 gene in tomato, and late blight R genes in potato, have been successfully used in breeding programs and many productive or resistant varieties have been released [8–10]. However, since the mid-1990s progress in conventional plant breeding has varied by region and by crop, and the relative rate of yield gain for the major food crops has decreased over time, together with evidence of upper yield plateaus in some of the most productive domains [11]. This is partially due to the narrow genetic base of elite germplasm in major crops because modern intensive breeding practices have exploited a very limited fraction of the available biodiversity. Although selection by plant

© 2017 Elsevier Ltd. All rights reserved.
breeders has resulted in increased genetic variation in some cases, genetic diversity within elite cultivars of major crops is generally limited compared to the overall available variability within those species and their relatives worldwide [12,13].

With intense focus and overuse of elite plant materials, ~75% of the genetic diversity of domestic agricultural crops has disappeared and 60,000 plant species could be lost by 2025 [14]. The narrowing genetic base of elite germplasm increases genetic vulnerability to new biotic and abiotic stresses, as well as to unpredictable environmental conditions [15]. Improved access to and more efficient utilization of a broad range of genetic variation that can be explored for novel traits would enable scientists to better address these challenges [6]. Exotic germplasm accessions including landraces, wild species, and wild relatives exhibit substantial genetic diversity, including resistance to stress environments, increased yield stability, food quality, and reduced dependence on agronomic inputs, and this diversity may have been lost in elite genetic pools by selective processes or because of access problems [16–18]. Moreover, rare alleles or allele combinations within exotic germplasm may provide important biological clues about adaptation [19]. Efficient exploitation of exotic germplasm in changing environments depends on (i) how well exotic germplasm has been collected, evaluated, and documented; (ii) how rapidly breeders can identify useful genetic variation; and (iii) how efficiently beneficial genetic variation can be incorporated into elite germplasm.

Challenges in Using Exotic Germplasm
Seeds of landraces, wild relatives of crop species, and modern varieties have been consciously collected and stored in genebanks worldwide, and serve as the important reservoirs of agricultural biodiversity [5]. To date, about 7.4 million accessions have been conserved in more than 1750 genebanks [20]. These accessions are accessible worldwide for the breeding community (Table 1). However, exotic germplasm is not sufficiently utilized by plant breeders because of a lack of efficient strategies to identify and transfer beneficial alleles into elite germplasm. From a genebank perspective, the key challenges include how to (i) be more efficient for the conservation and evaluation of a huge number of accessions with limited resources, and (ii) facilitate sharing accessions and information among genebanks and breeders. From a breeder’s perspective, it is difficult to (i) evaluate and

Table 1. International Genebanks and Information Networks for Main Crops

<table>
<thead>
<tr>
<th>Genebank or information network</th>
<th>URLs</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Rice Research Institute (IRRI)</td>
<td><a href="http://iri.org">http://iri.org</a></td>
</tr>
<tr>
<td>Bioversity International</td>
<td><a href="http://www.bioversityinternational.org">www.bioversityinternational.org</a></td>
</tr>
<tr>
<td>International Center for Tropical Agriculture (CIAT)</td>
<td><a href="https://ciat.cgiar.org">https://ciat.cgiar.org</a></td>
</tr>
<tr>
<td>International Maize and Wheat Improvement Center (CIMMYT)</td>
<td><a href="http://www.cimmyt.org">www.cimmyt.org</a></td>
</tr>
<tr>
<td>International Potato Center (CIP)</td>
<td><a href="http://cipotato.org">http://cipotato.org</a></td>
</tr>
<tr>
<td>International Center for Agricultural Research in the Dry Areas (ICARDA)</td>
<td><a href="http://www.icarda.org">www.icarda.org</a></td>
</tr>
<tr>
<td>The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)</td>
<td><a href="http://www.icrisat.org">www.icrisat.org</a></td>
</tr>
<tr>
<td>International Institute of Tropical Agriculture (ITA)</td>
<td><a href="http://www.ita.org">www.ita.org</a></td>
</tr>
<tr>
<td>The Africa Rice Centre – Africa Rice (WARDA)</td>
<td><a href="http://africanrice.org">http://africanrice.org</a></td>
</tr>
<tr>
<td>The World Vegetable Centre (AVRDC)</td>
<td><a href="http://avrdc.org">http://avrdc.org</a></td>
</tr>
<tr>
<td>The Global Gateway to Genetic Resources (Genesys)</td>
<td><a href="http://www.genesys-pgr.org">www.genesys-pgr.org</a></td>
</tr>
<tr>
<td>The Germplasm Resources Information Network (GRIN-Global)</td>
<td><a href="http://www.ars-grin.gov">www.ars-grin.gov</a></td>
</tr>
<tr>
<td>European Cooperative Program for Plant Genetic Resources (ECPGR)</td>
<td><a href="http://www.ecpgr.cgiar.org/resources/germplasm-databases">www.ecpgr.cgiar.org/resources/germplasm-databases</a></td>
</tr>
</tbody>
</table>
discover valuable materials among a ‘sea of seeds’, (ii) identify favorable alleles or haplotypes, and (iii) incorporate beneficial gene complexes into breeding materials considering issues such as maladaptation (e.g., flowering time asynchrony), linkage drag, and multigene introgression.

