

The conservation and use of crop genetic resources for food security

Colin K. Khoury

Thesis committee

Promotor

Prof. Dr P.C. Struik
Professor of Crop Physiology
Wageningen University

Co-promotor

Dr A. Jarvis
Research Area Director, Decision and Policy Analysis,
International Center for Tropical Agriculture (CIAT)
Flagship Leader for Climate Smart Agriculture,
CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS)

Other members

Prof. Dr M.E. Schranz, Wageningen University
Prof. Dr A. Graner, Leibnitz Institute of Plant Genetics and Crop Plant Research (IPK), OT
Gatersleben, Germany
Dr L. Visser, CGN, Wageningen University and Research Centre
Prof. Dr P.H. Feindt, Wageningen University

This research was conducted under the auspices of the C.T. de Wit Graduate School of
Production Ecology and Resource Conservation.

The conservation and use of crop genetic resources for food security

Colin K. Khoury

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol.
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 31 August 2015
at 11 a.m. in the Aula.

Colin K. Khoury

The conservation and use of crop genetic resources for food security
305 pages.

PhD thesis, Wageningen University, Wageningen, NL (2015)
With references, with summary in English

ISBN: 978-94-6257-442-7

Abstract

Khoury, C.K. (2015). *The conservation and use of crop genetic resources for food security*. PhD thesis, Wageningen University, Wageningen, The Netherlands. 305 pp.

Crop genetic diversity is a critical resource to address the nutrition and agronomic challenges facing global food security. Increases in use of this diversity are expected but are dependent upon conservation, availability, and access. Considerable erosion of crop genetic diversity has occurred *in situ*, i.e., in farmers' fields and natural habitats, and the variation conserved in *ex situ* genebanks is also vulnerable due to insufficient resources. The window of opportunity to resolve these deficiencies and thus accomplish a comprehensive global system for crop genetic diversity conservation and availability for use will not remain open indefinitely.

Among the factors hindering the conservation of crop genetic resources is a lack of essential information regarding this diversity. Questions include: (a) what is the status of diversity in our food systems, and where are the greatest vulnerabilities?, (b) where can genetic diversity be found that can be useful in increasing productivity and mitigating these vulnerabilities?, (c) is this genetic diversity available in the present and in the long term?, and (d) what steps are needed to improve the ability for researchers to access genetic resources critical for present and future crop improvement?

This thesis aims to contribute to the knowledge required to answer these questions through an exploration of the need for, potential of, challenges and constraints regarding, and necessary steps to enhance the conservation and use of crop genetic diversity. The research starts with an investigation of the state of diversity in global food supplies, finding that national food supplies around the world have become increasingly similar over the past 50 years, gaining in calories, protein, and fat, as animal-derived foods and high-calorie plant foods have risen in importance. The proportion of diets consisting of major cereals, sugar crops and oil crops has increased, while regionally and locally important cereals, root crops, and oil crops have generally become further marginalized.

The thesis then delves into the potential for utilization of a particular set of genetic resources of increasing interest globally – crop wild relatives. These wild cousins of cultivated species have been used for many decades for crop improvement. Their utilization is likely only to increase due to improvements in information on species and their diversity, advancements in breeding tools, and the growing need for exotic genetic diversity in order to address compounding agronomic challenges. As wild plants they are subject to a myriad of human caused threats to natural ecosystems. A focus on wild genetic resources is thus timely both for biodiversity conservation and food security objectives.

Research on crop wild relatives first concentrated on the identification of potentially important wild genetic resources at the national level in the United States. The resulting National Inventory listed close to 5000 taxa. A prioritization of species based on value to food security emphasized close to 300 native taxa that are most closely related to important food crops.

Once potentially valuable genetic resources are identified, subsequent information is needed regarding where they occur, what diversity they may possess, and how well conserved and therefore available to crop breeders they are. A 'gap analysis' methodology was proposed to answer these questions, capitalizing on developments in digital occurrence and eco-geographic data as well as species distribution modeling, intentionally utilizing freely available software and data, and incorporating a novel expert evaluation tool.

The thesis then advances the gap analysis methodology, taking advantage of improvements in species targeting, occurrence data, modeling, and expert feedback methods, and further utilizing eco-geographic information to identify traits of value to crop improvement objectives. In case studies on the wild relatives of bean (*Phaseolus* L.), sweetpotato [*Ipomoea batatas* (L.) Lam., I. series *Batatas*], and pigeonpea [*Cajanus cajan* (L.) Millsp.], related species were found to be highly under-represented in *ex situ* conservation systems and thus inadequately available to breeders and researchers. Species differed among themselves and in comparison to the associated crop in their adaptations to temperature, precipitation, and edaphic characteristics, and many species also showed considerable intraspecific variation. Taxa and specific geographic locations were prioritized for further collecting in order to improve the completeness of germplasm collections for these important crops.

While conservation of crop genetic diversity is fundamental to the availability of this diversity for breeding, it is national and international policies that determine the real capacity for researchers to acquire these resources. The thesis culminates in an exploration of the degree to which international collaboration is required in order to achieve access to genetic resources where they are needed. Countries were found to be highly dependent on crops whose genetic diversity largely sources from outside their borders. This reliance is evident even in countries located in regions of high indigenous crop diversity and has increased significantly over the past half century, bolstering evidence for the need for effective national and international policies to promote genetic resource conservation and exchange.

Keywords: Agricultural development, Climate change adaptation, Crop diversity, Crop genetic resources, Crop improvement, Crop wild relatives, *Ex situ* conservation, Food security, Gap analysis, Germplasm conservation, Germplasm exploration, Interdependence, National inventory, Plant genetic resources.

Preface

This thesis is dedicated to Luigi Guarino, Jane Toll, Andy Jarvis, and Hannes Dempewolf, for their guidance, inspiration, and remarkable support.

My profound thanks to Cary Fowler and Gary P. Nabhan, for catalyzing this path long ago, embracing my efforts along the way, and for continuing to inspire me with their visions.

My gratitude to my family for their patience and support.

My thanks to the Crop Wild Relatives team (Harold A. Achicanoy, Vivian Bernau, Nora P. Castañeda-Álvarez, Julian Ramírez-Villegas, Chrystian C. Sosa, Steven Sotelo, Hector Tobon, Alex Castañeda, Shirley Calderón, Ingrid Vanegas, and Ovidio Rivera) for your hard work, and my particular appreciation to Nora, for being my academic and professional sister over these years.

My gratitude to my advisors Paul C. Struik, Andy Jarvis, John H. Wiersema, and Stephanie Greene, for your guidance and support.

Global efforts in plant genetic resources are successful only through diverse collaborations. I thank the taxonomists, data managers, conservationists, curators, ecologists, plant breeders, and other genetic resource professionals with whom I have had the honor to work and/or benefit from their efforts in gathering and understanding pertinent data. Particular thanks to Harold A. Achicanoy, Vivian Bernau, Anne D. Bjorkman, Nora P. Castañeda-Álvarez, Marleen Cobben, Daniel G. Debouck, Stef de Haan, Hannes Dempewolf, Ruth J. Eastwood, Johannes M. M. Engels, Lauren A. Eserman, Cary Fowler, Ximena Flores-Palacios, Stephanie Greene, Luigi Guarino, Bettina Heider, Robert L. Jarret, Andy Jarvis, Henry Juarez, Michael Minion Kantar, Mulualet T. Kassa, Sandy Knapp, Nigel Maxted, Richard E. Miller, Jonas V. Müller, Carlos Navarro-Racines, Sally L. Norton, Julian Ramírez-Villegas, Loren H. Rieseberg, Genoveva Rossel, Tiina Särkinen, Robert W. Scotland, Chrystian C. Sosa, Steven Sotelo, David Spooner, Paul C. Struik, Mindy Syfert, Álvaro Toledo, Hari D. Upadhyaya, L. Jos G. van der Maesen, John H. Wiersema, John R.I. Wood, and G. Craig Yencho for the intensive collaborations.

Thanks also to the staff of the Statistics Division of FAO for providing literature and clarifying historical modifications to commodity and country listings, and Johannes M. M. Engels, Mark Vellend, Sarah Elmendorf, Jay Bost, Gea Galuzzi, Gary P. Nabhan, Tim Crews, Alder Keleman, Neil Palmer, and Carlos Navarro-Racines for inputs (Chapter 2); the USDA National Plant Germplasm System (NPGS) curators, members of the NPGS Crop Germplasm Committees, and Agricultural Research Service crop experts for their inputs on the Inventory, and Gary Kinard, Edward Garvey, Karen Williams, and Larry Stritch for helpful comments on the manuscript (Chapter 3); Nora P. Castañeda-Álvarez for preliminary inputs, and Silvia Castano, Ovidio Rivera and Carlos Nagles for georeferencing (Chapter 4); Shirley Calderón, Alex Castañeda, Ovidio Rivera, and Ingrid Vanegas for gathering and organizing occurrence data (Chapters 5 and 6); Flora Mer for data gathering, and Steven Prager for comments (Chapter 6); and Gary Kinard, Álvaro Toledo, and Jan Borring for comments (Chapter 7).

Thanks to Charis Khoury for the cover illustrations, one of which was inspired by a photo taken by Kasia Stepień.

Contents

Chapter 1. General introduction	1
Chapter 2. Increasing homogeneity in global food supplies and the implications for food security	17
Chapter 3. An inventory of crop wild relatives of the United States	51
Chapter 4. A gap analysis methodology for collecting crop genepools: a case study with <i>Phaseolus</i> beans	87
Chapter 5. Distributions, <i>ex situ</i> conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [<i>Ipomoea batatas</i> (L.) Lam., <i>I. series Batatas</i>]	129
Chapter 6. Crop wild relatives of pigeonpea [<i>Cajanus cajan</i> (L.) Millsp.]: distributions, <i>ex situ</i> conservation status, and potential genetic resources for abiotic stress tolerance	187
Chapter 7. Interdependence among countries on crop genetic diversity	237
Chapter 8. General discussion	275
Summary	291
Publications of author	297
PE&RC Training and Education Statement	301
Curriculum vitae	303
Funding	305

CHAPTER 1

General introduction

BACKGROUND

Human beings have been remarkably successful over the course of their history in expanding both in total population as well as distributional range across the planet. Perceived limits to population growth (e.g., Malthus 1803, Vogt 1948, Ehrlich 1975) have been overcome historically through innovation in the use of natural resources and their derivatives, e.g., via the transition from gathering and hunting to agricultural societies, improvements in hygiene and waste management in increasingly densely populated areas, antibiotics and other advancements in medicine, and the agricultural green revolution (Diamond 2004, Gepts 2006).

Global food production has thus far kept pace with demand (Gepts 2006), and both the number of stunted as well as underweight persons worldwide is declining (Kearney 2010, International Food Policy Research Institute 2014, WHO 2015), but a number of persistent and novel challenges bring renewed concern to the question of our ability to feed ourselves into the future. First, from the perspective of human nutrition, although adequate food energy (i.e., calories) is available as a global sum, disparities in access and challenges in distribution continue to create hunger (Hawkesworth *et al.* 2010, Kearney 2010, Remans *et al.* 2011, WHO 2015). Second, food system and demographic change, and the historical emphasis on macronutrients (i.e., carbohydrates, protein, and fat) in agricultural research and in development policies have had mixed effects on providing for essential micronutrients. Although improvements have occurred in some geographic areas, deficits in vitamin A, iodine, folate, iron, and/or zinc, among others, persist in diets in more than half of all countries, affecting more than two billion persons globally (Remans *et al.* 2011, CDC 2015, WHO 2015). Finally, economic development, urbanization, and many facets of globalization have increased consumer purchasing power and the availability particularly of energy dense foods (Popkin 2006, Pingali 2007, Hawkesworth *et al.* 2010, Kearney 2010). In concert with lifestyle change, this ‘nutrition transition’ toward diets high in animal products, sugars, and fat has globalized the epidemic of overweight and obesity and its associated non-communicable illnesses, including heart disease, Type 2 diabetes, and cancer (Popkin 2006, Pingali 2007, Kearney 2010, Tilman & Clark 2014). While

each of these branches of malnutrition individually occur more severely in particular areas worldwide, the so-called ‘triple-burden’ of malnutrition (i.e., the coexistence of food insecurity, undernutrition, and overweight) can now appear within the same regions, communities, and even households (Gomez *et al.* 2013).

Agricultural production challenges are equally concerning. Food demand may as much as double in the coming few decades, due both to the rising global human population and to changing dietary expectations (Hoisington *et al.* 1999, Tilman *et al.* 2011, Kastner *et al.* 2012). Yet projected future yield trends are no longer expected to keep pace with this increase in demand (Ray *et al.* 2013). Real limitations in the further expansion of arable lands and the use of water, phosphorus, and other natural resources are also being reached (Cordell *et al.* 2009, Foley *et al.* 2011), and pollution levels due to the excessive use of nitrogen fertilizers and other inputs are considered to be well over those which may be considered safe (Rockstrom *et al.* 2009, Bodirsky *et al.* 2014). Arable lands and other natural resources dedicated to food production are increasingly subject to competition from biofuel and other non-food crops (Ray *et al.* 2012). There is also an increasing urgency to balance the use of lands and inputs for agriculture with the maintenance of the integrity of wild habitat, biodiversity, and other ecosystem services (Foley *et al.* 2005, Matson & Vitousek 2006, Phalan *et al.* 2011). Compounding these challenges, global climatic change is generating greater uncertainty in reliable agricultural production in an increasing number of regions due both to greater extremes as well as to increased variation in temperatures and precipitation, and is additionally altering pest and disease occurrences (Lobell *et al.* 2008, Burke *et al.* 2009, Jarvis *et al.* 2009, Wheeler & von Braun 2013).

PROBLEM ANALYSIS

Realizing long-term global food and nutrition security will require achieving a level of balance in demand from an informed and empowered populace, and sustainably producing sufficient accessible food and nutrition in order to nourish this population (Ruel 2003, Alston *et al.* 2009, Tilman & Clark 2015, Graham *et al.* 2007, Foley *et al.* 2011, Remans *et al.* 2011, West *et al.* 2014). While the pathway to this end is not completely evident, balancing food demand is likely to require a wide range of actions including sustainable economic development and security, education, reductions in inequality in human societies, and availability and access to family planning services and technologies. In regard to food supply, it is expected to require

maximizing productivity within defined arable lands with the use of renewable inputs, i.e., sustainable intensification, as well as reducing food waste (Godfray *et al.* 2010, Foley *et al.* 2011, Tilman *et al.* 2011, Struik & Kuyper 2014).

Producing more food on the same amount of land may be achieved through two main pathways – making agronomic practices more effective to bring yields closer to potential, and increasing potential yield through crop genetic improvement (Tilman *et al.* 2011). Innovations in agronomy and in crop breeding have provided solutions to historical production constraints, but have also contributed to the present challenges. Through the green revolution, traditional practices of maintaining relatively low but reliable yields via the use of spatial crop genetic diversity (i.e., the employ of diverse crop landrace populations, and numerous crops grown in the same areas) have been replaced with technologies employing higher yielding, genetically uniform crop varieties typically grown over large areas and supported by considerable use of agricultural inputs (Gepts 2006, van de Wouw *et al.* 2009, 2010). Due to this spatial uniformity, such varieties are typically substituted every few years in order to avoid crop failure resulting from genetic vulnerability to pests and/or diseases (i.e., temporal diversity has mitigated the loss of spatial diversity). Given the current and projected food demand, the global food system is dependent upon the industry of modern crop varieties and agronomic practices to maintain high yields (Hoisington *et al.* 1999).

