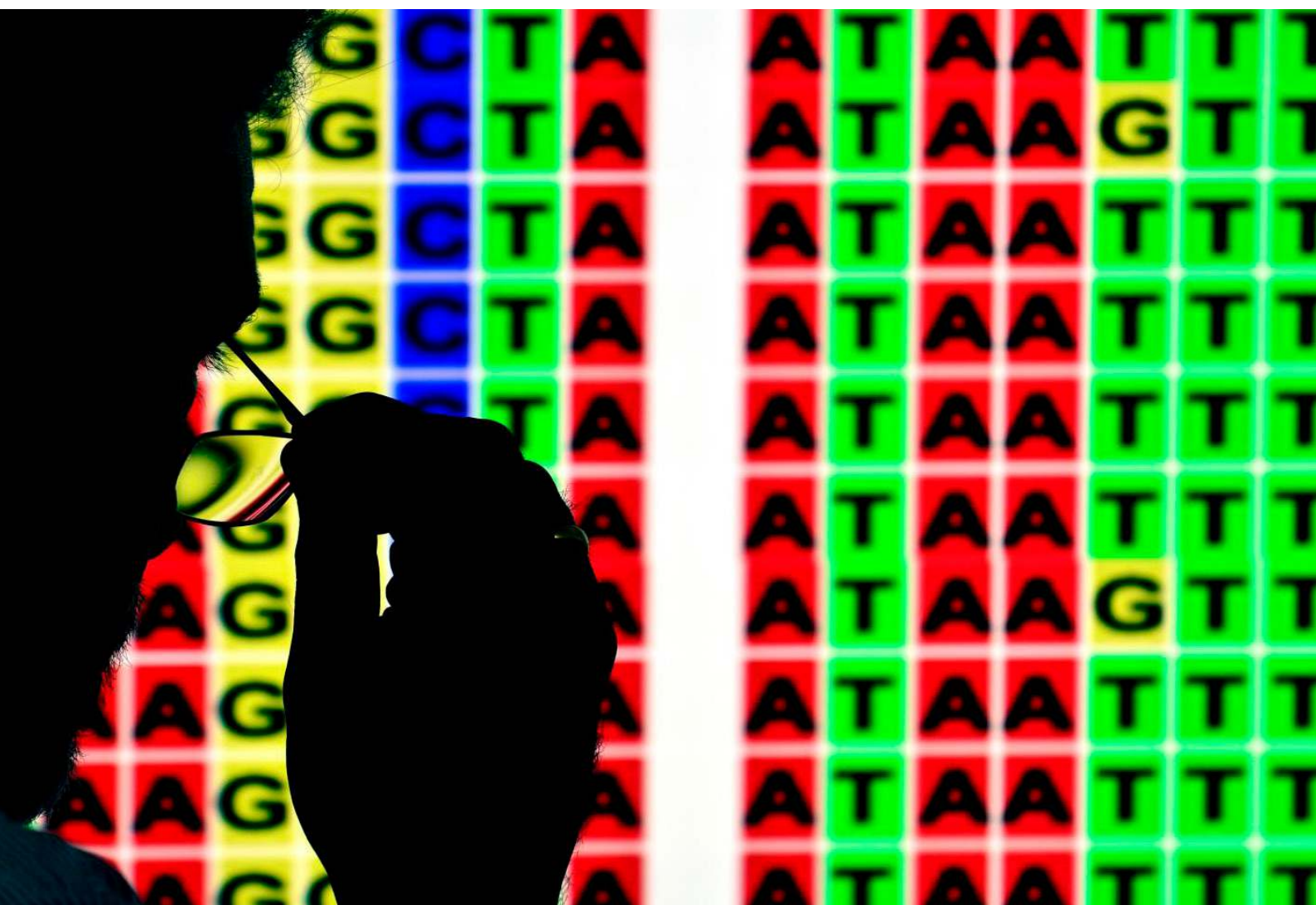


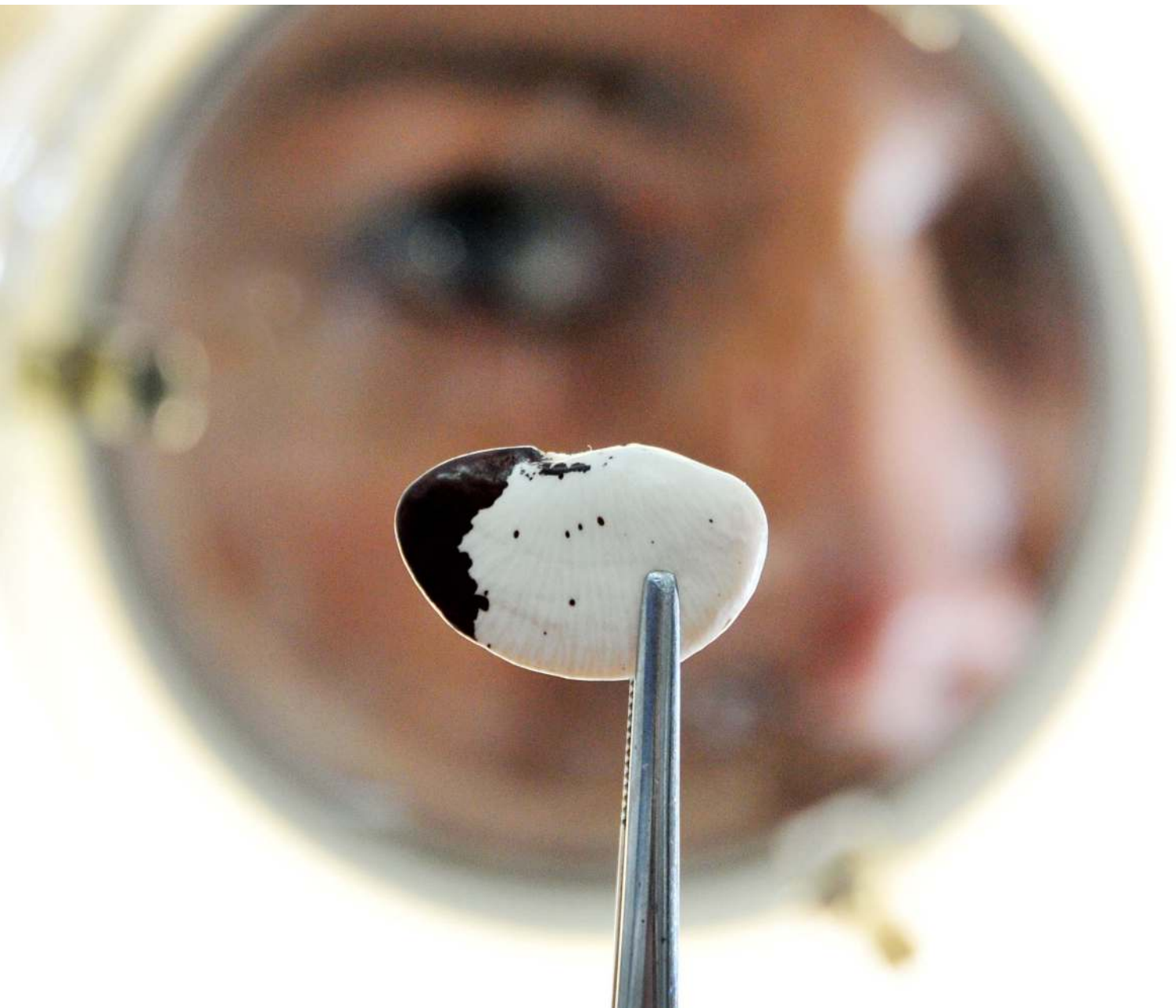


Harnessing the power of crop diversity to feed the future

White Paper



**DivSeek will mine the
wealth of genetic resources
to enhance food and
nutritional security.**



Executive Summary

Meeting the demand for more and more nutritious food around the globe, given climate change, and increased competition for land and water resources, is one of the most important global challenges of our time. Crop production needs to increase and more resilient crops need to be developed to provide food and agricultural products in a sustainable manner for a growing human population.

Assessing and identifying new sources of genetic variation is a critical part of any long-term strategy to enhance the productivity, sustainability and resilience of crop varieties and agricultural systems. Approximately seven million crop accessions are currently being conserved in genebanks collections worldwide. This resource represents one of the greatest, largely untapped, opportunities for accelerating yield gains and overcoming emerging crop productivity bottlenecks. However to access this wealth of diversity will require the characterization of this material via the application of state-of-the-art genomic, phenomic and molecular technologies, and the release of the subsequent data via an online, open-access portal.

The Diversity Seek initiative (DivSeek) will work with existing, emerging and future initiatives to characterize crop diversity and develop a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm. This initiative will help to bridge the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers, to support applied germplasm curation, forward-looking breeding programs and strategic research.