Collection and Conservation of Germplasm
Successful exploitation and utilization of exotic germplasm requires effective collection, maintenance, and evaluation of germplasm. Over the past several decades, genebanks have collected large numbers of samples for ex situ collections, complementing by in situ conservation efforts. Genebanks managers must consider: (i) sampling strategies to represent the breadth of genetic and geographic diversity for a crop and its wild relatives; (ii) storage condition requirements to provide for maximum longevity of viability; (iii) the number of accessions that can be collected and maintained, considering resources available for storage conditions, field or greenhouse facilities, and other factors; (iv) how often accessions need to be rejuvenated, considering their reproductive biology and need for controlled pollination; and (v) how to effectively identify duplicates [19]. It is estimated that <30% of the total number of accessions are distinct [20]. Finally, genebanks must take measures to minimize negative effects on the collection as a result of genetic drift, unconscious selection, contamination (unwanted pollen flow, unintentional mixing of samples), or other sources of error.

Characterization and Evaluation of Accessions
Many genebanks find it difficult to respond effectively to requests for accessions or information relevant to breeding objectives because of inadequate characterization and evaluation of germplasm in their collections. They strive to obtain reliable passport information about the provenance, origin, or source of those samples when they are collected. Phenotypic characterization usually focuses on a few morphological characters that are highly heritable and easy to access, such as the number of leaves, flowering, and maturity, based on internationally agreed crop-specific standards. Comprehensive characterization of large collections for multiple traits requires large field or greenhouse space at high costs for facilities and labor, beyond the current budget limitations of genebank managers. As a result, only limited information on the phenotypic performance of accessions in genebanks is available, particularly for complex traits with strong genotype by environment interactions, or with low heritability.

Genotypic information is very useful to characterize genebank accessions, evaluate genetic diversity, and to eliminate duplicates. Genotypic characterization of genebank collections has been undertaken for complete collections (e.g., the Dutch lettuce collection characterized with three amplified fragment length polymorphism (AFLP) primer combinations [21]), or a subset of accessions has been characterized in great depth (e.g., 50 rice varieties characterized at 6.5 M SNP loci [22]). The ~2800 maize inbred lines of the US Maize Collection have been genotyped using genotyping-by-sequencing [23]. By 2020 the genomes of all the major commodities under research by CGIAR will have been sequenced [24]. There is an urgent need to develop strategies and protocols to build a well-characterized database of collections including both genotypic and phenotypic information. Given budgetary constraints, there are concerns about whether existing collections should be maintained or expanded, or whether it would be better to reduce the number of accessions maintained and devote funding to either DNA-based or phenotypic characterization. While debated by some, it would be extremely foolish to reduce the diversity of a collection based on sequence information that is not linked to coding regions associated with trait functions, or non-coding regions important for gene regulation. Other crucial questions include: what is the desirable size of a collection, and based on what criteria? What is the relative value of a large collection without characterization, versus a smaller, well-characterized collection?
Evaluation of accessions is important to discover useful sources and predict the genetic potential and breeding value of the currently underutilized materials in genebanks. Phenotypic evaluation has been successful in identifying beneficial materials for simple traits in genebanks. However, many traits of interest are quantitative and multigenic, with low heritability and strong genotype by environment interactions, and this makes phenotypic evaluation challenging. Moreover, some important traits are only expressed at particular developmental stages, in specific environments, or in response to particular biotic or abiotic stresses, and are not easily captured or are not included in phenotypic evaluation schemes, such as hybrid vigor [25]. It is difficult, if not impossible to predict, which accession has the potential to face future (unknown) stress conditions in the absence of relevant data and analyses [15]. Thus, while it is important to develop characterized collections to facilitate utilization, it is a challenge to do this for a large number of traits and environments. In conclusion, maintaining large collections increases the chance of finding rare alleles or allele combinations for emerging phenotypes.