The question therefore is how production and associated food supply practices in the present and future can increase food availability while better conserving natural resources, maintaining critical ecosystem services, and being robust against climatic change and other shocks. The use of genetic diversity in crop improvement, with the aim of improving the ecological efficiency of crop production through traits such as drought and heat tolerance, input use efficiency, and extended storage life, is considered to be a key answer to this question (Xiao *et al.* 1996, Hoisington *et al.* 1999, Zhu *et al.* 2000, Gepts 2006, Guarino & Lobell 2011, McCouch *et al.* 2013).

Crop genetic diversity is generated through genetic mutation and recombination, and further transformed through natural and artificial selection, and is therefore the product of the evolution over time of crops and closely related wild plants in their agricultural and natural habitats. As the source of agronomic traits employed for adaptation to biotic and abiotic stresses and for

yield increases through breeding, and the palette from which food systems may be further diversified, this genetic diversity is a biological cornerstone of food security.

Expectations of increases in utilization of crop genetic diversity are made under the assumption that adequate variation will be available for exploration. Ironically, considerable erosion of crop genetic diversity has occurred through the adoption of improved crop varieties or substitute crop species and subsequent neglect of traditional varieties and crops, economic development and associated shifts in consumer demand, land use change and habitat destruction, and urbanization and the displacement of cultures associated with particular crops and varieties, among other factors (Harlan 1975, Hoisington *et al.* 1999, Gepts 2006, Wilkes 2007, van de Wouw *et al.* 2009, 2010). In some crops, only a fraction of the diversity once present is thought to still be found today in farmers' fields, e.g., in wheat varieties in some areas of the Fertile Crescent (Harlan 1971, Gepts 2006).

Speaking on the importance of crop genetic resources and expressing concerns regarding their state of conservation four decades ago, Jack Harlan, a preeminent American agronomist and professor of plant genetics, stated,

“These resources stand between us and catastrophic starvation on a scale we cannot imagine. In a very real sense, the future of the human race rides on these materials. The line between abundance and disaster is becoming thinner and thinner and the public is unaware and unconcerned. Must we wait for disaster to be real before we are heard? Will people listen only after it is too late?” (Harlan 1975).

Due to the disappearance of crop genetic diversity in farmer's fields and wild habitats, the world's genebanks originally established to make plant genetic resources readily available to breeders for crop improvement, have become essential repositories for crop diversity conservation (Hoisington *et al.* 1999, Gepts 2006, FAO 2010, Thormann *et al.* 2015). A substantial portion of the world's remaining heritage of food crop genetic resources is likely now conserved exclusively in genebanks, and important steps have been taken to begin to ensure reliable long-term funding for the conservation, documentation, and distribution of these resources (Global Crop Diversity Trust 2013), and to facilitate access to this diversity worldwide (FAO 2002).

Unfortunately, though, many genebank collections holding unique genetic diversity fail to comply with international standards due to degrading infrastructure, inconsistent power supplies, far from ideal regeneration cycles and protocols, poor safety duplication, unsophisticated documentation systems, and/or inadequate staff and resources to fulfill their conservation and distribution mandates (FAO 2010, Khoury *et al.* 2010). Two points regarding the current state of the world's heritage of crop genetic diversity are thus clear - much remains to be done to secure the diversity threatened *in situ* and in under-funded genebanks, and the window of opportunity to accomplish a comprehensive global system for genetic resources conservation will not remain open indefinitely (Wilkes 2007, FAO 2010).

RESEARCH QUESTIONS AND OBJECTIVES

Among the factors hindering the actualization of a global system for crop genetic resources conservation and availability is a lack of essential information regarding this diversity. A number of primary questions have not been fully answered, including:

- (a) what is the status of diversity in our food systems, and where are the greatest vulnerabilities?,
- (b) where can genetic diversity be found that can be useful in increasing productivity and mitigating these vulnerabilities?,
- (c) is this genetic diversity available in the present and in the long term?, and
- (d) if not available, what steps are needed to improve the ability for researchers to access genetic resources critical for present and future crop improvement?

This thesis aims to contribute to the knowledge required to answer these questions through an exploration of the need for, potential of, challenges and constraints regarding, and necessary steps to enhance the conservation and use of crop genetic diversity.

RESEARCH APPROACH

The research starts with an investigation of the current state of diversity in global food supplies. This also represents an exploration into a longstanding two-part assumption in the field of genetic resources – that humanity relies on relatively few crops for its survival, and that this list of important crops is growing smaller over time.

The thesis then delves into the potential for use of a particular set of crop genetic resources of

increasing interest globally – crop wild relatives. These weedy and wild cousins of cultivated crops have been used for many decades in order to introduce traits of value to crops through breeding, particularly for pest and disease resistance (Iltis 1988, Xiao *et al.* 1996, Gur & Zamir 2004, Hajjar & Hodgkin 2007, McCouch *et al.* 2007, Khoury *et al.* 2010, Maxted *et al.* 2012). Their use in crop improvement is likely only to increase for a variety of reasons. First, information is improving regarding the identity, potential, and conservation status of these plants, and digital data platforms are making this information more readily available (Harlan and de Wet 1971, Andersson & de Vicente 2010, The Plant List 2010, GBIF 2012, Wiersema *et al.* 2012, GRIN 2013, Vincent *et al.* 2013). Second, advancements in geographic information systems, both through the use of increasingly high resolution ecological (e.g., climatic, soil, ecosystem classification, and human impact) data and evolving models and methods, are enabling a more comprehensive conceptualization of the geographic distribution of crop diversity and its conservation needs (Hijmans & Spooner 2001, Hijmans *et al.* 2005, Jarvis *et al.* 2005, Elith *et al.* 2006; Phillips *et al.* 2006, Phillips & Dudik 2008, Costa *et al.* 2010, Ramírez-Villegas *et al.* 2010, Parra-Quijano *et al.* 2011, Hengl *et al.* 2014, Tapia *et al.* 2014). In addition, rapidly progressing classical and genomic tools, technologies, and methods are facilitating their use in crop breeding (Tanksley & McCouch 1997, Zamir 2001, Ford-Lloyd *et al.* 2011, Volk & Richards 2011, McCouch *et al.* 2012). Finally, there is a growing interest in the use of exotic genetic diversity in order to confront increasingly difficult pest and disease, abiotic stress, and other agronomic challenges (Hoisington *et al.* 1999, Gepts 2006, Ortiz *et al.* 2008, Jarvis *et al.* 2009, Tester and Langridge 2010, Guarino & Lobell 2011, McCouch *et al.* 2013, Henry 2014).

Crop wild relatives are genetic resources at the nexus of a number of critical global challenges. They have the potential to contribute significantly to crop improvement, helping to address food security and development goals, while improving the adaptation of crops to climate change (Hoisington *et al.* 1999, Guarino & Lobell 2011, Dempewolf *et al.* 2014). As wild plants, they are subject to a myriad of human caused threats to natural ecosystems, including habitat modification, urbanization, mining, logging, changing fire regimes, pollution, invasive species, overharvesting, and climate change (Stein *et al.* 2000, Wilkes 2007, Jarvis *et al.* 2008, Loarie *et al.* 2009, Brummitt & Bachman 2010, Khoury *et al.* 2010, Ureta *et al.* 2011). A focus on wild genetic resources is thus timely both for conservation and crop improvement objectives.

The research presented in this thesis was performed in contribution to a number of ongoing crop wild relative initiatives active at different scales globally. "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives", is a ten year global project aimed at identifying important crop wild relatives worldwide, assessing their representation in genebanks, filling gaps in these *ex situ* collections through targeted collecting, and pre-breeding collected samples in order to provide breeders with materials of interest for adapting their crop varieties to climate change (Dempewolf *et al.* 2014). On the national level, the United States is bringing together information needed to conceptualize the crop wild relative diversity present in the country, assess its conservation needs, and organize strategies for long-term management of this diversity by government, private, and other stakeholders (Plant Germplasm Operations Committee 2010). Within the global CGIAR Research Programs, including Roots, Tubers and Banana, and Grain Legumes, assessments are being made of the state of conservation of crop wild relatives of mandate crops, with the aim of making the diversity present in these species available to breeders worldwide (CGIAR 2015).

While conservation of crop genetic resources is fundamental to the availability of this diversity for present and future breeding, it is not the only major constraint to utilization. National and international policies on crop genetic resources determine the real capacity for researchers to acquire diversity of potential interest (Fowler & Hodgkin 2004, Esquinas-Alcázar 2005, Bjørnstad *et al.* 2013). The thesis culminates in an exploration of the implications of the global geographic distribution of crop genetic diversity for food security, in particular the level to which international collaboration is required in order to achieve access to genetic resources where they are needed. This research was performed in contribution to current negotiations for the potential expansion of scope and membership within the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty) (FAO 2002).

THESIS OUTLINE

This thesis is organized into eight chapters (Figure 1). Following this general introduction, Chapter 2 contextualizes the need for the use of diverse genetic resources in crop breeding in order to mitigate vulnerability in global food supplies. This chapter represents an exploration of central assumptions in the field of plant genetic resources – that humanity relies on relatively few crops for its survival, and that this list of important crops is growing smaller over time - inspired by a seminal article in the field that was published 25 years ago using data from 35

Chapter 1

years before the present (Prescott-Allen & Prescott-Allen 1990). Research questions include: what is the state of crop species diversity in global food supplies?, what is the trend over time in change in this diversity?, what are the most important crops globally?, and, what options might be available to diversify food supplies in the future?

From this global context, Chapter 3 concentrates on the identification of potentially important wild genetic resources at the national level. Focusing on the United States, a large and eco-geographically diverse country with relatively advanced conservation policy and active national genetic resources conservation efforts, the chapter presents an inventory of crop wild relatives and other wild species of potential use in agricultural research. The chapter also develops a method for prioritizing these wild species based upon their potential to contribute to food security. Research questions include: how can crop wild relatives and other useful wild plants be documented at the national level?, and, how can they be prioritized for conservation and research actions?

Once potentially valuable crop genetic resources are identified, subsequent information is needed regarding where they occur, what diversity they may possess, and how well conserved and therefore available to crop breeders they are. Chapter 4 offers a methodology to answer these questions at the crop genepool level, i.e., for the wild relatives associated with any particular crop. This chapter builds upon developments in the generation of and access to digital occurrence and eco-geographic data as well as improvements in modeling wild plant species distributions, intentionally utilizing freely available software and data. The method also includes a novel expert assessment methodology, using researchers knowledgeable in the distributions and conservation concerns of crop wild relatives to evaluate the results. Research questions include: how can modeled geographic distributions of crop wild relatives be used to assess the comprehensiveness of the diversity represented in existing genebank collections?, how can priorities for further collecting in order to fill gaps in the conservation of species in genebanks be outlined at the species and population level?, and, how valid are these results compared to expert opinion, and thus what are the potential limitations of the method?

Chapters 5 and 6 utilize the basic methodology outlined in Chapter 5, but also advance the method by capitalizing on knowledge learned over the four years of doctoral research completed since the publication of the original gap analysis case study. Along with

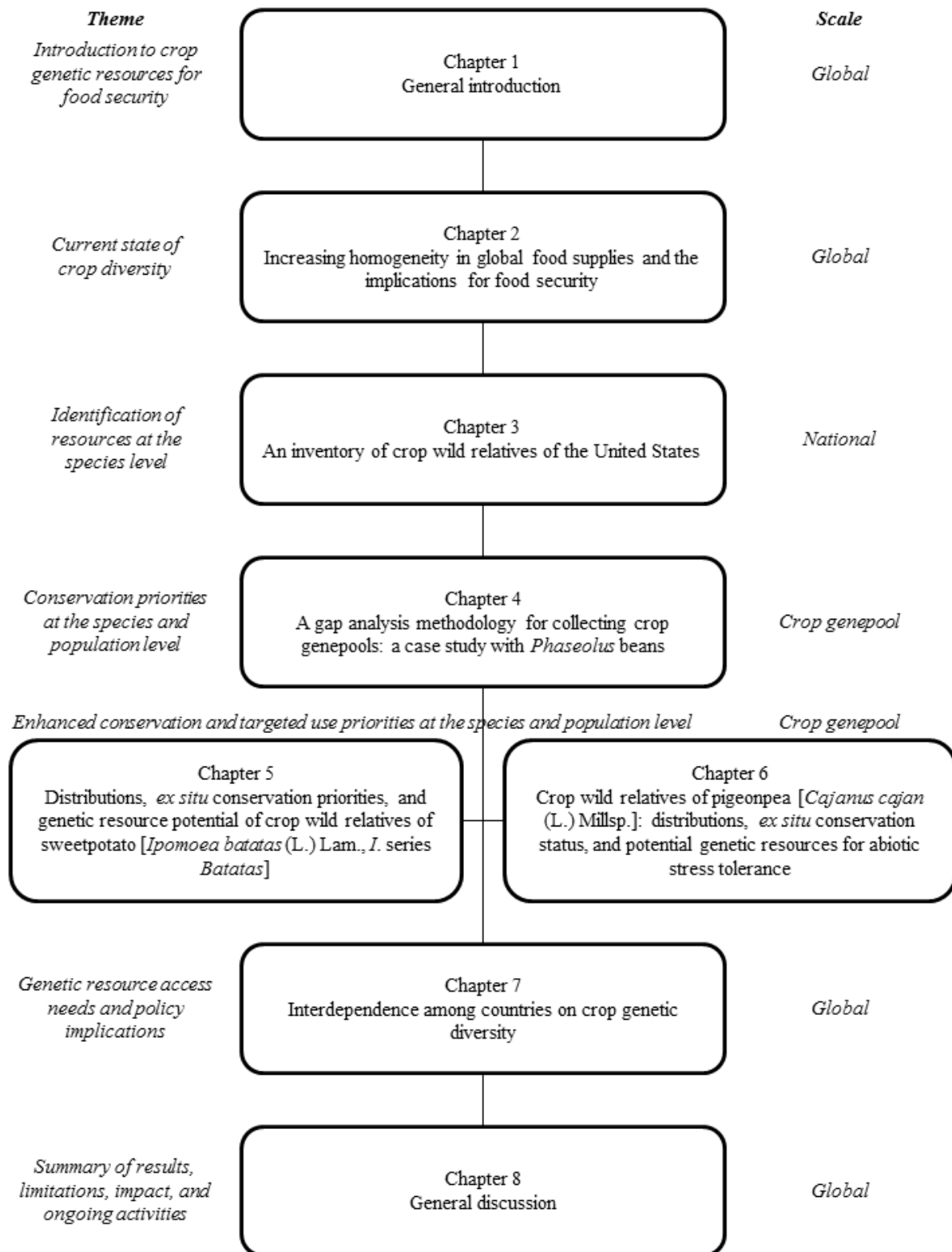


Figure 1. Thesis structure

Chapter 1

improvements in species targeting, occurrence data, modeling, and expert feedback methods, the chapters take an additional step by drawing upon eco-geographic information to indicate the potential for species and specific populations to possess traits of value to crop improvement objectives, particularly for abiotic stress tolerance. The chapters focus on crops differing substantially in the state of existing information regarding associated wild relatives, as well as historical use of the resources in crop improvement. Research questions include: what constitutes a potentially useful wild relative of sweetpotato and pigeonpea?, where are these species encountered?, what is the state of conservation and availability of these species to researchers?, what are the highest taxonomic and ecogeographic priorities for further collecting?, and, what traits do species and populations possess that may be valuable to specific crop improvement objectives?