DivSeek will bring together genebanks, breeders, plant and crop scientists, database and computational experts to enhance the use of genebank materials, promote effective utilization of genetic variation in plant improvement, and to better understand how components of genetic variation contribute to plant performance (i.e., growth, development, yield and nutritional composition) in diverse climatic environments. By creating a well-coordinated, international effort, based on experience and knowledge of relevant stakeholders, DivSeek will unlock the potential of crop diversity stored in genebanks around the globe and make it available to all so that it can be utilized to enhance the productivity, sustainability and resilience of crops and agricultural systems.

Introduction and Background

The world population is set to reach 7.2 billion by 2020 and will rise to 9.6 billion by 2050.¹ Agricultural production must increase by 60% by 2050 to meet the increasing demand for food around the globe. Populations in developed nations are predicted to remain steady at around 1.3 billion, while virtually all population growth will occur in the developing world. Climate change and increased competition for resources make this a challenge that affects all of us, but in particular those of us that live in developing nations.

We therefore need to increase crop production and develop more resilient crops that are able to generate step changes in yields, similar to that resulting from the Green Revolution. How will this be achieved? Genetic resources currently housed in existing genebanks represent one of the greatest, largely untapped, opportunities for accelerating yield gains and overcoming emerging crop productivity bottlenecks. Assessing and identifying new sources of genetic variation will be a critical part of any long-term strategy to enhance the productivity, sustainability and resilience of crop varieties and agricultural systems.