Availability of Genebank Materials and Information
The biggest challenge faced by genebanks perhaps is to identify and acquire useful materials and to make them available to users worldwide quickly. With the development of high-throughput genomics and phenomics, more genotypic and phenotypic information related to accessions, as well as passport information, pedigree relationships, and agronomic performance in diverse environments, can now be deposited in genebanks. How can genebanks make good use of these data, and capture value from massive amounts of information to predict the potential value of accessions? How can researchers or breeders find what they need in which genebank easily? It will be necessary to develop a user-friendly germplasm cooperation platform to organize, analyze, integrate different types of data, share information among genebanks, and meet the needs of researchers and breeders around the world. Phytosanitary restrictions, importation requirements, and national and international laws governing access present additional obstacles to international germplasm exchange [26]. An example of an effort to address a cooperative information platform is the DiSeek Initiative, which seeks to form a collaborative network of genebanks, breeders, plant and crop scientists, and database and computational experts [27].

Maladaptation
Utilization of exotic germplasm is often hampered by maladaptation. Unadapted germplasm may not grow normally outside its native region, hampering direct phenotypic trait evaluation at genebank locations. Beneficial alleles may not be expressed or may be masked by other maladaptation genes [28,29]. For example, tropical maize requires short-day conditions for flowering, and is delayed under temperate ‘long-days’, leading to poor ear development, tall plants, poor root and stalk strength, and high ear placement [30]. Therefore, useful tropical maize germplasm cannot readily be identified in temperate environments because of maladaptation. In addition, it is difficult to directly introduce maladapted exotic germplasm into breeding programs. It is thus important for plant breeders to develop efficient strategies to overcome these maladaptation problems, and to establish reliable protocols for exotic gene introgression.

Identification and Utilization of Beneficial Alleles
When an exotic accession is identified as containing desirable traits, such as biotic or abiotic stress tolerance, the underlying alleles or haplotypes need to be transferred into elite germplasm for trait improvement. Introgression of major genes from exotic germplasm can be performed by repeated backcrosses into elite germplasm for multiple generations using either phenotypic or genotypic selection to remove undesirable exotic genome contributions. However, it usually takes a long time from identifying a desirable accession to usage of beneficial alleles. For example, it took 50 years (1960–2010) from the discovery of submergence-tolerant
rice landrace to the successful release of submergence-tolerant rice varieties, and this included fine mapping of the SUBMERGENCE 1 (SUB1) locus from the FR13A landrace, the molecular characterization of SUB1, and the final introgression process [8]. Introgression breeding is very demanding when multiple genes need to be pyramided simultaneously [13]. The use of marker-assisted selection (MAS) is promising to shorten introgression time. However, if multiple genes are targeted, very large population sizes are necessary to ensure at reasonable certainty that an individual with the target genotype can be identified. In addition, marker-assisted backcrossing (MABC) is limited to oligogenic inherited traits. For complex traits that are controlled by a large number of genes, such as yield, MAS is not efficient and is associated with substantial linkage drag, resulting in a yield gap between introgression materials and elite cultivars [31].