Chapter 7 focuses on access to crop genetic resources. Given the geography of crop genetic diversity identified in previous chapters and additional literature stretching back to the works of N. I. Vavilov (1926), the chapter investigates the disjunct between where crop genetic resources originated and are particularly diverse, and where crops themselves are currently important in diets and production systems. The overall goal is to assess the degree to which countries are interdependent in regard to crop genetic resources, and therefore need to engage in international exchange in order to support the crop breeding that underlies their national food supplies and production systems. As with Chapter 2, this research was inspired by an important historical article (Flores-Palacios 1998), which contributed to the initial arguments for the creation of the Plant Treaty (FAO 2002). Publication of Chapter 7 is intended to increase membership and contribute to the rationale for expansion of the crops covered under the Plant Treaty. Research questions include: how does the distribution of crop genetic diversity associated with the world's major crops indicate interdependence among countries in regard to crop breeding efforts?, how dependent are countries on crop diversity from outside their borders in order to support their food supplies and production systems?, how is this dependence changing over time?, and what are the implications of this dependence in regard to current political efforts focused on access to crop genetic resources?

Chapter 8 provides a general discussion of the research and its main results. The chapter offers a summary of the current status of conservation of the crop genetic resources targeted in the thesis, identifying limitations in the study and suggesting future research in order to make

further progress on key questions in the field. The chapter also discusses the impact of the research to date, and highlights ongoing activities that are building upon the efforts documented here.

REFERENCES

- Alston, J.M., Beddow, J.M., Pardey, P.G. (2009). Agricultural research, productivity, and food prices in the long run. *Science* 325 (5945), 1209-1210.
- Andersson, M.S., de Vicente, M.C. (2010). *Gene Flow Between Crops and their Wild Relatives* (Baltimore: The Johns Hopkins University Press).
- Björnstad, Å., Tekle, S., Göransson, M. (2013). 'Facilitated access' to plant genetic resources: does it work? *Gen. Res. Crop Evol.* 60, 1959–1965.
- Bodirsky, B.L., Popp, A., Lotze-Campen, H., Dietrich, J.P., Rolinski, S., Weindl, I., *et al.* (2014). Reactive nitrogen requirements to feed the world in 2050 and potential to mitigate nitrogen pollution. *Nat. Commun.* 5, 3858.
- Brummitt, N., Bachman, S. (2010). *Plants Under Pressure: A Global Assessment. The First Report of the IUCN Sampled Red List Index for Plants* (London: Natural History Museum, UK).
- Burke, M.B., Lobell, D.B., Guarino, L. (2009). Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environ. Chang.* 19 (3), 317-325.
- CDC (Centers for Disease Control and Prevention). (2015). Micronutrient Facts. <http://www.cdc.gov/impact/micronutrients/>.
- CGIAR. 2015. CGIAR Research Programs. <http://www.cgiar.org/our-research/cgiar-research-programs/>.
- Cordell, D., Drangert, J-O., White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Chang.* 19 (2), 292–305.
- Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodivers. Conserv.* 19, 883-899.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sust. Food Syst.* 38 (4), 369-377.
- Diamond, J. (2004). The wealth of nations. *Nature* 429, 616-617.
- Ehrlich, P.R. (1975). *The Population Bomb* (Mattituck, New York: Ameron).
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- Esquinas-Alcázar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* 6 (12), 946–953.
- FAO (Food and Agriculture Organization of the United Nations). (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2002). *International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- Flores-Palacios, X. (1998). *Contribution to the Estimation of Countries' Interdependence in the Area of Plant Genetic Resources*. Commission on Genetic Resources for Food and Agriculture, Background Study Paper No. 7, Rev. 1 (Rome: FAO).
- Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., *et al.* (2005). Global consequences of land use. *Science* 309 (5734), 570-574.
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., *et al.* (2011). Solutions for a cultivated planet. *Nature* 478, 337–342.
- Ford-Lloyd, B., Schmidt, M., Armstrong, S.J., Barazani, O., Engels, J., Hadas, R., *et al.* (2011). Crop wild relatives - undervalued, underutilized and under threat? *BioScience* 61, 559-565.

- Fowler, C., Hodgkin, T. (2004). Plant genetic resources for food and agriculture: assessing global availability. *Annu. Rev. Env. Resour.* 29 (1), 143-179.
- GBIF (Global Biodiversity Information Facility). (2012). *GBIF Data Portal* (Copenhagen: GBIF). <http://data.gbif.org/welcome.htm>.
- Gepts, P. (2006). Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci.* 46 (5), 2278-2292.
- Global Crop Diversity Trust. (2013). *Fundraising Strategy 2014-2018* (Bonn: Global Crop Diversity Trust).
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., *et al.* (2010). Food security: The challenge of feeding 9 billion people. *Science* 327, 812–818.
- Gomez, M.I., Barrett, C.B., Raney, T., Pinstrip-Andersen, P., Meerman, J., Croppenstedt, A. *et al.* (2013). *Post-Green Revolution Food Systems and the Triple Burden of Malnutrition*. ESA Working Paper No. 13-02 (Rome: Agricultural Development Economics Division, FAO).
- Graham, R.D., Welsh, R.M., Saunders, D.A., Ortiz-Monasterio, I., Bouis, H.E., Bonierbale, M., *et al.* (2007). Nutritious subsistence food systems. *Adv. Agron.* 92, 1-74.
- GRIN (Germplasm Resources Information Network). (2013). USDA, ARS, National Genetic Resources Program. <http://www.ars-grin.gov/cgi-bin/npgs/html/wep.pl>.
- Guarino, L., Lobell, D.B. (2011). A walk on the wild side. *Nat. Clim. Change* 1 (8), 374-375.
- Gur, A., Zamir, D. (2004). Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol.* 2(10), e245.
- Hajjar, R., Hodgkin, T. (2007). The use of wild relatives for crop improvement: a survey of developments over the past 20 years. *Euphytica* 156, 1-13.
- Harlan, J.R. (1971). Agricultural origins: centres and noncentres, *Science* 174, 468-474.
- Harlan, J.R. (1975). *Crops and Man* (Madison: American Society of Agronomy).
- Harlan, J.R., de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20 (4), 509-517.
- Hawkesworth, S., Dangour, A.D., Johnston, D., Lock, K., Poole, N., Rushton, J., *et al.* (2010). Feeding the world healthily: the challenge of measuring the effects of agriculture on health. *Philos. T. R. Soc. B.* 365 (1554), 3083–3097.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., *et al.* (2014). SoilGrids1km - global soil information based on automated mapping. *PLoS One* 9 (8), e105992.
- Henry, R.J. (2014). Genomics strategies for germplasm characterization and the development of climate resilient crops. *Front. Plant Sci.* 5, 68.
- Hijmans, R.J., Cameron, S., Parra, J., Jones, P.G., Jarvis, A. (2005). *WorldClim*, Version 1.3 (Berkeley: University of California).
- Hijmans, R.J., Spooner, D. (2001). Geographic distribution of wild potato species. *Am. J. Bot.* 88, 2101-2112.
- Hoisington, D., Khairallah, M., Reeves, T., Ribout, J.-M., Skovmand, B., Taba, S., *et al.* (1999). Plant genetic resources: what can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci.* 96, 5937-5943.
- Iltis, H.H. (1988). “Serendipity in the exploration of biodiversity. what good are weedy tomatoes?,” in: *Biodiversity*, ed. E.O. Wilson (Washington D.C.: National Academy Press), 98-105.
- International Food Policy Research Institute. 2014. *Global Nutrition Report 2014: Actions and Accountability to Accelerate the World’s Progress on Nutrition* (Washington, DC: International Food Policy Research Institute).
- Jarvis, A., Lane, A., Hijmans, R.J. (2008). The effect of climate change on crop wild relatives. *Agr. Ecosyst. Environ.* 126 (1-2), 13–23.

- Jarvis, A., Ramirez-Villegas, J., Hanson, J., Leibing, C. (2009). "Crop and forage genetic resources: international interdependence in the face of climate change," in: *The Impact of Climate Change on Countries' Interdependence on Genetic Resources for Food and Agriculture*. Background Study Paper No.48, eds. S. Fujisaka, D. Williams, M. Halewood (Rome: Commission on Genetic Resources for Food and Agriculture, FAO).
- Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet. Resour. Crop Ev.* 52 (6), 671-682.
- Kastner, T., Rivas, M.J.I., Koch, W., Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. *Proc. Natl. Acad. Sci.* 109 (18), 6868-6872.
- Kearney, J. (2010). Food consumption trends and drivers. *Philos. T. R. Soc. B.* 365 (1554), 2793-2807.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57 (4), 625-639.
- Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B., Ackerly, D.D. (2009). The velocity of climate change. *Nature* 462 (7276), 1052-1055.
- Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* 319 (5863), 607-610.
- Malthus, T.R. (1803). *An Essay on the Principle of Population; or, a view of its past and present effects on human happiness; with an enquiry into our prospects respecting the future removal or mitigation of the evils which it occasions* (London: John Murray).
- Matson, P.A., Vitousek, P.M. (2006). Agricultural intensification: Will land spared from farming be land spared for nature? *Conserv. Biol.* 20 (3), 709-710.
- Maxted, N., Kell, S.P., Ford-Lloyd, B.V., Dulloo, E., Toledo, Á. (2012). Toward the systematic conservation of global crop wild relative diversity. *Crop Sci.* 52, 774-785.
- McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., *et al.* (2013). Agriculture: feeding the future. *Nature* 499 (7456), 23-24.
- McCouch, S.R., McNally, K.L., Wang, W., Sackville Hamilton, R. (2012). Genomics of gene banks: A case study in rice. *Am. J. Bot.* 99 (2), 407-423.
- McCouch, S.R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septinginsih, E., *et al.* (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica* 154, 317-339.
- Ortiz, R., Sayre, K.D., Govaerts, B., Gupta, R., Subbarao, G.V., Ban, T., *et al.* (2008). Climate change: can wheat beat the heat? *Agr. Ecosyst. Environ.* 126, 46-58.
- Parra-Quijano, M., Iriondo, J.M., Torres, E. (2011). Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodivers. Conserv.* 21, 79-96.
- Phalan, B., Onia, M., Balmford, A., Green, R.E. (2011). Reconciling food production and biodiversity conservation: Land sharing and land sparing compared. *Science* 333 (6047), 1289-1291.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231-259.
- Phillips, S.J., Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and comprehensive evaluation. *Ecography* 31, 161-175.
- Pingali, P. (2007). Westernization of Asian diets and the transformation of food systems: implications for research and policy. *Food Pol.* 32 (3), 281-298.
- Plant Germplasm Operations Committee. (2010). *Minutes of the Annual Meeting, Geneva, NY. 28-29 July 2010* (Beltsville: USDA National Germplasm Resources Laboratory).
- Popkin, B.M. (2006). Technology, transport, globalization and the nutrition transition food policy. *Food*

- Pol.* 31 (6), 554-569.
- Prescott-Allen, R., Prescott-Allen, C. (1990). How many plants feed the world? *Conserv. Biol.* 4 (4), 365-374.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. *PLoS One* 5, e13497.
- Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8 (6), e66428.
- Ray, D.K., Ramankutty, N., Mueller, N.D., West, P.C., Foley, J.A. (2012). Recent patterns of crop yield growth and stagnation. *Nat. Commun.* 3, 1293.
- Remans, R., Flynn, D.F.B., DeClerk, G., Diru, W., Fanzo, J., Gaynor, K., *et al.* (2011). Assessing nutritional diversity of cropping systems in African villages. *PLoS One* 6 (6), e21235.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., *et al.* (2009). A safe operating space for humanity. *Nature* 461 (7263), 472-475.
- Ruel, M.T. (2003). Operationalizing dietary diversity: a review of measurement issues and research priorities. *J. Nutr.* 133 (11), 3911S-3926S.
- Stein, B.A., Kutner, L.S., Adams, J.S. (2000). *Precious Heritage: The Status of Biodiversity in the United States* (New York: Oxford University Press).
- Struik, P.C., Kuyper, T.W. (2014). Editorial overview: sustainable intensification to feed the world: concepts, technologies and trade-offs. *Cur. Op. Environ. Sustain.* 8, vi – viii.
- Tanksley, S.D., McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277 (5329), 1063-1066.
- Tapia, C., Torres, E., Parra-Quijano, M. (2014). Searching for adaptation to abiotic stress: ecogeographical analysis of highland Ecuadorian maize. *Crop Sci.* 55 (1), 262-274.
- Tester, M., Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327 (5967), 818-822.
- The Plant List. (2010). Version 1. <http://www.theplantlist.org/>.
- Thormann, I., Fiorino, E., Halewood, M., Engels, J.M.M. (2015). Plant genetic resources collections and associated information as a baseline resource for genetic diversity studies: an assessment of the IBPGR-supported collections. *Genet. Resour. Crop Evol.* doi 10.1007/s10722-015-0231-9.
- Tilman, D., Balzer, C., Hill, J., Belfort, B.L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci.* 108, 20260-20264.
- Tilman, D., Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature* 515 (7528), 518-522.
- Ureta, C., Martínez-Meyer, E., Perales, H.R., Alvarez-Buylla, E.R. (2011). Projecting the effects of climate change on the distribution of maize races and their wild relatives in Mexico. *Glob. Change Biol.* 18, 1073-1082.
- van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., Visser, B. (2009). Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources* 8, 1-15.
- van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* 120 (6), 1241-1252.
- Vavilov, N.I. (1926). Studies on the origin of cultivated plants. *Bulletin of Applied Botany, Genetics and Plant Breeding* 16, 1-248.
- Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., *et al.* (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265-275.
- Vogt, W. (1948). *Road to Survival* (New York: W. Sloane Associates).
- Volk, G.M., Richards, C.M. (2011). Integration of georeferencing, habitat, sampling, and genetic data

- for documentation of wild plant genetic resources. *Hortic. Sci.* 46 (11), 1446–1449.
- West, P.C., Gerber, J.S., Engstrom, P.M., Mueller, N.D. Brauman, K.A., Carlson, K.M., *et al.* (2014). Leverage points for improving global food security and the environment. *Science* 345 (6194), 325–328.
- Wheeler, T., von Braun, J. (2013). Climate change impacts on global food security. *Science* 341 (6145), 508-513.
- WHO (World Health Organization). (2015). Micronutrient deficiencies. <http://www.who.int/nutrition/topics/vad/en/>.
- Wiersema, J.H., León, B., Garvey, E.J. (2012). Identifying wild relatives of subtropical and temperate fruit and nut crops. *Acta Hort.* 948, 285-288.
- Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. A modest proposal for teosinte evolution and conservation *in situ*: The Balsas, Guerrero, Mexico. *Maydica* 52, 49-58.
- Xiao, J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D. (1996). Genes from wild rice improve yield. *Nature* 384, 223-224.
- Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Genet.* 2, 983-989.
- Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., *et al.* (2000). Genetic diversity and disease control in rice. *Nature* 406 (6797), 718-722.