Over the last 50 years there has been an increasing homogeneity of the global food supply, world diets have become more similar and as a consequence we are dependent on fewer crops². Today, production of rice, wheat, maize and potatoes provides 60% of the human calorie needs, and about 30 crops account for 95% of the world's food and energy consumption. Despite the small number of crop species that forms the basis of today's

agriculture, the diversity available within each of these crops is immense. For example, in Andean farming communities alone, at least 500 local potato varieties are still grown and marketed, while over 100,000 varieties of Asian rice are available in international genebanks.

However, over time this wealth of diversity has been lost from agricultural production through genetic erosion driven by the demand for high crop productivity on the farm, the desire for crop uniformity in the marketplace, habitat destruction, pests and diseases and the loss of traditional farming cultures. Although modern agriculture has achieved notable increases in overall yield, relying on a small number of crops creates a more vulnerable agricultural system, which is less resilient and adaptable to threats such as climate change and pests and diseases.

To counter the trend towards genetic uniformity global efforts intensified in the 1970s to collect, preserve, characterize and disseminate germplasm of the world's major food crops. Many of these germplasm collections, known as genebanks, are held in trust for the benefit of mankind in genebanks around the world under Article 15 of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA). Crop biodiversity stored in these international as well as many national genebanks, collectively represents thousands of years of evolutionary history, and illuminates the multiplicity of adaptive genetic responses that plant populations have evolved to cope with environmental and physiological stress.



Until recently, it has been impossible to genetically and phenotypically characterize genebank holdings on a large scale, but it is now both feasible and necessary if we are to unlock this genetic diversity for the benefit of human society.

Using the emerging deluge of omics data along with mathematical models and systems biology approaches, we can now study this diversity to reveal the many ways in which nature has solved these problems, and how humans have shaped the genomes of crop plants through domestication and breeding. This information will provide unprecedented insights into the intricate and finely tuned genetic networks that enable plants to respond to subtle or catastrophic perturbations in the environment. It will also present novel solutions and ingenious templates that can serve as a blueprint for accelerating plant breeding to provide sustainable food security for humanity.

To access this wealth of diversity we advocate the immediate application of state-of-the-

art genomic, phenomic and molecular technologies to characterize the world's genebanks, and release of the generated data via an online, open-access information platform. Such an endeavor will unleash the true value of genebanks, and promote active adoption in both public and private sector institutions. This, in turn, will help to improve the conservation status of the conserved germplasm and lower its vulnerability to further genetic erosion *ex situ*.

¹United Nations Population Division (2013), World Population Prospects, the 2012 Revision (Medium Variant).

²Khoury, Colin K, et al. "Increasing homogeneity in global food supplies and the implications for food security." *Proc Natl Acad Sci* 111 (2014): 4001-4006.

Untapped Potential

Approximately seven million crop accessions are being conserved in genebanks collections worldwide.

Despite their promise and successful use by farmers in the past, a lack of information about their genetic potential impedes their utilization in plant improvement. To assess how best to harness this untapped potential to make plant breeding faster, more efficient and more cost effective, a workshop was held on 9-10th January 2014 in San Diego. The workshop brought together 90 experts from 63 institutions and 21 countries (see Annex A) and was funded in part through the Global Crop Diversity Trust, the CGIAR consortium office as well as Genome Canada and several Canadian provincial Genome programs. The list of participants included a wide range of researchers from the Asilomar (see Annex B)³ and Digital Seed Bank initiatives (see Annex B)⁴, breeders, genebank managers, bioinformaticians, data managers, policy specialists, lawyers, funders, donors, private sector participants, and representatives of international organizations and treaties. The participants discussed (1) how the wealth of crop diversity across the globe could be utilized to help meet global challenges and (2) whether existing and planned initiatives could collaborate under one umbrella with a primary goal of increasing crop genetic diversity in breeding programs around the world. The workshop provided the organizers with a plethora of topics and suggestions to take forward, as well highlighting barriers that would need to be overcome.

The workshop concluded that to make progress the following key areas would need to be addressed:

1. Developing common standards and descriptors that allow data to be structured, shared and reconciled with existing data management schemes and that facilitate use by diverse scientific, crop and breeding communities.
2. Establishing new tools and approaches that enable data exploration, mining and utilization.
3. Seeking new and substantial investment to thoroughly characterize crop diversity using state-of-the-art genomic, phenomic and molecular technologies.
4. Facilitating a dialogue amongst national governments, research communities, institutions and individuals to promote awareness of the value of crop diversity and mobilize the diverse resources required to accomplish these goals.

To meet these needs the DivSeek initiative has been established.

³McCouch, Susan, et al. "Agriculture: Feeding the future." *Nature* 499.7456 (2013): 23-24.

⁴Gruissem, Wilhelm, et al. "The global plant council: Increasing the impact of plant research to meet global challenges." *Journal of Plant Biology* (2012): 1-6.