Emerging Avenues for Utilization of Exotic Germplasm
Utilization of exotic germplasm depends on efficient and rapid discovery, and introgression of favorable alleles from exotic resources into breeding projects. The rapidly developing high-throughput genotyping and phenotyping methods, pangenome sequencing, together with advances in statistical methods and biotechnology, provide opportunities to address various challenges for utilization of exotic germplasm. With ‘big data’ generated from high-throughput genotyping and phenotyping platforms, omics-based technologies together with emerging statistical genomic methods can be used for the identification and exploitation of major genes, minor alleles, or haplotypes affecting complex traits [32,33]. Novel biotechnologies such as genome editing allow direct transfer of beneficial genes or gene complexes into an elite genetic background, or manipulation of existing genes, in a very efficient way to obtain expected phenotypes without lengthy backcrossing [34]. In addition, genomic selection (GS) can be used to identify pre-breeding materials with beneficial genetic variation for complex traits, potentially addressing linkage drag problems [35].

Considerations in Determining Appropriate Phenotyping and Genotyping Strategies
Faster and cheaper advanced high-throughput genotyping and sequencing technologies including pangenome sequencing provide opportunities to genotype all collected samples in a genebank to monitor genetic diversity within and between different accessions more efficiently, rapidly, and accurately [36–38]. Different genotyping platforms such as genotyping-by-sequencing (GBS) and various SNP genotyping assays, such as KASP, GoldenGate, and Infinium chips, have been widely applied in genome-wide SNP discovery and population genotyping in various plant species [36]. Sequence information may help to identify duplicates and aid in control of contamination when used in a proper context, and may provide greater insights into the identification and evaluation of genetic variation in germplasm. However, the genome of a single individual is insufficient to study the gene diversity within a species, and determination and characterization of the pangenome is necessary to understand variation within a species. Pangenome analysis can help to harness the variation in exotic germplasm by identifying core genes, that are present in all individuals, and dispensable genes that are present in some but not all individuals [39]. In conjunction with phenome information, it can expedite the discovery and use of new and useful alleles [40].

Phenotyping remains expensive but necessary to establish trait–gene associations. Given budget limits, instead of evaluating all materials in genebanks, a ‘core crop reference set’ of materials based on available genomic, passport (origin, geographical, and ecological), and provenance information can be used for in-depth phenotyping and trait dissection [41]. A core collection may include only 5–10% of the original accessions, but should represent at least 70% of the genetic diversity of the collection [42]. With this approach, the number of samples to be phenotyped can be dramatically reduced [43]. The use of core collections has shown great potential for major staple crops such as rice [44,45], soybean [46], wheat [38], and barley [47] to mine favorable alleles and haplotypes associated with important traits, and promising
accessions have been selected for further screening for specific uses [48,49]. For example, the Seeds of Discovery (SeeD) project was initiated by the International Maize and Wheat Improvement Center (CIMMYT) to characterize core sets of maize landrace accessions for facilitating both gene discovery and the establishment of distinct haplotypes, with the goal to establish foundations for harnessing favorable variation from landraces [31]. So far the SeeD project has established a breeder’s core of 4000+ maize landrace accessions from the original 40 000 accessions, and these were genotyped using a large number of molecular markers and extensively phenotyped, including testcross performance (http://seedsofdiscovery.org). A cost-effective strategy was recently proposed by integrating genomic prediction into a broad germplasm evaluation process to ‘turbocharge’ genebanks by selecting representative accessions to form a training set, which will be phenotyped in replicated trials to develop a prediction model to be used to predict the performance of untested accessions [50]. Recently, a hybrid strategy was proposed to be used in evaluating breeding potential and in uncovering the hidden breeding value of exotic germplasm. Per se performance of accessions has a low correlation with their combining ability, and the true breeding values might be masked by a high genetic load. Testcrosses of exotic germplasm may help to reduce the effects of deleterious alleles and increase the accuracy of evaluation [51,52]. Recent advances in high-throughput phenotyping technologies offer new opportunities for genebanks to phenotype accessions in a more efficient, cost-effective, and informative manner. The rapid progress made in field or controlled environmental phenotyping platforms utilizes novel sensors, image analysis, modeling, robotics, and data-mining techniques. These enable genebank managers or plant breeders to monitor traits across developmental stages, rapidly capture dynamic images, and quantify corresponding environmental conditionals to evaluate genotype by environmental interactions. For example, Chen et al. proposed and validated an analysis framework using dynamic phenotypic data of the drought responses of 18 different barley cultivars during vegetative growth collected in an automated plant transport and imaging platform [53]. The analytical framework is useful to bridge the gap between plant phenomics and genomics to efficiently uncover the genetic basis of complex agronomic traits [53]. High-throughput phenotyping platforms can be placed into three categories: (i) proximal (remote) sensing and imaging, (ii) laboratory analyses of samples, and (iii) near-infrared reflectance spectroscopy (NIRS) analysis in the harvestable portion of the crop. As a non-destructive and noninvasive approach, proximal (remote) sensing and imaging systems can be employed to dynamically record a wide range of field traits, including yield potential, adaptation to abiotic (water stress, extreme temperatures, salinity) and biotic (susceptibility to pests and diseases) limiting conditions, and even biomass quality traits [54]. To capture field heterogeneity, imaging techniques in parallel with wireless sensor networks and geographical information systems will allow more-precise mapping and monitoring of spatial variability. Recently, several field-based platforms have been developed, ranging from ground-based to aerial systems [55–57]. High-throughput phenotyping technologies will increase the efficiency of the characterization, evaluation, and utilization of germplasm.