CHAPTER 2

Increasing homogeneity in global food supplies and the implications for food security

Colin K. Khoury^{1,2}, Anne D. Bjorkman^{3,4}, Hannes Dempewolf^{4,5,6},
Julian Ramirez-Villegas^{1,7,8}, Luigi Guarino⁶, Andy Jarvis^{1,7}, Loren H. Rieseberg^{4,5,9}, and
Paul C. Struik²

¹ International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia.

² Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

³ Department of Geography, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

⁴ Department of Botany, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

⁵ The Biodiversity Research Centre, University of British Columbia, Vancouver, BC, Canada V6T 1Z4.

⁶ Global Crop Diversity Trust, Platz der Vereinten Nationen 7, 53115 Bonn, Germany.

⁷ CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Cali, Colombia.

⁸ School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, United Kingdom.

⁹ Department of Biology, Indiana University, Bloomington, Indiana, 47405.

Published in *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) 111 (11), 4001-4006. doi: 10.1073/pnas.1313490111

Abstract

The narrowing of diversity in crop species contributing to the world's food supplies has been considered a potential threat to food security. However, changes in this diversity have not been quantified globally. We assess trends over the past 50 years in the richness, abundance, and composition of crop species in national food supplies worldwide. Over this period national per capita food supplies expanded in total quantities of food calories, protein, fat, and weight, with increased proportions of those quantities sourcing from energy dense foods. At the same time, the number of measured crop commodities contributing to national food supplies increased, the relative contribution of these commodities within these supplies became more even, and the dominance of the most significant commodities decreased. As a consequence, national food supplies worldwide became more similar in composition, correlated particularly with an increased supply of a number of globally important cereal and oil crops, and a decline of other cereal, oil, and starchy root species. The increase in homogeneity worldwide portends the establishment of a global standard food supply, which is relatively species rich in regard to measured crops at the national level, but species poor globally. These changes in food supplies heighten interdependence among countries in regard to availability and access to these food sources and the genetic resources supporting their production, and give further urgency to nutrition development priorities aimed at bolstering food security.

Keywords: Crop diversity, Plant genetic resources, Agricultural development, Global analysis

INTRODUCTION

A shared axiom of ecology and nutrition is that, within certain ranges, diversity enhances the health and function of complex biological systems. Species diversity has been shown to stimulate productivity, stability, ecosystem services, and resilience in natural (Hooper *et al.* 2005, Cadotte *et al.* 2012, Hooper *et al.* 2012, Zhang *et al.* 2012, Gamfeldt *et al.* 2013) and in agricultural ecosystems (Kirwan *et al.* 2007, Hajjar *et al.* 2008, Picasso *et al.* 2008, Bonin & Tracy 2012, Cabell & Oelofse 2012, Davis *et al.* 2012, Kremen & Miles 2012, Mijatović *et al.* 2013). Likewise, variation in food species contributing to diet has been associated with nutritional adequacy (Frei & Becker 2004, Kennedy *et al.* 2005, Graham *et al.* 2007, Negin *et al.* 2009) and food security (Ruel 2003).

The development of sedentary agricultural societies and further rise of modern agriculture is generally considered to have led to a decline in the total number of plant species upon which humans depend for food (Harlan 1975, Gepts 2006), particularly the wild, semi-domesticated, and cultivated vegetables and fruits, spices, and other food plants that supplemented staple crops with the provision of micronutrients and that historically bolstered food security during crop failures (Doughty 1979). Harlan (1975) warned that

“most of the food for mankind comes from a small number of crops and the total number is decreasing steadily. In the United States in the past 40 years, many vegetables and fruits have disappeared from the diet, and the trend is going on all over the world. More and more people will be fed by fewer and fewer crops.”
(Harlan 1975).

More recent analyses of dietary transition in developing countries in association with globalization have noted increases in the diversity of plants contributing to diets locally, along with a Westernization transition in preference of energy dense foods (i.e. animal products, plant oils and sugars) over cereals, pulses, and vegetables, and of particular major crop plants within these food categories over traditional crops (Kearney 2010, Pingali 2007). The impact of such changes on overall crop diversity worldwide has not been comprehensively documented, although recent changes in varietal and allelic level diversity of some crops have been investigated (van de Wouw *et al.* 2009, 2010, Bonneuil *et al.* 2012). Given the potential food security implications of narrowing of the diversity of crop species both in production systems

and in food supplies, an assessment of the global state of crop plant species diversity is warranted.

Here we examine changes in the diversity of the portfolio of crop species upon which humans primarily depend for food security in regard to calories, protein, fat, and food weight. Using national per capita food supply data published by the Food and Agriculture Organization (FAO) of the United Nations, we analysed trends in the richness, abundance, and composition of measured crop commodities in the food supplies of 152 countries comprising 98% of the world's population from 1961-2009.

MATERIALS AND METHODS

We analyzed FAO national per capita food supply data for all available measurements [calories (kcal/capita/day), protein (g/capita/day), fat (g/capita/day), and weight (g/capita/day)], from 1961-2009, utilizing the full set of commodity and country listings, standardized across all years. Food supply from plants represents national production plus imports plus or minus stock changes over the survey period; minus exports, quantities used for seed, animal feed, and in the manufacture of non-food products, and losses during storage and transport (Prescott-Allen & Prescott-Allen 1990). Plant commodities clearly comprised of the same crop species were aggregated into single commodities representing the crop species, e.g. olive oil and olives. After aggregation, 52 crop commodities remained. Animal foods were included within the crop importance analysis as a single aggregated commodity.

In order to align all time periods and include as much of the world's population as possible, the current nations formerly comprising the USSR, Yugoslav SFR, Ethiopia PDR, and Czechoslovakia were aggregated into their former states, with national data merged by weighted average based upon population of the respective states at the respective reporting year. Belgium and Luxembourg were reported together during 1961-1999 and therefore recent years listing the countries separately were merged as above. Data for two countries (Occupied Palestinian Territories and Netherlands Antilles) did not appear consistently reported over the study period and were excluded from the analysis. The resulting 152 comparable countries comprised 98% of the world's population across the study period.

Changes in the spread of crop commodities [based on the presence (>0) or absence of a crop in

a given country in each year] were analyzed using generalized estimating equations with a binomial error distribution, “country” as a grouping factor, and an autoregressive (AR-1) correlation structure (i.e., time-lag correlation), using the “geeM” package in R. Slope parameters from these models were extracted to show the direction and magnitude of change for each commodity. Changes in globally aggregated absolute abundance were non-linear, and were therefore analyzed by subtracting the 1961 abundance value from the 2009 value. Changes in country-level absolute abundance were analyzed the same way, for consistency, and the per-country difference was then averaged across countries. We analyzed changes in relative abundances of crops (i.e., relative to the total value of all crops within a given country) using a generalized linear mixed model with a binomial error distribution, with fixed effects for “year” and “crop” and a random effect of “country.” Comparisons between crops were performed by estimating the slopes of the predicted values for each crop in the model. Error bars were calculated by estimating the maximum and minimum slope values based on the standard errors of the predicted values. Crops for which fat or protein make up less than 1% of their weight were removed from the analysis for those respective measurements. All mixed-effects models were performed using the “lme4” package in R.

Changes in crop commodity richness in national food supplies were analyzed with generalized linear mixed effects models with a Poisson error distribution and “country” as a random effect, with random slopes for “year”, to account for repeated measures over time. The significance of the fixed-effect “year” term in predicting richness was ascertained by conducting a Chi-squared likelihood ratio test on nested models with and without the term. In order to estimate the differences between countries in change in richness we extracted slope coefficients from the random effects of the mixed-effects model for one measurement type, calories. These coefficients represent the magnitude and direction of the change in richness for each country but do not provide estimates of error. One country (Namibia) displayed a markedly inconsistent jump in richness. The slope for this country was therefore derived from data only from consistent years, with slope parameters estimated by weighted averaging of the slope in consistent time periods. Analyses were performed using the “lme4” package in R.

We used Pielou's evenness index to measure how equally crops contributed to a country's food supplies in a given year, with linear mixed effects models in the “lme4” package to measure change in evenness over time, as described above. Because Pielou's evenness index can also

reflect changes in richness, we additionally analyzed changes in dominance, defined as the change in the proportion of the most abundant crop commodity in each country in each year, using generalized linear mixed models with a binomial error distribution with the “lme4” package in R. For evenness and dominance, slope parameters were extracted from the random effects of the mixed models to estimate the direction and magnitude of the change in each country.

We quantified the homogenization of crop commodity composition using Bray-Curtis distance (dissimilarity) from each country to the global centroid (comparable to the global mean commodity composition, which is inclusive of abundance) in each year using the “betadisper” function in the R package “vegan”, with similarity derived by subtracting dissimilarity from 1. A mean increase in similarity-to-centroid therefore represents a decrease in country-to-country variation, and a homogenization of commodity composition over time. We then used linear mixed models with “country” as a random factor to assess the significance of the change in similarity-to-centroid over 1961-2009. Slope parameters were extracted from the random effects of the mixed-effects models to estimate the direction and magnitude of the change in similarity in each country. We also assessed homogenization using a non-metric multidimensional scaling analysis (NMDS) with Bray-Curtis dissimilarities to compare the mean and variance of commodity composition between three measurement years (1961, 1985 and 2009). Circles for each year represent 95% confidence intervals around the mean; smaller circles indicate lower country-to-country variation in composition.

The number of crop species currently important to food supply was estimated by listing for the most recent year (2009) the plant and animal commodities by decreasing importance until the total contribution equaled 90% of each country’s food supply for each measurement, a threshold which is inclusive of major contributors to supply and exclusive of commodities contributing very small quantities (Prescott-Allen & Prescott-Allen 1990) (Supplementary Figure 1). One biological species per commodity was recognized, except in cases where additional species were clearly likely to contribute to that commodity in at least one listed country. For general commodities, the minimum number of species most likely to contribute was identified (Supplementary Table 1).

RESULTS

As a global trend, national per capita food supplies from both plant and animal sources consistently increased over the past 50 years for all variables, with animal foods becoming increasingly important in contribution to protein and oil crops dominating fat food supplies (Supplementary Figure 2).

From 1961-2009 all 52 measured crop commodities, which included both individual crops and groups of crops producing similar products, with the exception of cottonseed oil, increased in geographic spread (i.e. were counted as contributing to food supply in an increasing number of countries) (Figure 1A, Supplementary Figure 3). The major cereals wheat, rice, maize and barley, along with sugar, potatoes, and the general vegetables and fruits commodities maintained their ubiquity in food supplies across the planet over the past 50 years. Crop commodities with the greatest relative changes in spread over this period included oil crops such as soybean, sunflower, palm oil, and rape & mustard. Commodities showing relatively small changes in spread and not already globally ubiquitous included minor cereals such as millets, sorghum, and rye, and starchy root crops such as yams, cassava, and sweet potatoes.

Wheat, rice, maize and other ubiquitous crop commodities were among those with the greatest gains in both relative and absolute abundance in national per capita food supplies over the past 50 years (Figure 1B, Supplementary Figure 4-5). In addition, the degree of increase in spread was generally a good predictor of change in the abundance of the crop commodities in food supplies. For example, oil commodities such as soybean, sunflower, palm oil, and rape & mustard were among the crops showing the greatest average increase in relative abundance in national food supplies, whereas millets, rye, sorghum, yams, cassava, and sweet potatoes showed the largest declines. Such global changes in abundance of measured crops in national food supplies are similarly evident in global aggregated per capita food supply data, which assesses global quantities in relation to total world population (Supplementary Figure 5).

The richness of national per capita food supplies in regard to the 52 measured crop commodities increased consistently over the past 50 years for all variables (Figure 2A-B, Supplementary Figure 6). National food supplies also increased slightly in evenness, indicating greater equality in the relative abundance of each of the crop commodities contributing to per capita food supply within each country, although mixed trends were evident worldwide (Figure 2C-D,

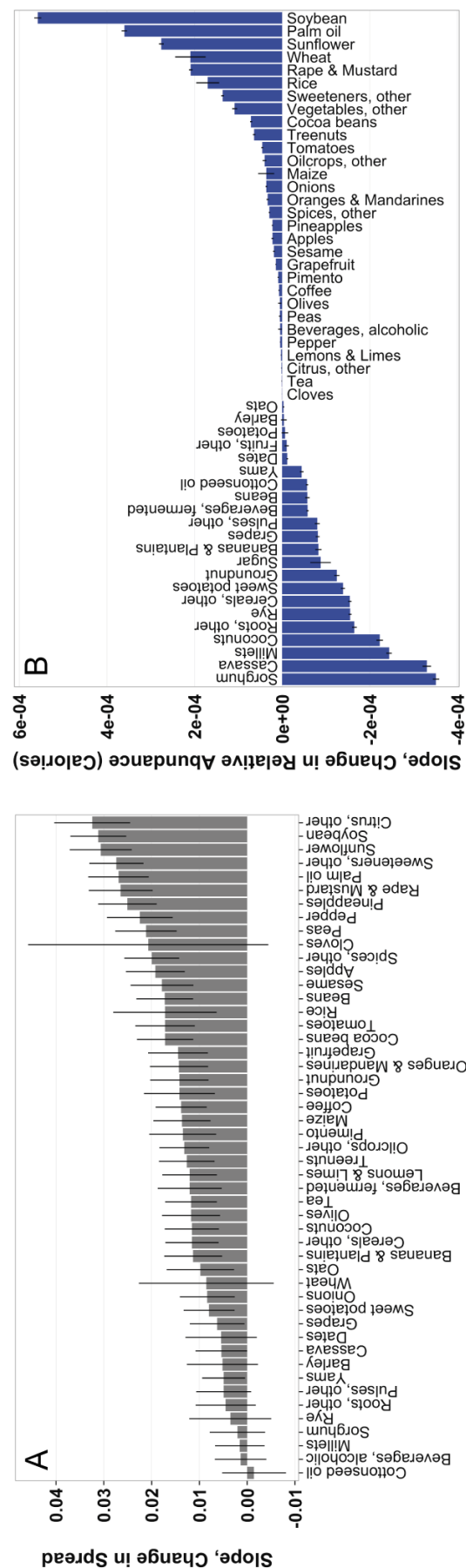


Figure 1. Global change in spread and abundance of crop commodities in national food supplies from 1961-2009. (A) Slope of the relative change in the geographic spread of crop commodities, defined as the change over time in the presence (i.e., value >0) of a crop commodity in contribution to any variable of food supply in a country in each year. Bars represent slope parameters (\pm 95% CI) from generalized estimation equations with a binomial error distribution, country as a grouping factor, and an auto-regressive correlation structure. Sugar; vegetables, other; and fruits, other commodities are not depicted as they were nearly ubiquitous in spread globally throughout the study period and therefore did not change significantly. (B) Slope of the change in relative abundance of crop commodities in contribution to calories, as derived from the value contributed by a particular crop relative to the sum of all crops within a given country within a given year. Bars represent slopes (\pm 95% CI) of the predicted values (1961-2009) for each crop from generalized linear mixed models with “year” and “crop” as fixed effects and country as a random effect. Slopes for change in relative abundance for all measurements are depicted in Supplementary Figure 3.