The Initiative

The Diversity Seek (DivSeek) Initiative will accelerate the utilization of crop diversity for the development of more resilient crops and sustainable agricultural production systems to meet the needs of the growing human population.

To achieve this goal, DivSeek will work with existing, emerging and future initiatives to characterize crop diversity, and to form a collaborative network of genebanks, breeders, plant and crop scientists, database and computational experts that will develop a unified, coordinated and cohesive information management platform to provide easy access to genotypic and phenotypic data associated with genebank germplasm.

As a first step, DivSeek will develop a system of internationally agreed-upon standards, protocols, tools, resources and best practices for generating, organizing, structuring, indexing, retrieving, sharing and analyzing genotypic and phenotypic data related to genebank holdings. This will be done in collaboration with existing networks and projects to avoid duplication of effort.

In the next steps, DivSeek will build an information platform that provides access to genotypic, phenotypic, and other types of information linked to physical germplasm as well as to "digital genomes" that represent an essential framework for linking genotype with phenotype. The platform will provide access to

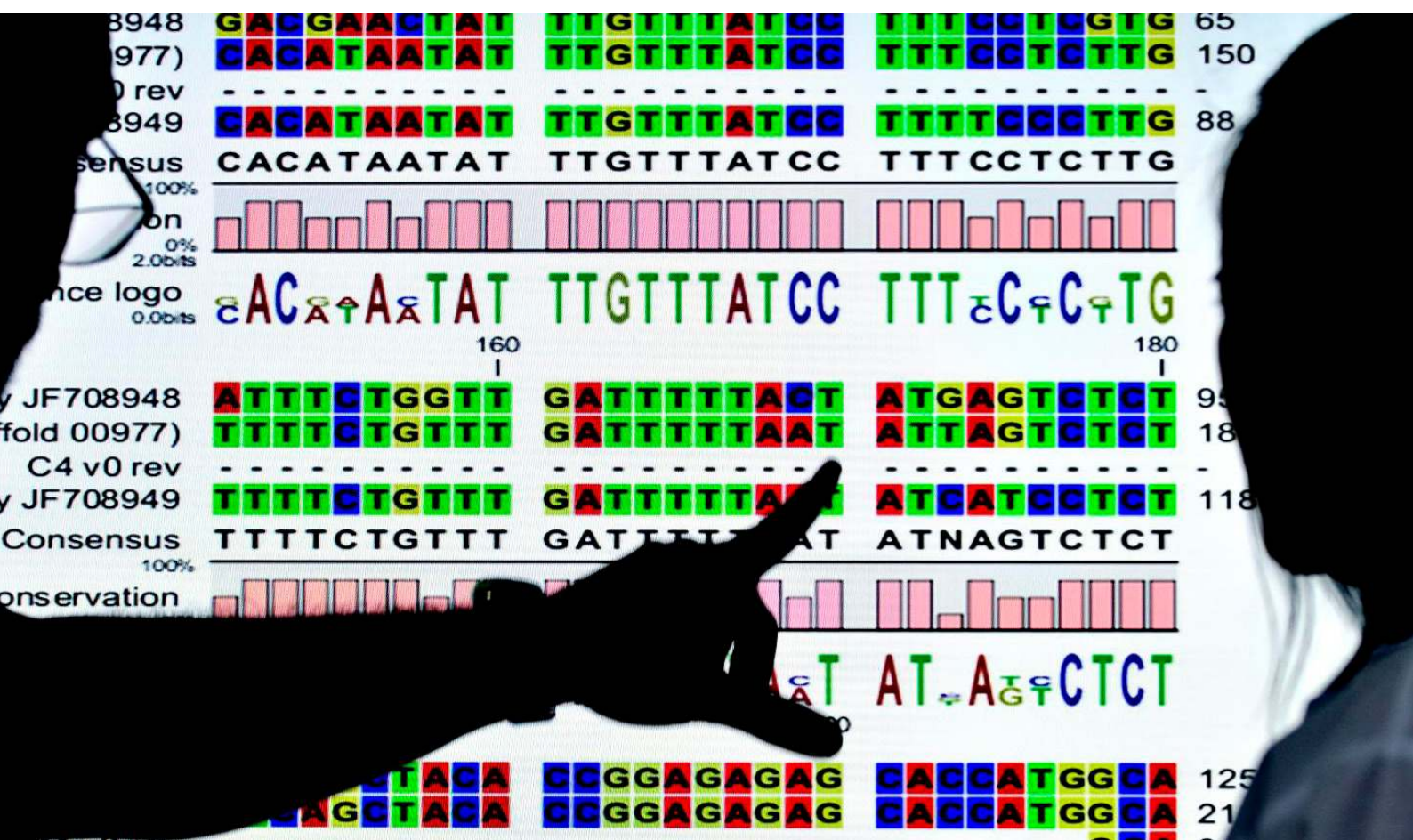
an array of genotypic and phenotypic datasets (stored in a single database or set of federated databases), which will be complemented by a suite of computational tools and resources that allow users such as researchers, genebank curators or plant breeders to visualize and analyze the data in the course of their everyday work.