Development and Implementation of a Globally Coordinated Information Platform and Network

It will be urgently necessary for local or national genebanks to coordinate unified standards and common monitoring protocols to integrate genomic with phenotypic data, including ecogeographic and other types of information. The capacities of extensive regional and international crop-based networks such as the Global Gateway to Genetic Resources (Genesys), European Cooperative Program for Plant Genetic Resources (ECPGR), Global Rice Science Partnership (GRISP), and GRIN-Global should be enhanced [58–60]. Free and powerful online global platforms can be used by breeders to quickly search for and request germplasm accessions held in hundreds of genebanks around the world [61]. Valuable feedback information from users who evaluate accessions in specific environments can be shared with the global
network. In this way, the genetic potential and breeding value of currently underutilized materials in genebanks can more readily be discovered and utilized by scientists and breeders.

**Solutions to Maladaptation Syndrome**

One of the most efficient methods to overcome maladaptation is to use controlled field or greenhouse environments. For example, to control delayed flowering of tropical maize germplasm in temperate areas, fabrics that preclude light transmission can be used to cover plants, providing short-day conditions for photoperiod control. However, this method is only possible for small-scale experiments or for a limited number of individuals owing to the costs and labor requirements for setting up and providing short-day conditions over weeks. Pre-breeding is important as a means to explore the genetic potential of unadapted exotic germplasm and make it attractive to breeders. The Germplasm Enhancement of Maize project (GEM) offers a good model of using unadapted exotic germplasm to increase the diversity of US elite germplasm through a well-coordinated effort between public sections and private seed companies [62]. The Raleigh, NC component focuses on identification of useful exotic germplasm and develops 50% tropical lines, while the Ames, IA component focuses on establishing introgression libraries of 25% exotic germplasm, which are evaluated by breeders in the network. The 25% exotic GEM materials are improved and adapted to Midwest photoperiod conditions, while the 50% exotic materials are suited to the Southern USA. In this way, many exotic germplasm accessions are adapted to Midwest photoperiod conditions.

However, the backcrossing process costs multiple seasons, and controlled pollinations are very labor-intensive. If there is prior knowledge about key biological elements or metabolic pathways contribute to the maladaptive syndrome, molecular strengthening (MOST) [63] treatments could potentially be applied to control trait expression. Instead of changing the genetics of germplasm using MAS or transgenic methods, MOST can be used to control traits temporally by external application of effectors to control molecular ‘targets’. For example, gibberellin is used to efficiently induce flowers in biennial crops or long-day plants. Commercial usage of Pro-Gibb T&O (GA₃) can induce flowering on tropical foliage plants [64,65], and exogenous gibberellin (GA) application accelerates spike development under short days for wheat lines expressing VRN1 (VERNALIZATION1) [66]. Research in Arabidopsis has shown that the effects of gibberellins on flowering control correlate with expression of LEAFY, a floral meristem identity gene. LEAFY is one of the floral meristem identity genes which control the transition from leaves with associated lateral shoots (paraclades) to bractless flowers [67]. There are other pathways that affect flowering time. By applying multiple effectors to address different molecules, flowering time can be further fine-tuned.