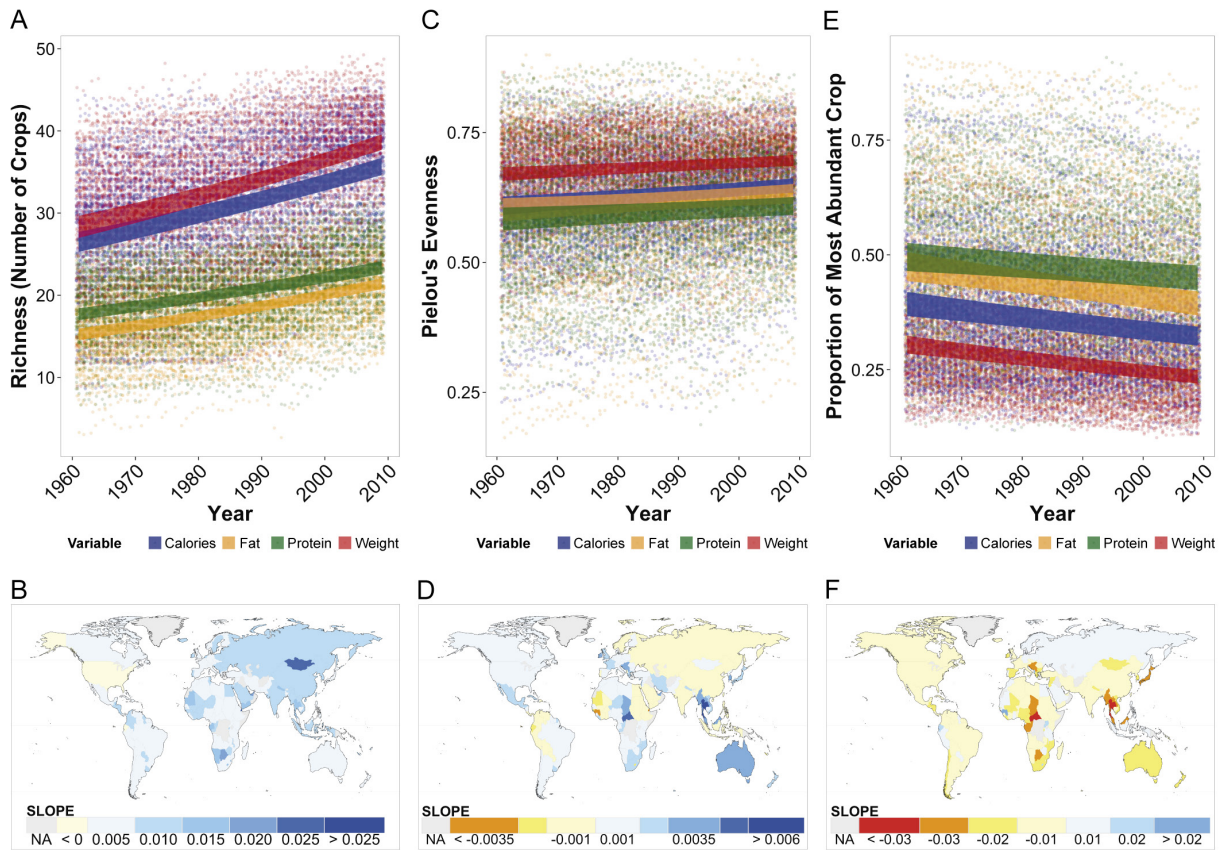


Figure 2. Global change in richness, evenness, and dominance of crop commodities in national food supplies from 1961-2009. (A) Global mean change in crop commodity richness (i.e., number of crops) in national food supplies. Points represent actual data, and lines are 95% prediction intervals from generalized linear mixed models with a Poisson error distribution. (B) World map displaying the slope of change in crop commodity richness per country for calories. (C) Global mean change in evenness of contributing crop commodities within national food supplies calculated by Pielou's evenness index. Points represent actual data, and lines are 95% prediction intervals for a linear mixed effects model. (D) World map displaying the slope of change in evenness of national food supplies for calories. (E) Global mean change in the dominance of the most abundant crop commodity in each country in each year. Points represent actual data, and lines are 95% prediction intervals from a generalized linear mixed model with a binomial error distribution. (F) World map displaying the slope of change in dominance in national food supplies for calories.

Supplementary Figure 7). Dominance (estimated as the proportion of a country's per capita food supply comprised of the most abundant crop commodity) declined as a global trend over the study period (Figure 2E-F, Supplementary Figure 8). Notable reductions in originally very high levels of dominance were visible in rice in contribution to calories in Southeast Asia, coconut for fat in Pacific Island countries, and groundnut for fat in Central African countries.

As national food supplies became richer for the 52 measured crop commodities and relative abundance patterns of contributing commodities shifted, food supplies worldwide became

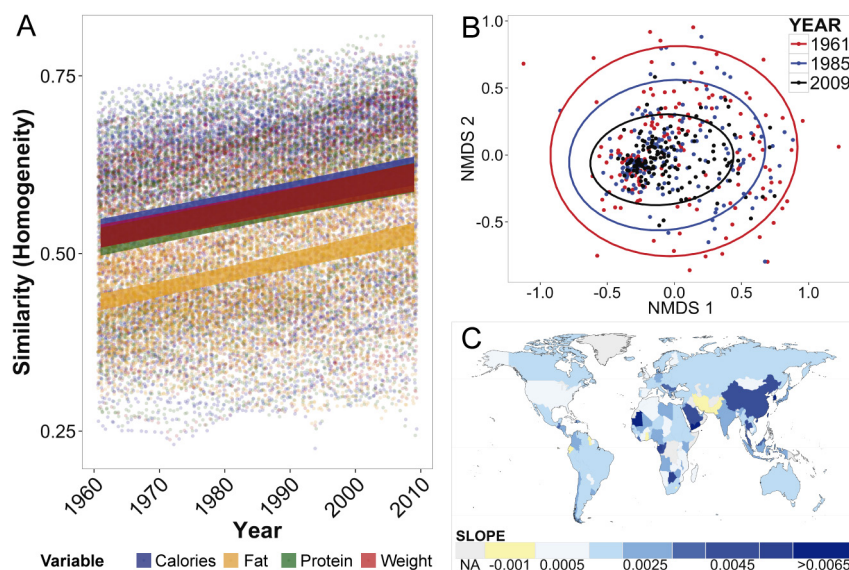


Figure 3. Global change in similarity (homogeneity) of food supplies, as measured by Bray-Curtis dissimilarity from each country to the global centroid (mean composition) in each year, converted to similarity. (A) Global mean change in similarity-to-centroid of national food supplies. Points represent actual data, and lines are 95% prediction intervals from linear mixed effects models. (B) Multivariate ordination of crop commodity composition in contribution to calories in national food supplies in 1961, 1985 and 2009. Red points represent the multivariate commodity composition of each country in 1961, blue points in 1985, and black points in 2009. Circles represent 95% confidence intervals around the centroid in each year. Between 1961 and 2009, the area contained within these 95% confidence intervals decreased by 68.8%, representing the decline in country-to-country variation of commodity composition (i.e., homogenization) over time. (C) World map displaying the slope of change in similarity-to-centroid of national food supplies for calories.

more similar in composition (inclusive of abundance) for all variables (Figure 3, Supplementary Figure 9). This increase in similarity brought national food supplies around the planet closer to a global standard composition. Between 1961 and 2009, homogeneity increased by 16.7%, as measured by the mean change in similarity between each country and the global standard composition, with a maximum (single-country) change of 59.7%. Likewise, mean among-country similarity increased by 35.7%. East and Southeast Asian as well as Sub-Saharan African countries as regional groups displayed the greatest changes in composition toward the global standard over 1961-2009, in association with the greatest increases in measured crop commodity richness and decreases in dominance.

As a measure of the relative importance of crop commodities in total global food supply, inclusive both of plant and animal food sources, we found that 50 of the measured crop commodities currently contribute to the top 90% of calories, protein, fat and weight around the world. We estimate that these crop commodities are composed of 94 crop species from 70

genera from 37 plant families (Supplementary Table 1).

DISCUSSION

The national-level data and multiple variables analyzed here represent the greatest degree of disaggregation of globally comparable food supply data available and thus permit the highest resolution analysis currently feasible for analyzing trends in food crop species utilization worldwide. Food supply data is not directly equivalent to consumption, as food losses at the household level are not measured, but represents a superior measure of the importance of food crops to diets than does production data (Prescott-Allen & Prescott-Allen 1990). National level food supply data both generalizes and underestimates total existing food crop species diversity, as sub-national dietary variation, crops primarily encountered in home gardens and local markets, seasonally important food plants, and culinary herbs, spices and other crops consumed in relatively small quantities are not comprehensively reported in national statistics (FAO 2001). Furthermore, FAO food supply data does not specifically report statistics in regard to micronutrients, where species richness may be particularly significant (Doughty 1979), nor does it assess the indirect contribution to food supply of animal feed and forage plant species. Along with listed crops such as maize and soybean that are major food sources for livestock, global production statistics (FAO 2010a) indicate that crops such as pumpkins, alfalfa, and clover should be recognized as critical indirect contributors to food supply.

More importantly, because of the limited number of individual crop commodities reported by FAO combined with the aggregation of numerous crops into several general commodities, the analyses lack the resolution necessary for elucidating trends in those geographically restricted cereals, legumes, fruits, vegetables and other crops that may be especially sensitive to changes in the global food system. Although some progress has been made in listing and proposing monitoring strategies for such species (Hammer & Khoshbakht 2005, Padulosi & Dulloo 2012), substantial improvements in the systems for reporting food production and supply are needed to adequately monitor such crops globally. Ideally assessments of the state of crop diversity in food supplies would integrate interspecific, varietal, and allelic level analyses of contributing crop plants (Bonneuil *et al.* 2012) in combination with nutritional information (Remans *et al.* 2011) and recorded at subnational levels inclusive of regional, community, household, and individual variation (Pingali 2007, Hawkesworth *et al.* 2010).

The increased food energy, protein, fat, and weight from plants in food supplies worldwide over the past 50 years appear to primarily have been sourced from globally dominant crops specifically reported by FAO. The total number of important crop species we identified remained relatively consistent in comparison to a previous point estimate based on national-level data from 1979-1981 (Prescott-Allen & Prescott-Allen 1990), but the spread and abundance values of these crops have changed measurably. The rate of movement toward homogeneity in food supply compositions globally continues with no indication of slowing. This trend implies a likely deterioration in importance of un-reported minor and geographically restricted food plants, along with the measured cereal, oil, starchy root and other crops that displayed significant declines in abundance in national food supplies. Thus even as the number of measured crops available to the consumer in a given country has increased over the past half-century as a global trend, the total diversity of crops contributing significantly worldwide has narrowed.

A series of interrelated factors are considered to have driven change in food supply compositions worldwide. A primary driver for both diversifying diets and shifting toward increased consumption of energy dense foods entails wealth increase and associated gains in purchasing power. Transitions in dietary preferences, particularly toward Western diets (e.g. meat and dairy, wheat, temperate vegetables and fruit, and sugary beverages) have been associated with many facets of globalization and urbanization, including trade liberalization and the development of extensive commodity transport systems, multinational food industries, food quality and safety standardization, mass media, labor changes, smaller family sizes, supermarkets, fast food, processed foods, and human migration. These drivers are detailed elsewhere (Popkin 2006, Pingali 2007, Hawkesworth *et al.* 2010, Kearney 2010). The modernization and globalization of agriculture through the replacement of human labor with machinery, investments in the breeding and distribution of high yielding major crops as a development strategy, and subsidies dedicated to a narrow range of crop commodities, among other factors, have further contributed to the increasing global availability of a limited number of major crop plants, with lesser priority given to nutritional diversity (Davis *et al.* 2004, Negin *et al.* 2009).

Increases in measured crop species richness and evenness in food supplies may be considered as indicative of enhanced food security on the local level, particularly in regard to availability

and access. On the other hand, the increased ability globally to realize within diets the human preference for energy dense foods (Kearney 2010) based upon a limited number of global crop commodities and processed products has been associated with the rise in non-communicable diseases such as adult onset diabetes, heart disease, and certain forms of cancer (Popkin 2006, Pingali 2007, Kearney 2010), a trend which is impacting rapidly developing countries such as China more quickly than projected (Kearney 2010). Such dietary changes have also contributed to reductions in diversity in human oral and gut microbiota, which in return have negatively impacted health (Lozupone *et al.* 2012, Adler *et al.* 2013). With the number of over-nourished worldwide surpassing the number of under-nourished (Popkin 2006), such diseases are becoming epidemics, including within countries still grappling with significant constraints in food availability. Making available and accessible adequate nutritional diversity within and in supplement to the major crops comprising an ever-greater proportion of global food supplies is critical (Ruel 2003, Graham *et al.* 2007, Remans *et al.* 2011). Engendering consumption patterns cognizant of the impact of food crop choice on health is equally pressing.

The increasing reliance on a suite of truly global crop commodities implies a narrowing in the diversity present in global agricultural systems as a whole, necessitating an equivalent expansion in attention to production stability for these crops (Alston *et al.* 2009). The importance of this stability is exacerbated by the necessity of increasing yields of major crops in order to keep pace with demand, the future success of which has been questioned (Ray *et al.* 2013). From a genetic diversity perspective, increasing homogeneity in global food supplies highlights the importance of the breeding and cultivation of varieties of these crops with diverse genetic backgrounds (National Research Council 1972, Zhu *et al.* 2000). While varietal diversity particularly of major cereal crops can in some regions be relatively high, for other key crops such as banana only a handful of varieties are widely cultivated, despite substantial diversity in the gene pool as a whole (Perrier *et al.* 2011).