Developing such an information management platform is an ambitious task, and will require an approach that is sufficiently modular, 'evolvable' and 'future proof' to allow it to adapt to new requirements and technologies, whilst maintaining a set of core standards that allow data exchange, integration and interoperability. One model to consider would be an "apps-based" paradigm, in which applications are developed around a minimalistic platform that largely focuses on specifying data and metadata standards for the annotation of large datasets and to also facilitate data transfer between apps.

To achieve such transformative change, a substantial shift in mindset is needed. Genebank managers, crop scientists and plant breeders need to be able to comfortably handle diverse, high-dimensional datasets in their pursuit of improved or better-adapted crop varieties. DivSeek will therefore promote the development of new interdisciplinary training and capacity building programs to help educate the next generation of plant scientists.

DivSeek's Mission

To enable breeders and researchers to mobilize a vast range of plant genetic variation to accelerate the rate of crop improvement and furnish food and agricultural products to the growing human population.



DivSeek's Strategy

DivSeek will unlock the wealth of diversity that exists in genebanks by linking two complementary sets of approaches, (1) preparing genebank collections effectively to promote their use and to facilitate comprehensive genotypic and phenotypic analyses, (2) performing a detailed characterization of the diversity that exists in

genebanks so that it can be targeted for use in crop breeding programs. The initiative will not only bridge approaches, it will also provide a link between multiple communities of experts, particularly those whose work centers on the physical germplasm and those whose work focuses on digital information and cyber infrastructure.

The Phases

Phase 0

Scoping Exercise (0-12 months)

To ensure that the initiative takes into account the wide range of users that will engage with DivSeek, and to structure an outcome that is flexible enough to deal with future needs we need to fully understand who is doing what, where and how is it being done. What projects, networks and initiatives are currently involved in this sphere? What are their strengths, and weaknesses? Who are the key players? What are the main barriers to progress? How can DivSeek promote collaboration and integration of projects to prevent duplication of effort and encourage knowledge sharing and the spread of best practice? To uncover and collate this information the DivSeek initiative will undertake a landscaping analysis to record the breadth and depth of relevant efforts in this field, identify potential partners and begin a dialogue with its potential user community.

Phase 1

Case Studies and Pilot Projects (6-48 months)

Initially DivSeek proposes to focus on a set of case studies and pilot projects, which will assess the type of data the platform will need to manage; how these data are currently used, accessed, analyzed and integrated; and develop solutions for improved and increased utilization of crop diversity within genebanks for a wide range of users. These efforts undertaken in phase 1 will build upon current methodologies to establish a set of common standards for data formatting, data annotation (i.e. ontologies), meta data, data-sharing protocols, and for linking genotypic and phenotypic data with germplasm. DivSeek

will actively engage with the wider community to discuss and deliberate on possible standards that should be considered for assessment and where necessary will also develop new tools, standards and resources.

The pilot projects will undertake test studies on both existing and new data to ensure that standards, tools and protocols that are developed are fit for purpose and can provide the foundation upon which a community-wide information platform will be based. During phase 1 DivSeek will focus on the use of standards that enable data exchange among different players without requiring individual parties to adopt a specific set of standards as a precondition for their participation. By testing a number of different standards and approaches in an iterative manner we expect that 'a common set of protocols' will emerge that are sufficient to facilitate the work of the majority of users.

The pilot projects will also provide an environment to test and refine mechanisms for formalized partnerships, management structures, communication links and IP. The case studies and pilot projects will draw upon the range of expertise and current initiatives identified in phase 0, and will be selected based on the interests of the participants and the relevance of their activities to the DivSeek initiative.

Examples of topic areas for case studies and pilot projects could include:

1. Bridging current gaps between the physical germplasm resource, "passport" information and associated genomic/genotypic information.

The Phases Contd.

This area will aim to enhance digital catalogues of genetic resources via the inclusion of genomic data, and the utilization of this genomic data to develop tools for improving genebank management. This will greatly enhance our ability to efficiently manage, make available and utilize genetic resources for selection in plant breeding. It will also endeavor to make genebank accessions uniquely identifiable and help to meet the requirements of genomic analysis in terms of germplasm homogeneity and authenticity.

2. Developing mechanisms to enrich genebank data with large-scale phenotypic data.

This area will aim to build up on recent developments in phenotypic technologies (image analysis, automation, and quantitative analysis of seeds and seedlings) to collect phenotypic information such as seed color, shape, weight, size, and other structural features. Assessing which phenotyping approaches can most easily be adapted and developed for inclusion in genebank workflows would help to promote the implementation of phenotypic evaluation in the standard procedures of genebanks.

3. Facilitating allele mining via effectively linking germplasm, genotypic and phenotypic information.

This area will aim to generate a compendium of genetic, molecular, biochemical and metabolic information of crop diversity on a test set(s) of crop varieties/accessions to assess how this data needs to be structured, integrated and analyzed to provide a useful source of new

variation for traits of agronomic importance, such as nutrient use efficiency, photosynthetic efficiency, disease resistance, and abiotic stress tolerance.