**Beneficial Allele Identification and Cloning**

*Forward Genetics for Identifying Beneficial Alleles* When genotypic and phenotypic data are available, genome-wide association studies (GWAS) can be used to identify beneficial alleles for target traits [68]. GWAS as a strategy for trait dissection is ideal for analyzing genebank accessions because, unlike classical QTL mapping, for which biparental crosses are needed, a diverse collection of unrelated accessions can be used for GWAS studies [19]. With good control of population structure and family relatedness, which is calculated based on molecular markers, false positives and false negatives can be balanced to identify valuable gene–trait associations [69]. However, current algorithms for GWAS do not favor the identification of rare alleles. To address this concern, the focused identification of germplasm strategy (FIGS) can be employed to accumulate target trait genetic variation [70]. FIGS identifies subsets of accessions that predict the potential contribution of a trait based on habitat and ecographical data associated with collection sites that are most likely to impose selection pressure for the trait of interest. In this way, the likelihood of identifying accessions with specific traits of interest can be maximized [71]. This method is very efficient for tolerance to biotic or abiotic stresses. FIGS
has been successfully used to identify subsets of barley and wheat accessions resistant to various fungal pathogens [72] and to the Russian wheat aphid [73]. Resequencing or RNA-seq methods could subsequently be used to identify the underlying alleles controlling the trait. Another challenge for gene identification is for traits that are affected by strong gene by genetic background interaction (epistasis). In this case, phenotyping potential donors per se is not sufficient to determine their breeding value. For example, small gene segments from a wild relative of tomato Lycopersicon pennelli showed strong effects for fruit phenotype when introgressed into lycopene-rich red fruits of Lycopersicon esculentum, but the valuable wild allele cannot be identified with association methods if we use the wild species phenotype per se [9]. To address this problem, progeny testing of intercrossed derivatives is essential, in conjunction with advanced backcross QTL analysis [74].

Reverse Genetics for Identifying Beneficial Alleles Information about gene function, genome structure, biological pathways, and metabolic regulatory networks is rapidly accumulating [75,76]. Reverse genetic methods can be employed to identify novel alleles for characterized genes. For example, the differences in leaf gene transcripts, proteins, and metabolite accumulation in gln1.3 and gln1.4 mutants were studied at two key stages of plant development controlling grain production in maize. Results showed that several key plant processes can be targeted to identify candidate alleles, affecting metabolic traits such as carbon metabolism and transport, cell wall metabolism, several metabolic pathways, and stress-responsive and regulatory elements [77]. For epistatic effects, large-scale omics projects are ongoing to develop predictive models about which exotic alleles are most promising to improve trait performance when introgressed into elite lines, mainly through exploring key heterosis and transgressive genomic regions [7]. By comparing transcriptomes of different tissue types, stages of development, and expression quantitative trait loci (eQTL) mapping, a deeper understanding of the underlying mechanisms of phenotypic variation can be gained. These methods provide clues about how changes in transcription may affect the phenotype [78]. Systematic genetics combining genomic, metabolomic, proteomic, and transcriptomic information has been successful to explore novel genes and potential networks controlling complex traits. This information was used to increase genetic gain for plant breeding [79–81]. In addition, allele mining can be used to identify superior alleles. For example, ecotype targeting induced local lesions in genomes (EcoTILLING) is a well-established and inexpensive allele-mining approach that has been used to identify novel alleles for genes controlling agronomic traits in diverse crop germplasm accessions [82]. A strategy of multiple accession, pooling-based EcoTILLING coupled with agarose gel detection has been found to accelerate the process of high-throughput allele mining and functional allelic genotyping, as well as the identification of accessions exhibiting favorable natural allelic variants, such as discovery of new alleles conferring resistance to biotic and abiotic stress [83,84]. The cloning of candidate genes underlying agronomically valuable loci identified either by forward or reverse genetics is necessary to uncover the function of candidate genes and explore their mechanisms of regulation and further potential in genome editing.