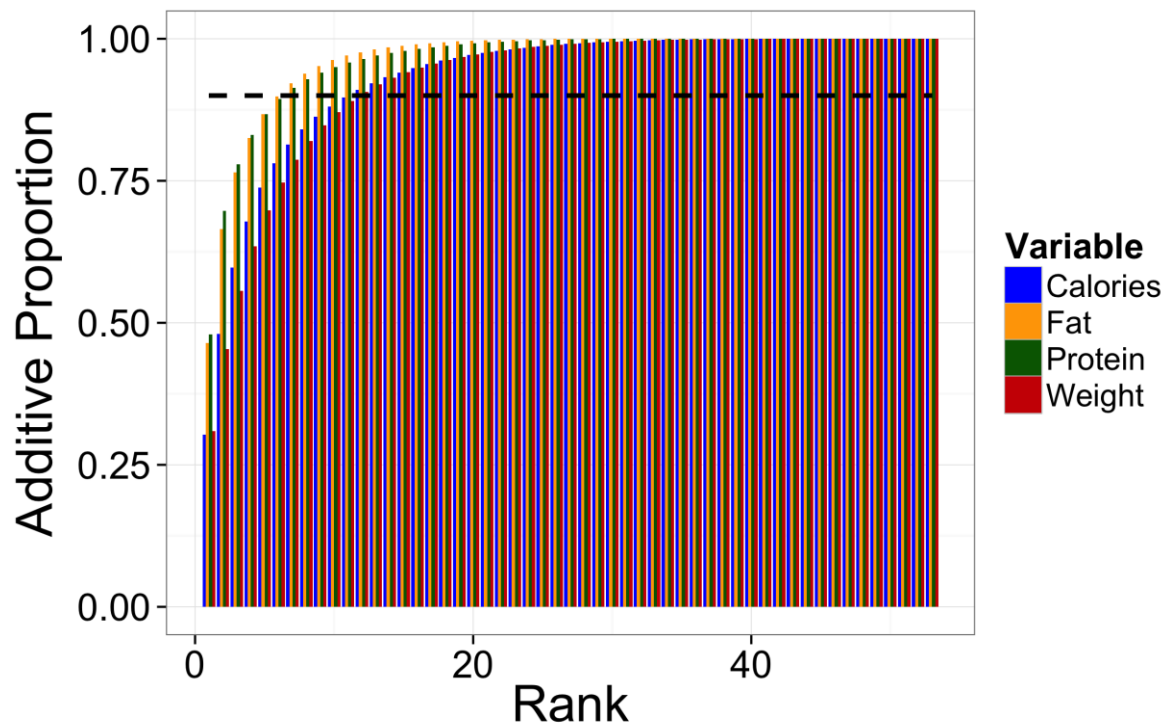
As crop development efforts rely upon the utilization of genetic resources (Gepts 2006), it is a policy imperative to ensure the conservation of, and access to, as wide a range of genetic diversity within these global crops as possible, along with the genotypic and phenotypic information necessary to effectively utilize these resources (McCouch *et al.* 2012, 2013). Unfortunately, significant gaps remain in the conservation of crop genetic resources (FAO 2010b) and access to this diversity requires improvement (Fowler & Hodgkin 2004). A

number of crops considered here are not sufficiently covered under the pertinent international treaty (FAO 2002), including plants significant to food supplies globally, such as sugar cane, soybean and groundnut.

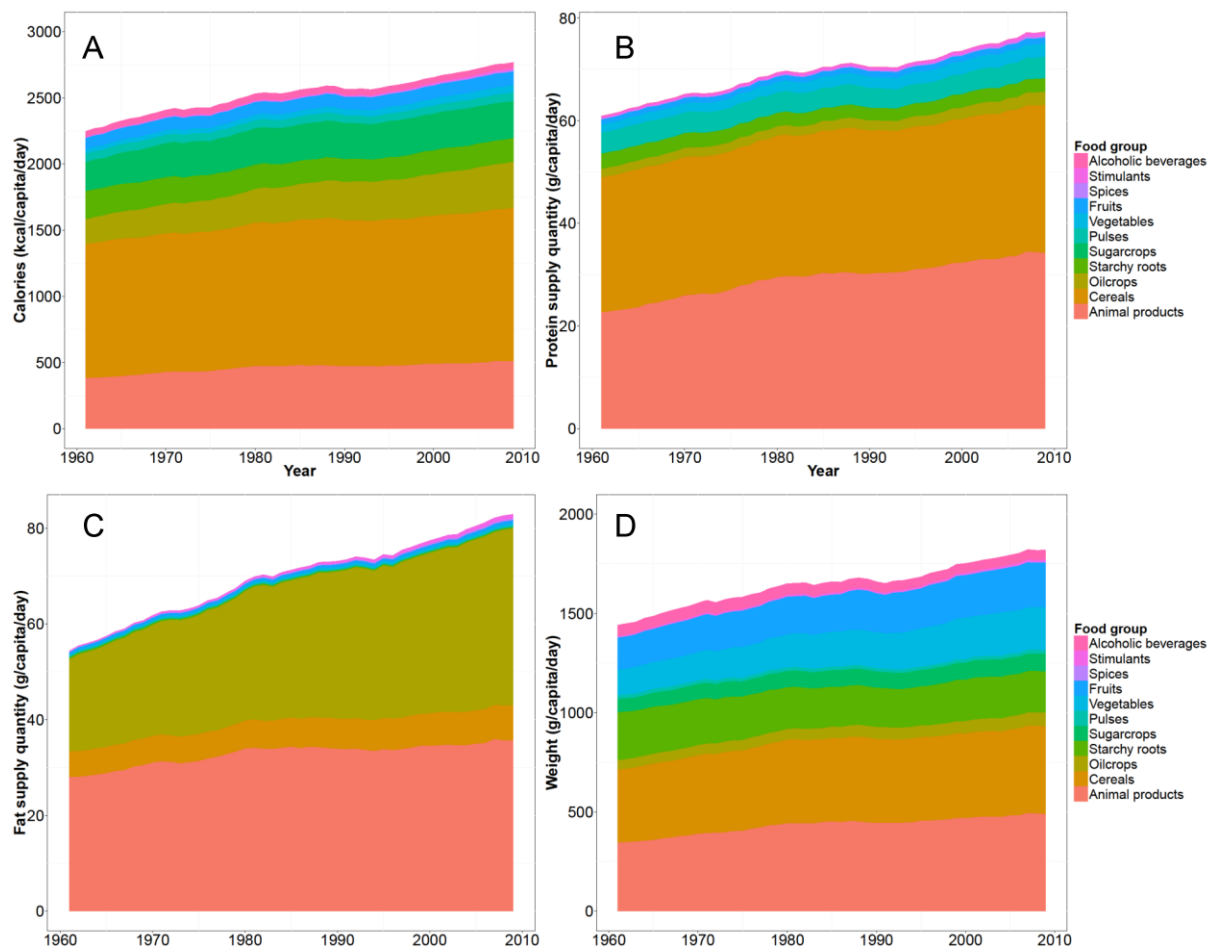
While our study identifies the major crop commodities critical to national food supplies worldwide, current patterns of production of these crops are not guaranteed given ongoing and predicted changes in climate (Lobell *et al.* 2008, Burke *et al.* 2009, Wheeler & von Braun 2013), the decline in availability of non-renewable inputs (Cordell *et al.* 2009), and increasingly severe impacts of agriculture on soil, water quality and biodiversity (Foley *et al.* 2005). Such trends may impact food security in regard to crop commodity trade (Fader *et al.* 2013), decrease the nutritional quality of major crops (Dwivedi *et al.* 2013), and enhance the attractiveness of under-researched crop species (Naylor *et al.* 2004) that are productive and nutritious given limited inputs, particularly under marginal or variable conditions (Jarvis *et al.* 2012).

Moreover the importance of crop commodities in food supplies, particularly in contribution to protein and fat, may shift in response to health, natural resource, and climate pressures (Kearney 2010, Stehfest *et al.* 2009, Kastner *et al.* 2012, Pradhan *et al.* 2013), counteracting the trend demonstrated over the past 50 years in increased animal as well as energy-dense plant food commodity consumption. The trajectory of Northern European food supplies appears to be demonstrating such a trend (Kearney 2010). Providing that alternative food crops may still be encountered, a further diversification of food supplies with interesting and nutritious crop plants may bolster this evolution.

SUPPLEMENTARY INFORMATION

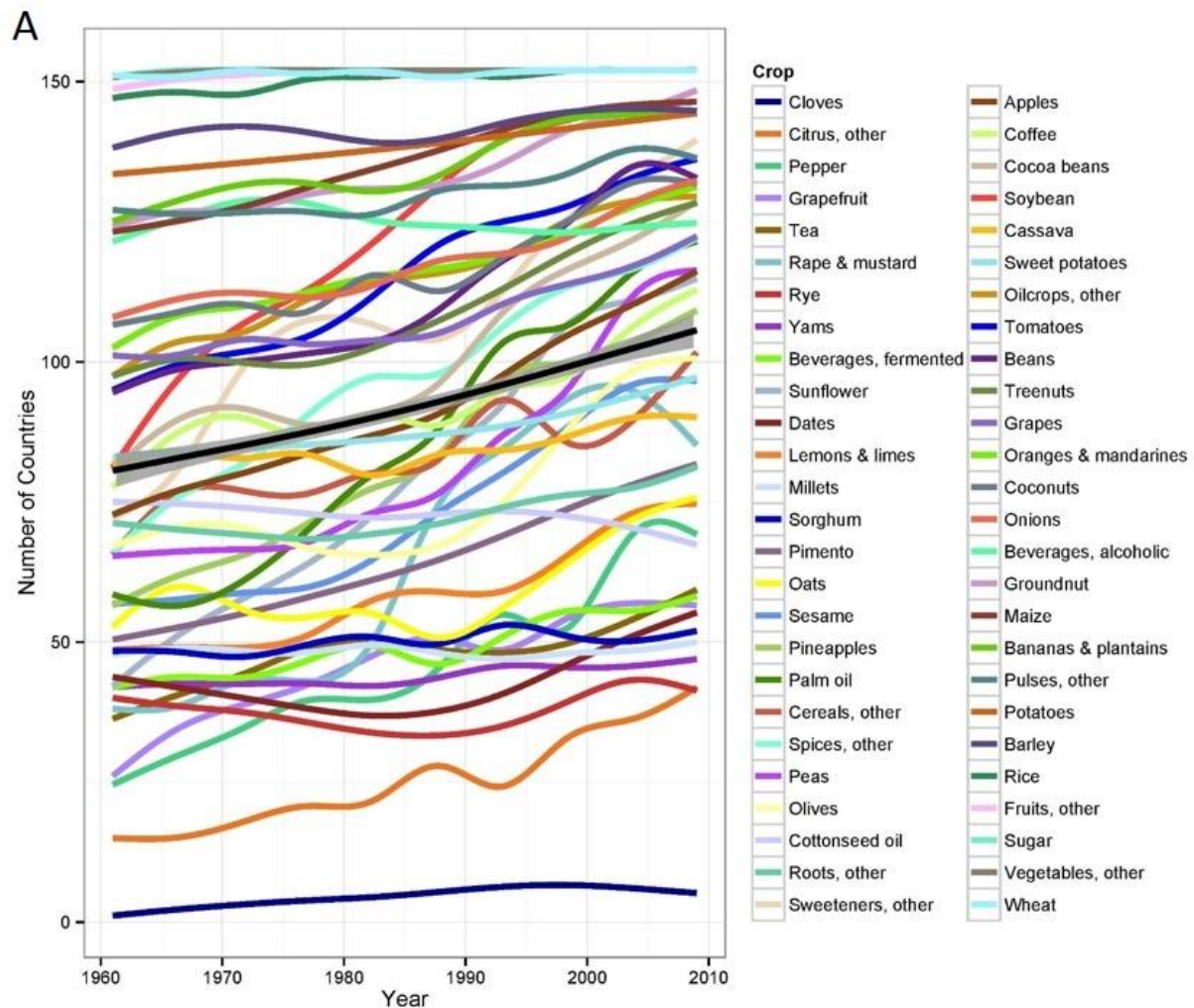


Supplementary Figure 1. Global average additive contribution by ranked abundance of crop plant and animal commodities to calories, protein, fat and weight in food supply, 2009. The dotted line shown at 90% displays the threshold utilized to determine the number of crops considered important to each national food supply in 2009 (Supplementary Table 1), which is inclusive of major contributors to supply and exclusive of commodities contributing very small quantities (Prescott-Allen & Prescott-Allen 1990).

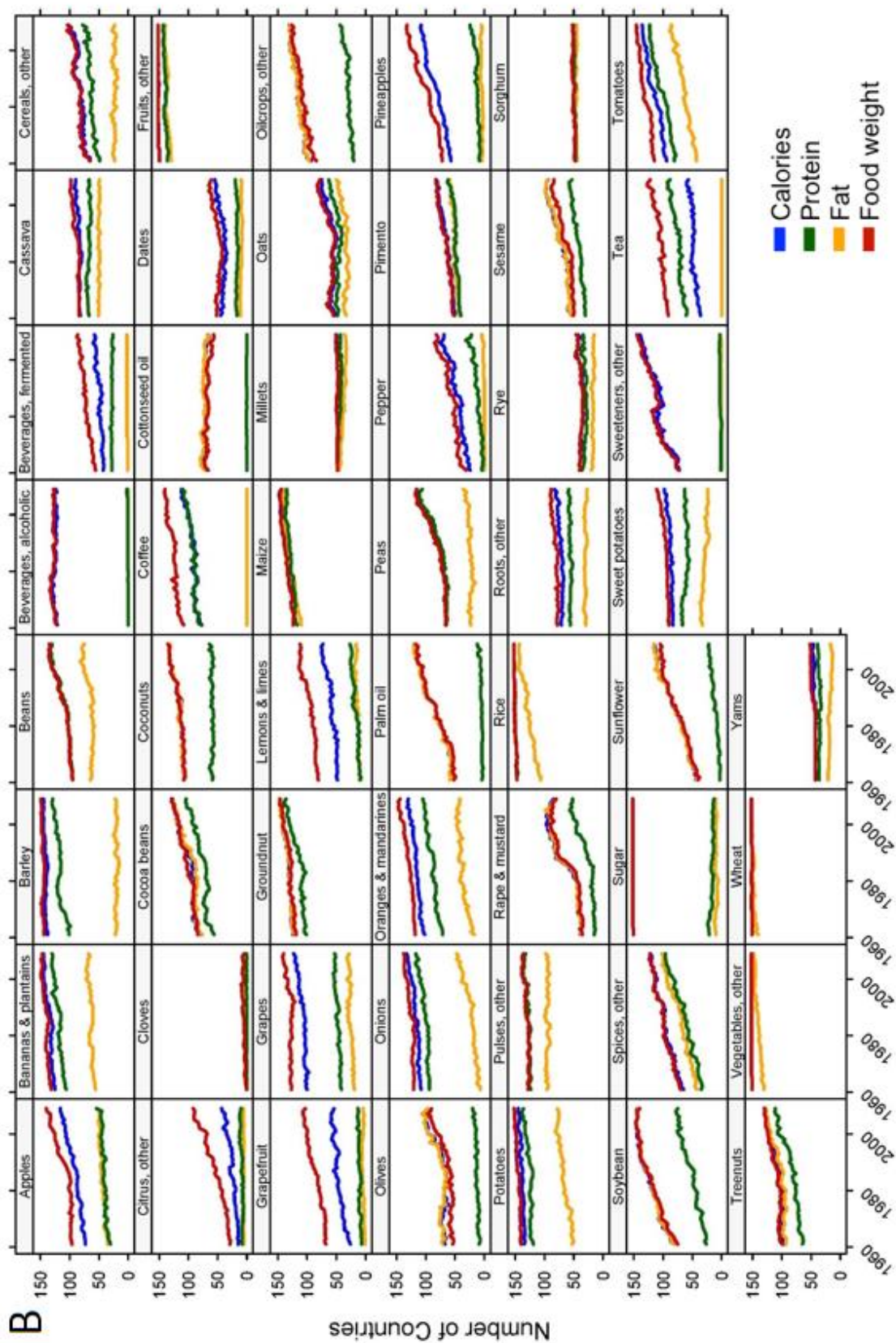


Supplementary Figure 2. Global change in national food supplies by food groups from plant and animal sources, for (A) calories, (B) protein (C) fat and (D) weight, 1961-2009. Data displayed is mean per capita values across all (152) countries.