Phase 2 **Establish the Information Platform (42–60 months)**

Based on the common protocols established in the pilot projects, a global information management platform providing transparent, integrated access to genomic and phenomic data will be linked to the physical germplasm or other genetic resources, and accompanied by a suite of tools and resources; the information management platform will be available for all to access.

The platform will comprise common data standards, structures and sharing protocols for diverse genotypic, phenotypic, environmental and meta data that will govern the way sequencing data (generated using diverse platforms), phenotyping data (organismal, biochemical, molecular, etc.) and environmental data (field, greenhouse, growth chamber, etc.) are annotated, accessed, shared and integrated by all users. If time and effort is devoted to developing these underlying standards and protocols during the pilot projects, they are likely to be useful to the larger community, assisting germplasm curators, breeders and other plant science researchers in their day-to-day work, helping them reach their goals and ultimately making their lives easier by streamlining what is currently a laborious and often confusing process. DivSeek envisages that the sets of standards, which emerge from this initiative,



will help empower a wide range of users and become adopted over time by the broader community. Such an achievement would be an important indicator of success for the DivSeek initiative.

There are many diverse use cases for the core data that will be stored and/or made available in the DivSeek platform, and it is expected that the availability of the platform will drive the development of innovative applications of the data. It is necessary that the platform be flexible, modular and 'evolvable' to allow for the development of novel applications and interfaces aimed at different communities. We anticipate the development of a broader

DivSeek community, encompassing multiple groups, including those in the developing world, contributing to and exploiting the platform and developing custom tools for specific use cases.

By creating a well-coordinated, international effort, based on experience and knowledge of all relevant stakeholders, projects and other partner consortia, DivSeek aims to facilitate the generation of high quality reusable data and promote access to a new pool of genetic diversity that will facilitate the development of crop varieties that are more resilient to current global challenges.

Launching DivSeek and Developing Governance Principles

DivSeek aims to provide an ‘umbrella’ under which many different relevant ongoing and planned projects can work together to develop solutions to harmonize their outputs and products. DivSeek will facilitate the curation, integration and utilization of relevant data and germplasm, and promote international exchange.

Development of the consortium will proceed in phases and will, by necessity, involve extensive consultation with stakeholders and potential partners. The DivSeek initiative recognizes at the outset the already established communities with their populations of germplasm, genetic, genomic, and phenotypic resources and their own governance structures. DivSeek will aim to harness these existing resources while developing new structures for access, information sharing and innovation that facilitate the goals of the initiative.

As part of the scoping exercise (above), DivSeek will survey existing systems, governance structures and policies, and assess their implications for this initiative, and map, inter alia, (1) where there may be restrictions on use of data or germplasm, (2) where there are tailing obligations arising from use of data or germplasm, (3) where there may be limitations on any germplasm or innovations arising from use of data, and (4) where use of data is without restriction of any kind. This work will inform the decisions on case studies and pilot projects in Phase 1. By dealing with different access restraints, pre-existing rights and

legal requirements attached to different data coming from partners and other providers, the survey findings should help this initiative develop beneficial innovation and facilitate the transfer of know-how related to data use.

DivSeek will develop its own internal sharing practices to pave the way for an efficient and coordinated interaction with partner projects, initiatives and organizations. DivSeek will establish transparent expectations that materials shared with it will be openly shared with collaborators, in exchange for obligations such as attribution, benefit sharing, and technology transfer, while at the same time ensuring a level of discretion and control by the participating partners regarding the distribution of information and the monitoring of its exchange and use.

To initiate the creation of a consortium-type structure, participating projects, initiatives and organizations will be asked to sign an initial letter of intent in which they would agree to commit in principle to working with DivSeek, as equal partners to attain its mission and goals.

For credible community-wide standards and best practices to be adopted by a large group of diverse stakeholders, it is important to clearly articulate incentives. The following set of value propositions could be powerful arguments for many projects, institutions and initiatives to associate themselves with DivSeek and agree to work together:



1. Value will be added to genebank collections by better linking the germplasm with data that will directly (1) facilitate its mining by breeders and other users, and (2) empower genebank managers to curate and prepare their collections for further future use by large-scale 'omics' research efforts.
2. Transaction costs will be lessened by providing an accepted framework for 'rights management' that enables projects to ensure compliance with internationally agreed policy frameworks and data sharing principles and regulations. Sound compliance procedures are an important component for the risk management of any project.
3. Agreement on data standards will facilitate interoperability and broaden usability, and common software tools will streamline project work-flows and reduce the need for de novo software development efforts.
4. A framework and program for capacity building and training will result in the development of a pool of qualified and well-trained professionals to implement and manage future initiatives.
5. If DivSeek reaches a critical mass, funding agencies may require their grantees to join to ensure adherence to best practices for long-term data storage and curation and thus maximum impact of project outputs.
6. Likewise, publishers may want to encourage authors to use the DivSeek framework to make data available for re-analysis and peer validation.
7. By becoming part of DivSeek, partners will bring research efforts to the attention of their peers, thereby providing the basis for future collaborations and synergies.