Introgression of Beneficial Alleles

Introgression of Major Genes into Elite Germplasm Once genomic regions are identified that associate with important traits, and the key genes underlying these traits have been characterized, favorable alleles with major effect in selected accessions can be introgressed into elite germplasm following established marker-assisted backcross procedures. Alternatively, genome editing provides opportunities for targeted mutagenesis and is a powerful tool for functional characterization of plant genes and for genetic improvement of agricultural crops [35]. The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR-associated 9) technology has successfully been used for the introgression of favorable alleles, with precise manipulation of specific genomic sequences to modify crop traits, and has been
successfully applied in economically important crops including sorghum, rice, wheat, soybean, and maize [85,86]. In addition, the CRISPR/Cas9 technique also allows multiplex gene editing by the simultaneous expression of two or more small/single guide RNA (sgRNAs), and this is important because it will allow breeders to address linkage drag and introgress multiple genes simultaneously [87]. A recently proposed method ‘promotion of alleles by genome editing’ (PAGE) shows potential to enable rapid introgression of multiple beneficial alleles [88]. In

**Figure 1. New Avenues for Utilizing Exotic Germplasm.** Core collections were determined based on analysis of passport, provenance, and phenotypic and genotypic data. A dynamic information platform can be built by integrating genomic data with phenotypic, eco-geographic, and other types of information such as gene–trait association information and valuable feedback information from users. Forward or reverse genetics can be employed to identify gene–trait associations which are essential to improve simple inherited traits or complex traits of the plant genetic resources, and could be used as feedback information to genebanks to select additional valuable accessions. The enhanced plant genetic resources can subsequently be used for variety improvement and can be added back to genebanks for further worldwide utilization. Abbreviation: MAS, marker-assisted selection.
comparison to other transgenic technologies, the CRISPR systems are currently considered to be non-GM (genetically modified) technologies in many countries because Cas9 can be deleted from the host plant via backcrossing in subsequent generations once the mutation is accomplished [89,90]. Nonetheless, it remains to be seen whether consumers will accept crops developed using genome-editing technology. For countries which regard genome-editing methods as non-GM methods, the amount of regulatory fees will be significantly reduced, and this will make this method more broadly available to plant breeders. Genetic engineering is one of the options to improve the effects of native alleles.

**Improvement of Complex Traits** For complex traits controlled by numerous genes with small effect, mining for favorable alleles and introgression into elite germplasm or manipulation by genome editing is unfeasible. Instead, the paradigm for polygenic trait improvement has shifted from allele mining towards pre-breeding based on genome-wide selection, and this may help to maximize selection gains and bridge the yield gap between elite and exotic materials [51]. The focus should be on genomic selection to increase the frequencies of favorable alleles. For genome-wide selection, the approach of assessing the breeding potential of exotic germplasm by evaluating test-cross performance with elite testers is more effective and accurate than evaluating their performance per se [52]. Once the most promising accessions have been identified by genome-wide selection, doubled haploid (DH) technology can be used to accelerate line development and eliminate the high genetic load of exotic germplasm by purging deleterious alleles before introgression into elite backgrounds [91,92]. To make best use of exotic germplasm, we propose collaboration between seed companies and public sector institutions in the genotyping, data analysis, and production of DH lines that can be accomplished by public institutions, and phenotyping in multiple locations can be carried out by seed companies. In this way, pre-breeding projects can be developed to provide new germplasm which can also be used to train plant breeders. This concept has been successful to improve complex traits such as plant height using GEM project germplasm [93] – alleles from tropical germplasm with tall plant height were introgressed into temperate elite maize lines to establish pre-breeding materials useful for bioenergy maize breeding. Plant height was significantly increased without affecting flowering time [94]. The SeeD (Seeds of Discovery; http://seedsofdiscovery.org) project has successfully characterized untapped variation in maize landraces and developed bridging germplasm with 75% or more elite and 25% or less landrace genome proportions, with the objective to provide early breeding lines carrying novel, landrace-derived genetic variation, and to breed for high-value characteristics such as nutritional quality, heat and drought tolerance, disease resistance, and tolerance to soil infertility. At the discovery phase, genotype data are used to discover the best landraces to initiate pre-breeding germplasm, based on genomic estimated breeding values (GEBVs) from genomic selection models, using well-characterized core collections as a training population, to maximize the number of beneficial alleles incorporated into pre-breeding materials [32,95,96]. Selected accessions are subsequently crossed with elite germplasm to create pre-breeding populations for future improvement of complex traits (Figure 1).