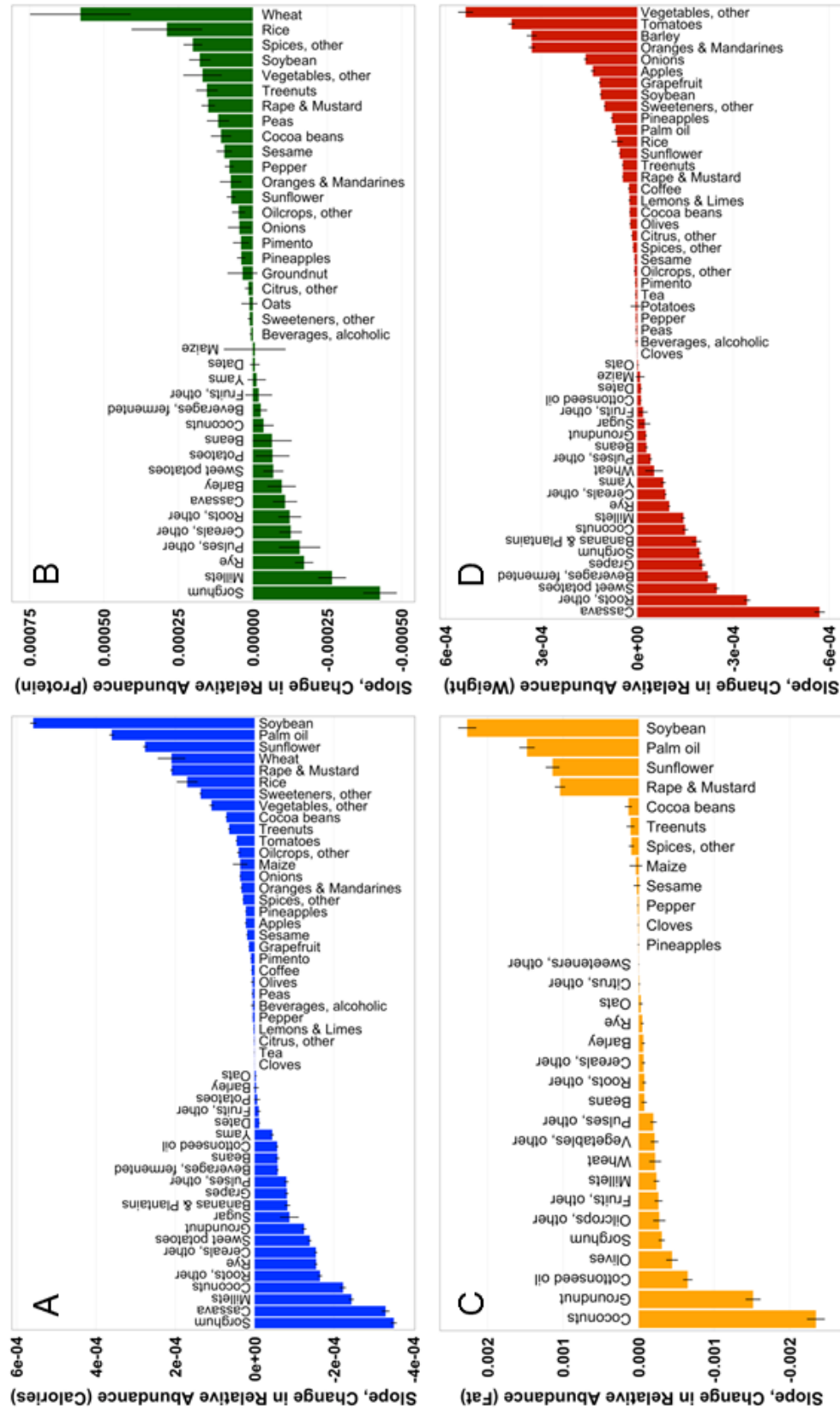
Supplementary Figure 3. Change in the total number of countries reporting each crop commodity, per crop, 1961-2009.



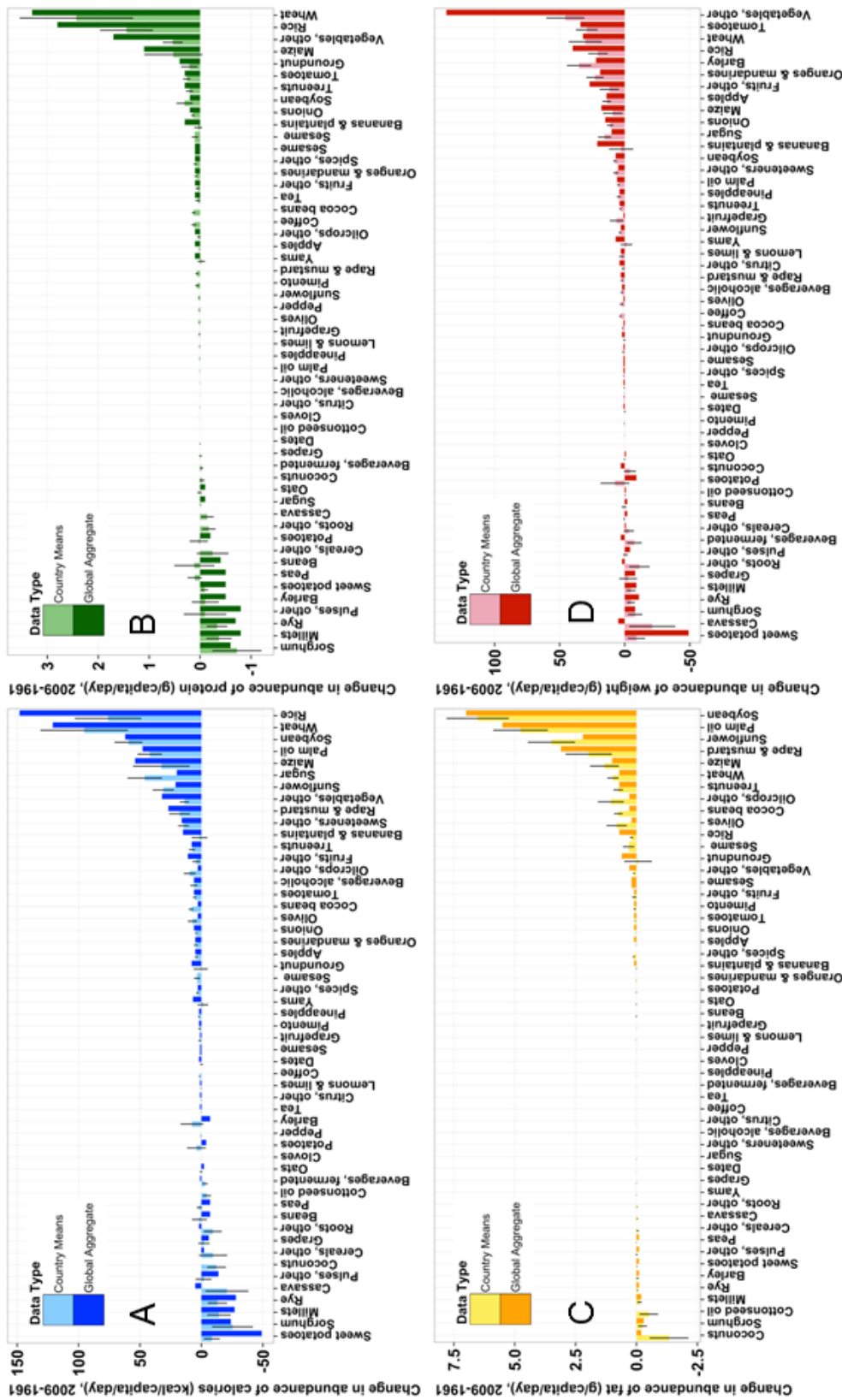
Supplementary Figure 3A. Change in the total number of countries reporting each crop commodity using generalized additive modeling (GAM). Crop commodities are counted as present within a given country in a given year when contributing to food supply (>0) for any variable. Crop commodities are listed in ascending order from the value in 1961. The black line displays the mean value across crop commodities.



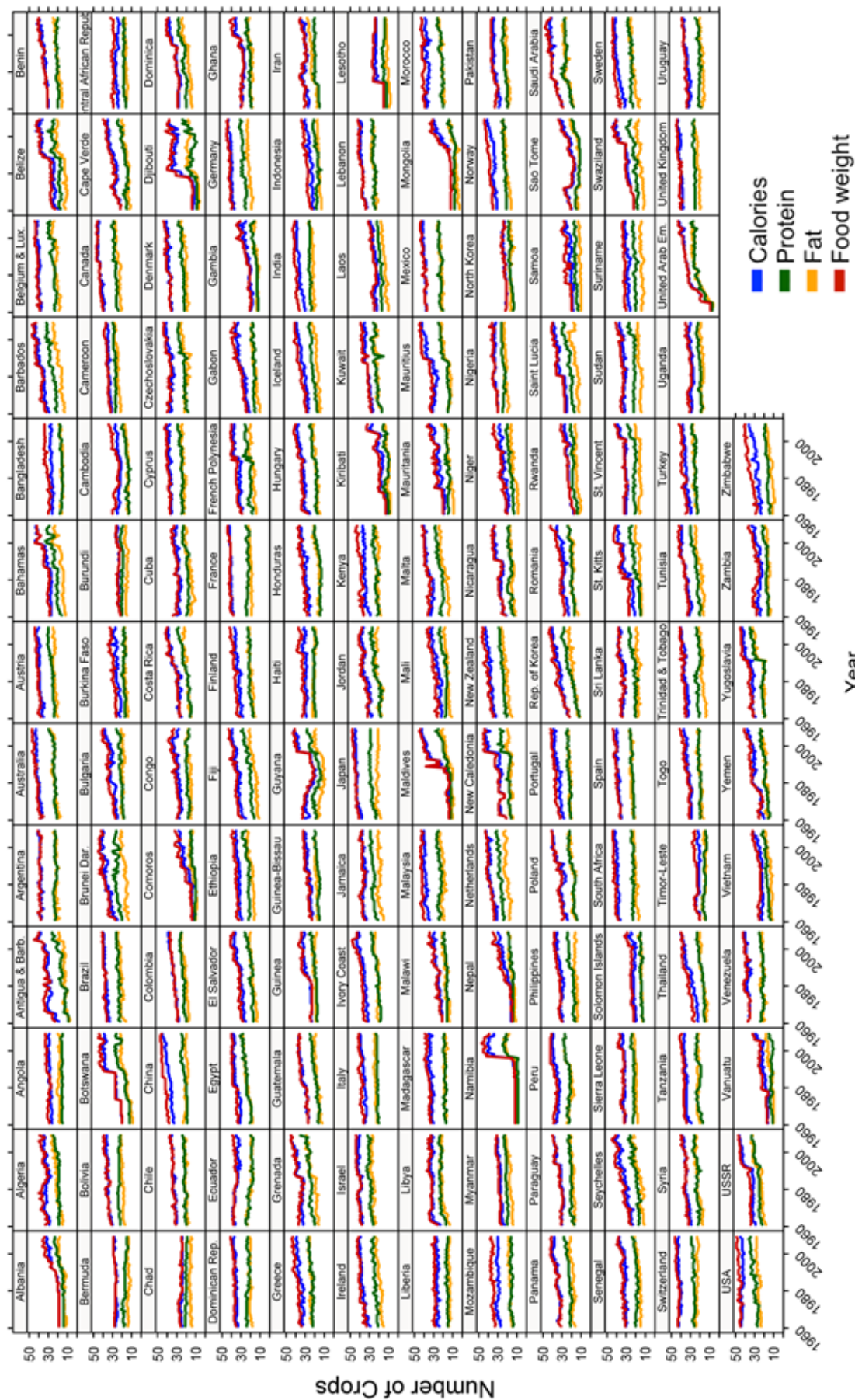
Supplementary Figure 3B. Change in the total number of countries reporting each crop commodity, per crop, for calories, protein, fat and weight.



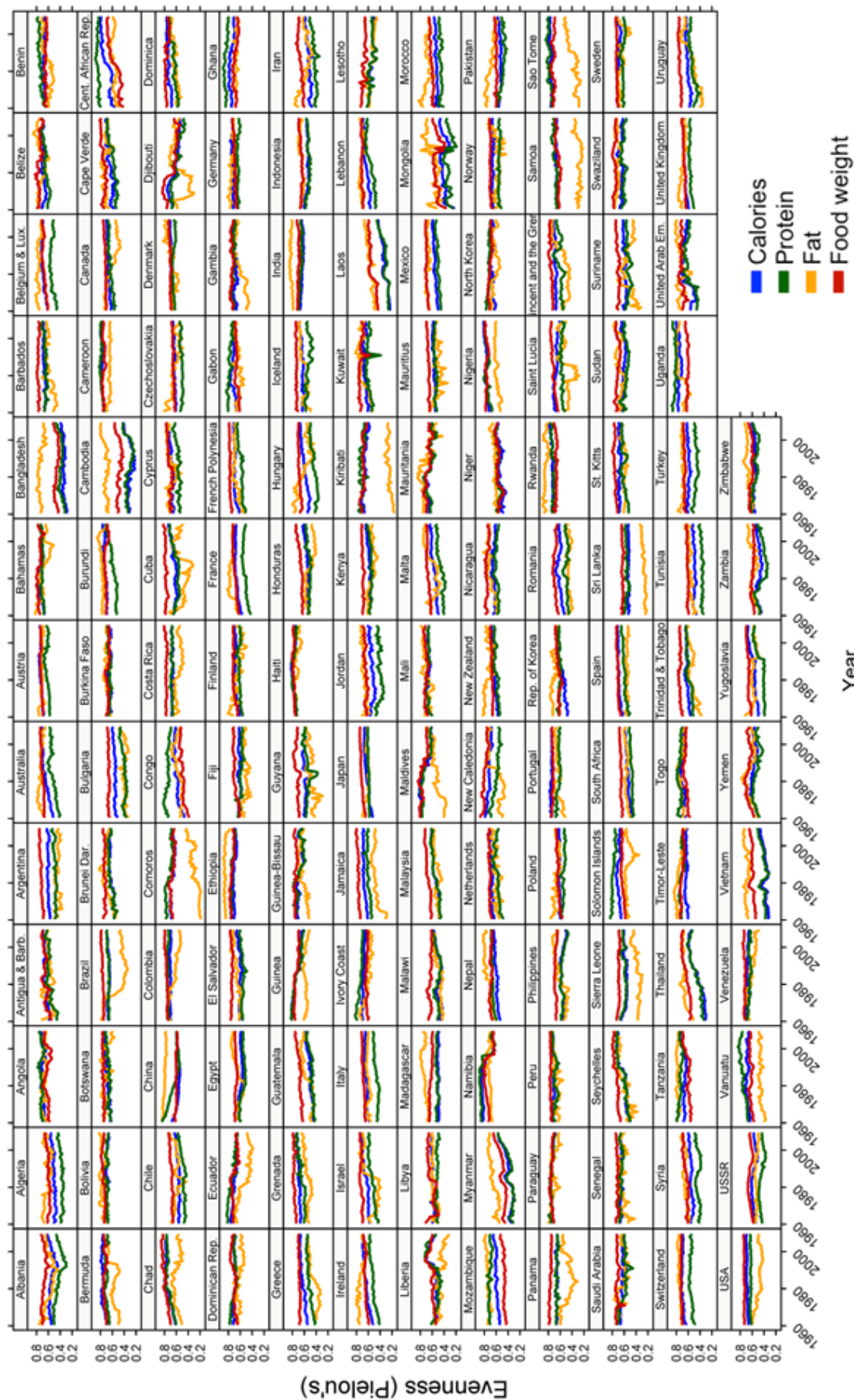
Supplementary Figure 4. Slope of the change in relative abundance probability of crop commodities in contribution to (A) calories, (B) protein, (C) fat, and (D) weight in national per capita food supplies from 1961-2009. Change in relative abundance probability was analyzed using a generalized linear mixed model with a binomial error distribution, “year” and “crop” as fixed effects and “country” as a random effect. Bars represent the slopes (\pm 95% confidence interval) of the predicted values derived from the model for each crop species. Commodities contributing minimally to protein/fat (i.e. <1 gram of protein/fat per 100 g of the item) were not included in (B) and (C) here.