Financing the Initiative

Fundraising for this initiative will have to focus on the truly global nature of the effort. It is likely that the accomplishment of the goals will require diverse funding and donors from the public and private realms. The DivSeek initiative will initially partner with existing activities and projects and will facilitate the establishment of new pilot projects with selected national and international partners who develop funding proposals to contribute directly to the mission and goals of DivSeek. A facilitation unit has been established while a permanent governance structure is being developed by the community.

The Global Crop Diversity Trust hosts and implements the facilitation unit jointly with the **Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture**, and operates it on a day to day basis with additional inputs provided by the CGIAR consortium, the Global Plant Council and other experts/organizations. To ensure that the goals and objectives of DivSeek are sustainably maintained, hosting arrangements for the facilitation unit need to be formalized. The secretariat function of the consortium may continue to be hosted with the same modalities.

Acknowledgements

We greatly appreciate the time, thoughts and ideas from all those that contributed to the white paper, with particular thanks to Ruth Bastow (Global Plant Council) and Hannes Dempewolf (Global Crop Diversity Trust) for their co-ordination and editorial efforts.

Images provided by Neil Palmer/CIAT and Prof John Doonan UK National Plant Phenomics Centre, Aberystwyth University, UK.

Conclusion

DivSeek is an initiative driven by the need to apply the advances in 'omics' tools and technologies to mine the resources in genebanks for the benefit of plant breeding and thus the world food supply. The initiative will coalesce around common principles, interests and standards building on those that have already taken root in the plant science community. **DivSeek bridges the gap between the information requirements of genebank curators, plant breeders and more targeted upstream biological researchers.**

At its inception, the initiative aims to serve these three distinct clienteles: supporting applied germplasm curation, forward-looking breeding programs and strategic research. By linking comprehensive digital genomic and phenotypic information to germplasm, DivSeek aims to expand the repertoire of plant materials that are used by plant breeders and the research community at large. By encouraging information sharing through 'open data principles', establishing common standards, and competitive development of user interfaces, DivSeek will encourage the plant science community to explore innovative solutions related to data generation, data deposition and data mining in order to help

develop more resilient crop cultivars and furnish food and agricultural products to the growing human population.

The intersection of "omics" data with germplasm resources and breeding trials has the potential to transform crop improvement into a predictive science, and shorten lead times for the development of new crops. While this transformation is already underway, activity remains fragmented and data are often undiscoverable (in a lab notebook or unique database) or inaccessible (generated in the private sector). DivSeek will therefore bring together genebanks, breeders, plant and crop scientists, database and computational experts to enhance the use of genebank materials, promote effective utilization of genetic variation in plant improvement, and to better understand how components of genetic variation contribute to plant performance (i.e., growth, development, yield and nutritional composition) in diverse climatic environments. DivSeek will unlock the potential of crop diversity and make it available to all so that it can be utilized to enhance the productivity, sustainability and resilience of our crops and agricultural systems.



Annexes

Annex A

Institutions Represented at the San Diego Meeting

Agriculture and Agri-Food Canada, Canada	IBM, USA
Australian National University, Australia	ICARDA, Morocco
Bayer CropScience, USA	ICRISAT, India
Biosciences/Abbott Labs, USA	IITA, Nigeria
Bioversity International, France	INRA, France
Centre Kasetart University, Thailand	Integrated Breeding Platform, CGIAR, Mexico
CGIAR Consortium Office, France	Iowa State University, USA
CIAT, Colombia	IPK, Germany
CIMMYT, Mexico	IRRI, The Philippines
CIP, Peru	KWS Saat AG, Germany
Cornell University, USA	Monsanto, USA
CropLife Canada, Canada	Natural History Museum, UK
CSIRO, Australia	NIAS, Japan
Dalhousie University, Canada	NIG, Japan
Du Pont Pioneer, USA	RWTH Aachen, Germany
EMBRAPA, Brazil	SAP, Germany
ETH Zurich, Switzerland	Syngenta, Switzerland
European Bioinformatics Institute, UK	TGAC, UK
FAO, Secretariate of the ITPGRFA, Italy	The Government of Queensland, Australia
Forschungszentrum Juelich, Germany	The James Hutton Institute, Scotland
Gates Foundation, USA	Tokyo University of Agriculture, Japan
Genome Alberta	UBC, Canada
Genome Atlantic	University at Georgia, USA
Genome BC, Canada	University of Adelaide, Australia
Genome Canada, Canada	University of Hohenheim, Germany
Genome Institute BIOTEC, Thailand	University of Missouri, USA
Genome Prairie, Canada	University of Queensland, Australia
Genome Quebec, Canada	University of Saskatchewan
German Association of Plant Breeders, Germany	University of Toronto, Canada
Global Crop Diversity Trust, Germany	USDA, USA
Global Plant Council, UK	Wageningen University, The Netherlands
Helmholtz Zentrum Munich, Germany	

Annex B

Foundational Initiatives

The Asilomar initiative

What is the best strategy to mine the untapped potential in global genebanks? This topic was the focus of a meeting about crops and their wild relatives held in Asilomar (California) in December 2012. The workshop convened a group of about 50 plant breeders, germplasm managers, plant biologists and physiologists from 14 different countries. It culminated in a publication "Feeding the Future" which advocated establishing

a large-scale initiative to create an internationally accessible information resource that would dynamically link sequence and phenotypic information with germplasm samples housed in the world's genebanks (Nature 499:23-24). The group also advocated moving beneficial diversity from wild and unadapted plants into elite, adapted germplasm for immediate use by plant breeders to hasten their efforts to develop improved crop varieties to meet global challenges.