**Concluding Remarks and Future Perspectives**

High-throughput genotyping and phenotyping combined with novel biotechnologies provide new avenues for the efficient management, characterization, and utilization of exotic germplasm. In the short term, genebanks are responsible for developing and maintaining large numbers of collections to ensure the availability of genetic diversity of different species as a public good. With the availability of large sets of phenotypic and genotypic data, and using interconnected information networks to provide knowledge from different academic areas such as plant breeding and biology, data science can be used to help breeders to search for useful accessions. Beneficial alleles can be subsequently introgressed into elite genetic pools using either direct genome-editing techniques or pre-breeding activities to generate more stable,
high-yielding crop varieties to address food security challenges. In the long term, full discovery of trait–allele–sequence associations and the maturation of rapidly developing genome-editing techniques may enable reduction of the numbers of accessions necessary to provide the needed biodiversity for a given crop. Ideally, a cyclical progression of activities would be designed to complement collection development and result in the determination of a set of well-characterized and genetically diverse samples to be incorporated or maintained for a given species. This would require a dynamic process that can respond not only to changes in available technologies and analytical tools but also to the need to understand how germplasm evolves in response to climate and ecosystem changes. As a starting point, a set of samples could be used to introduce useful genetic diversity by genome editing, and subsequently be used for breeding purposes (Figure 2). It is equally conceivable, as our understanding of the

---

**Figure 2. Comparison Between Current Strategy and Future Strategy for Future Crop Improvement.** The current strategy of utilizing plant genetic resources focuses on selecting core collections and introgressing beneficial genetic variation into breeding programs. Based on a globally coordinated information platform and network, plant genetic resources have the potential to be more efficiently characterized, evaluated, and utilized for future crop improvement. On the one hand, well-characterized plant materials can still be maintained and utilized. On the other, novel genetic variation can be generated based on existing accessions by genome-editing techniques for crop improvement.
inherent value of plant genetic resources grows, and our ability to apply new technologies and create useful new genetic resources matures, that collection size will not be an issue. Instead, genetic value coupled with collaborations and cooperation between institutions may drive the size and scope of collections, and how they are managed and accessed (see Outstanding Questions).

Acknowledgments
The authors would like to thank the USDA National Institute of Food and Agriculture (Project Numbers: IOW04314, IOW01018), as well as the Raymond F. Baker Center for Plant Breeding and the K.J. Frey Chair in Agronomy at Iowa State University for funding this work. This work was supported by funding from the State Key Laboratory of Wheat and Maize Crop Science at Henan Agricultural University (SKL2014KF-05) and the National Natural Science Foundation of China (Grants 31101153 and 31371628).

Supplemental Information
Supplemental information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tiplants.2017.04.002.

References
tion of the United Nations.
using variable length, paired-end genotyping-by-sequencing data. BMC Bioinform. 17, 192.


44. Kim, T.-S. et al. (2016) Genome-wide resequencing of KIRICE - CORE reveals their potential for future breeding, as well as functional and evolutionary studies in the post-genomic era. BMC Genom. 17, 1

45. Li, J.-Y. et al. (2014) The 3,000 rice genomes project: new opportunities and challenges for future rice research. GigaScience 3, 8


50. Yu, X. et al. (2016) Genetic prediction contributing to a promising global strategy to turbocharge gene banks. Nat. Plants 2, 16150


52. Bölh, J. et al. (2014) Breeding potential of European flint maize landraces evaluated by their testcross performance. Crop Sci. 54, 1665–1672


64. Wen, W. et al. (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nat. Commun. 5 (3438)