A unique feature of such an initiative would be the emphasis on linking information derived from sequencing and performance evaluation directly to the seed stocks and germplasm resources that are the starting point for variety development. In most information repositories today, sequence information is aligned to a digital “genome” in a database, but not directly linked to the physical germplasm resources that are needed by the plant breeding community.

As a follow up to the Asilomar meeting and in preparation for the San Diego workshop, a ‘Delphi’ survey was distributed to workshop participants and a relevant expanded group of multi-stakeholder experts by a research group at the University of Saskatchewan with the declared aim of gathering responses on the desired features and actions of the emerging initiative to enhance decision-making. The survey consisted of two rounds that addressed topics such as the mission statement, value/issue frameworks, governance structure, usage and operations, and related efforts. The surveys were completed by 84 (58%) and 79 (54%) respondents, respectively. The first survey asked questions related to the scope and priorities of the initiative while the second focused on specific features such as types of governance and financing/funding models. The survey identified 49 unique initiatives or efforts deemed relevant to the initiative. This long list reflects the many disparate and often fragmented efforts that are currently underway in the plant community, further emphasizing the need to link initiatives and enable information exchange and shared use of data.

Digital Seed Bank Initiative

Digital Seed Bank is led by the Global Plant Council, (<http://globalplantcouncil.org>) which is a coalition of 28 plant and crop science societies from across the globe, representing over 50,000 researchers over 6 continents. The Digital Seed Bank, proposed in 2011 (J Plant Biol 55:343-348), aims to capture quantitative information about the expression of genes, proteins and metabolites and combine it with genomic data and phenotypic information of crops in their native environments. Collecting and integrating ‘omics’ and trait datasets on thousands of accessions that span the broadest possible diversity of a single species, will uncover new genes, alleles and networks and allow a better understand how components of genetic variation contributes to plant performance under varying climatic conditions. The data generated by the Digital Seed Bank will be key to the development of resilient crop varieties that meet future agricultural and societal needs.

Seeds of Discovery Project

The Seeds of Discovery (SeeD) project, initiated by CIMMYT as part of the Mexican Government-funded MasAgro initiative in 2011, harnesses the power of genetic resources to facilitate an expansion of the

genetic base of maize and wheat breeding programs (<http://seedsofdiscovery.org>). To date, SeeD has genome-profiled more than 22,000 maize and 48,000 wheat accessions using robust, standardized genotyping-by-sequencing (GbS) protocols. Molecular data are linked to traits evaluated under field conditions, with an emphasis on climate resilience, diseases and grain quality. Genetically diverse accessions with favorable traits are used to develop ‘bridging germplasm’ harboring novel, favorable alleles in breeder-ready genetic backgrounds, to accelerate the utilization of untapped diversity in the development of high-yielding, climate-ready cultivars. Ongoing capacity-strengthening efforts and a genetic-analysis service that assists breeders in the implementation of genomic-selection methods further enhance the adoption and use of the germplasm and knowledge generated by the project.

Rice Diversity Projects

The Rice Diversity Project was initiated in 2006 as a collaboration between Cornell University, the USDA’s National Rice Research Center in Arkansas, and IRRI to establish a genome wide association (GWA) mapping platform for rice. The project has genotyped 2,000 purified rice accessions with 700,000 high quality SNPs using a high-density rice array. Samples have been phenotyped for numerous traits, and a GWA analysis pipeline has been established as the basis for gene and QTL discovery.

The 3000 Rice Genomes Project was initiated in 2011 as a collaboration between the International Rice Research Institute (IRRI), the Chinese Academy of Agricultural Sciences (CAAS) and BGI to sequence the genomes of 3000 rice accessions from IRRI and the China National Crop Genebank. Prior to being sequenced, the 3,000 accessions were purified, and genetic stocks were established of the purified lines to be available to the research community worldwide through the IRRI genebank for gene discovery. The Global Phenotyping Network, led by CIRAD, is coordinating large-scale multi-location phenotypic evaluation and phenomics screening of genetic stocks, to be combined with the genetic data from both projects for gene discovery. With both the genotyping and the sequencing work completed in 2013, the International Rice Informatics Consortium (IRIC) was formed to bring together leading laboratories in rice genetics around the world through a portal providing integrated access to germplasm, genetic data, phenotypic data, and the tools required to curate and analyze the data. The data and commentary are available at these webpages:

Datanote
<http://www.gigasciencejournal.com/content/3/1/7>
Commentary
<http://www.gigasciencejournal.com/content/3/1/8>
Data
<http://gigadb.org/dataset/200001>



www.divseek.org
divseek@croptrust.org