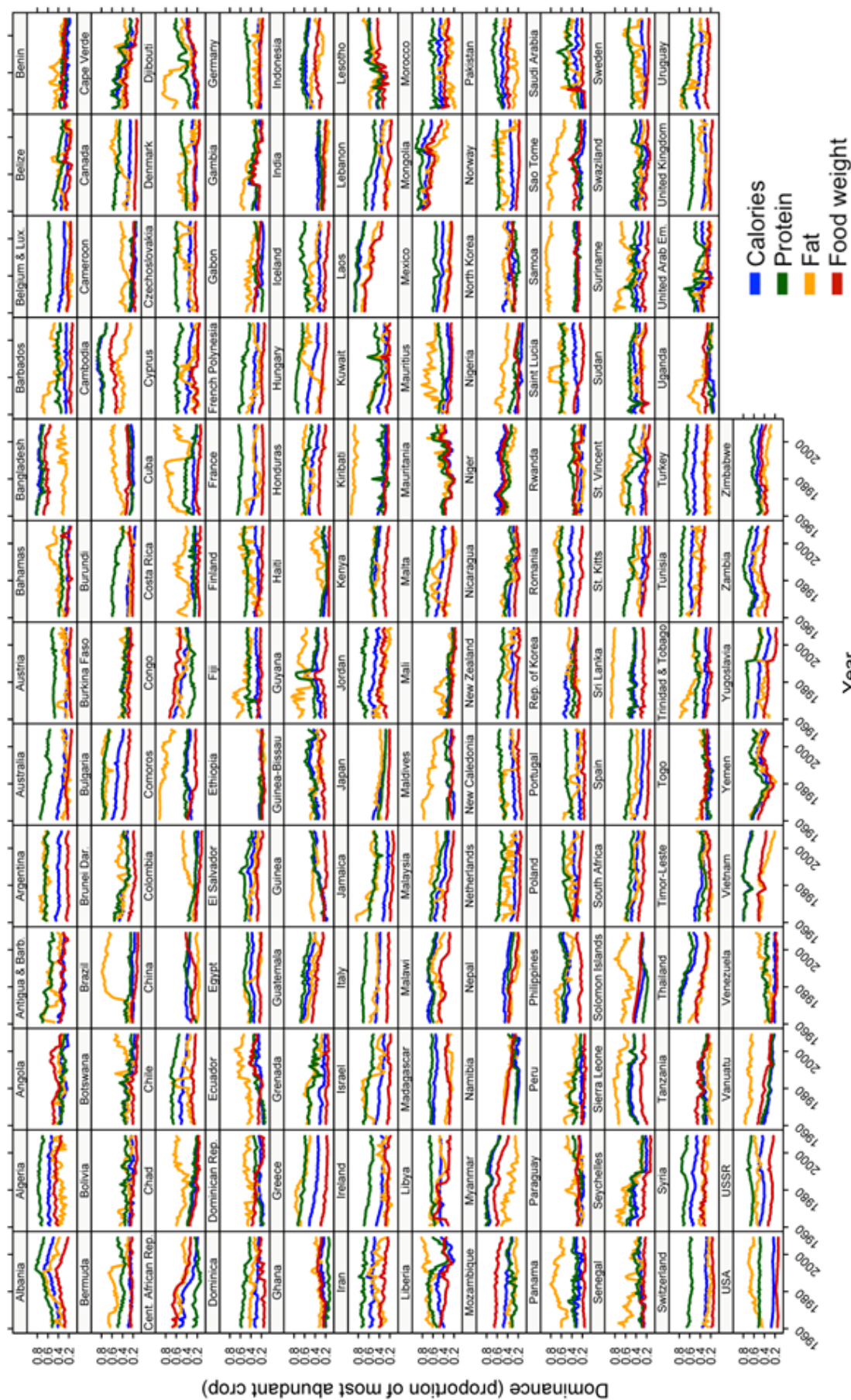
Supplementary Figure 5. Comparison of absolute abundance trends for crop commodities reported in contribution to (A) calories, (B) protein, (C) fat, and (D) weight in global aggregate food supply data versus mean national food supply data, 1961-2009. In order to derive these values, the absolute abundance of the crop in 1961 was subtracted from the 2009 value. For country-level changes, the between-years difference was first calculated for each country and then averaged across all countries (\pm 95% CI). Global aggregate abundance is a single global value and therefore no estimate of error is possible.



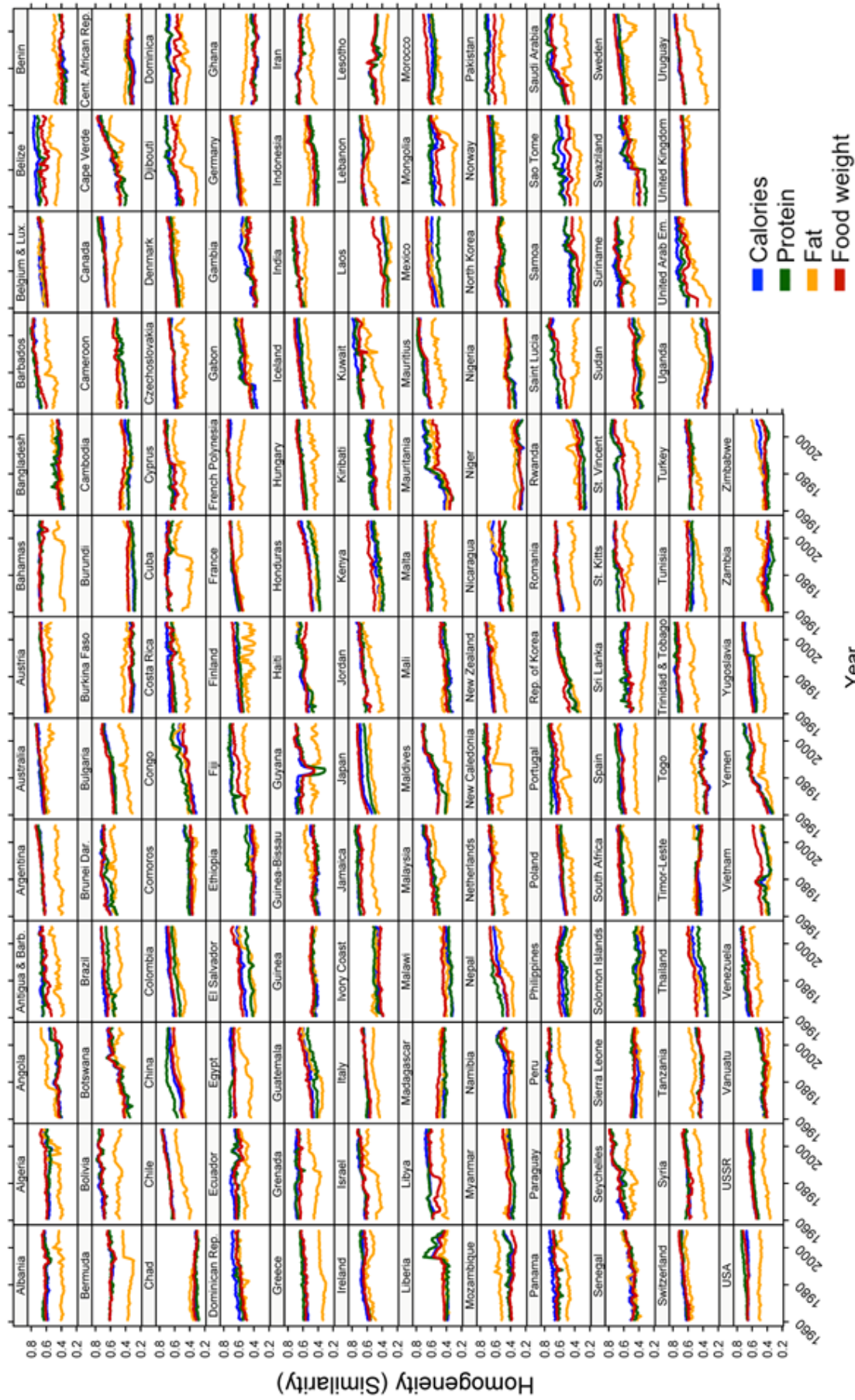
Supplementary Figure 6. Change in the total number of crop commodities reported in contribution to national food supplies, per country, for calories, protein, fat and weight, 1961-2009.



Supplementary Figure 7. Change in evenness in contributing crop commodities in national food supplies using Pielou's evenness index, per country, for calories, protein, fat and weight, 1961–2009.



Supplementary Figure 8. Change in the proportion of the most abundant crop commodity in national food supplies, per country, for calories, protein, fat and weight, 1961-2009.



Supplementary Figure 9. Per-country contribution to global homogenization, as measured by the change in similarity of the national food supply crop commodity composition in comparison to the global mean composition (centroid), for calories, protein, fat and weight, 1961-2009.

Supplementary Table 1. Relative importance of crop commodities worldwide, as measured by the number of countries in which each specific crop commodity counted within the top 90% of national per capita food supply for that variable (total countries = 152) in the mostcurrent year (2009).

Commodity	Countries for Calories	Countries for Protein	Countries for Fat	Countries for Food Weight	Comments
Wheat (<i>Triticum aestivum</i> L., <i>T. turgidum</i> L. (or <i>Triticum durum</i> Desf.))	148	146	84	144	We counted the major species of bread wheat (<i>Triticum aestivum</i> L.) and pasta wheat (<i>T. turgidum</i> L. or <i>T. durum</i> Desf.).
Rice (<i>Oryza sativa</i> L., <i>O. glaberrima</i> Steud.)	138	106	29	93	We counted both Asian rice (<i>Oryza sativa</i> L.) as well as African rice (<i>O. glaberrima</i> Steud.), as all West African countries in which this rice is consumed list the commodity as important, and due to the increasing adoption of Asian-African hybrid (NERICA) rices in the same region (Linares 2002). Commodities grouped included Rice (Milled Equivalent) and Ricebran Oil.
Maize (<i>Zea mays</i> L.)	116	92	71	83	Commodities grouped included Maize and Maize Germ Oil.
Barley (<i>Hordeum vulgare</i> L.)	69	16	0	103	Commodities grouped included Barley and Beer.
Sorghum (<i>Sorghum bicolor</i> (L.) Moench)	28	29	16	23	
Millet- Pearl millet (<i>Pennisetum glaucum</i> (L.) R. Br.), Finger millet (<i>Eleusine coracana</i> (L.) Gaertn., Tef (<i>Eragrostis tef</i> (Zucc.) Trotter))	15	13	8	14	Countries listing the commodity as important are from West, Central, and East Africa, as well as India and Nepal, therefore we counted both pearl millet (<i>Pennisetum glaucum</i> (L.) R. Br.) and finger millet (<i>Eleusine coracana</i> (L.) Gaertn.) (Rachie 1975). Tef (<i>Eragrostis tef</i> (Zucc.) Trotter) was also counted as the commodity is important in Ethiopia PDR (Zeid <i>et al.</i> 2011). Fonio (<i>Digitaria exilis</i> (Kippist) Stapf, is listed under the Cereals, other commodity (FAO 2013) and was therefore not counted here.
Cereals, other- Fonio (<i>Digitaria exilis</i> (Kippist) Stapf)	9	6	1	6	This commodity encompasses Popcorn, Buckwheat, Quinoa, Fonio, Triticale, Canary seed, and a variety of non-specific mixed grain and flour commodities (FAO 2013). We counted fonio as the commodity is significant in Guinea (Dendy & Dobraszyk 2001). The rest of the specific crops included in this general commodity are not likely to be important in listed countries. Instead, listed countries indicate that the general commodity is likely important due to the inclusion of mixed cereals and flours (probably wheat and other cereals).
Rye (<i>Secale cereale</i> L.)	9	7	0	2	
Oats (<i>Avena sativa</i> L.)	3	2	0	0	
Potatoes (<i>Solanum tuberosum</i> L.)	73	51	0	110	
Cassava (<i>Manihot esculenta</i> Crantz)	42	21	2	52	
Roots, other- Taro (<i>Colocasia esculenta</i> (L.) Schott), Yautia (<i>Xanthosoma sagittifolium</i> (L.) Schott)	17	16	0	27	We counted taro (<i>Colocasia esculenta</i> (L.) Schott) and yautia (<i>Xanthosoma sagittifolium</i> (L.) Schott), the specific crops listed under this general commodity, as the commodity is listed as important in the countries in which each is produced in the greatest quantity (e.g. Nigeria, Ghana, and Cameroon for taro; and Peru, Dominica, and Trinidad and Tobago for yautia) (FAO 2010a). Both of these crops were listed as important in Prescott-Allen & Prescott-Allen (1990).
Sweet potatoes (<i>Ipomoea batatas</i> (L.) Poir.)	15	6	2	24	
Yams (<i>Dioscorea rotundata</i> Poir., <i>D. cayenensis</i> Lam., <i>D. alata</i> L.)	14	12	2	15	The countries in which this commodity is important are in West Africa, the West Indies and the Pacific. We counted here the most common major species in these regions (International Institute for Tropical Agriculture 2009).

Commodity	Countries for Calories	Countries for Protein	Countries for Fat	Countries for Food Weight	Comments
Sugar- Sugar cane (<i>Saccharum officinarum</i> L.), Sugar beet (<i>Beta vulgaris</i> L.)	149	0	0	137	Commodities grouped included Sugar (Raw Equivalent), Sugar Beet, Sugar Cane, and Sugar, Non-Centrifugal. As the raw equivalent sugar commodity may derive from either species (FAO 2013), we counted both crops (<i>Saccharum officinarum</i> L. and <i>Beta vulgaris</i> L.).
Sweeteners, other	24	0	0	9	The commodity includes sugar maple (<i>Acer saccharum</i> Marshall) as well as a number of non-specific commodities (e.g. Fructose, Maltose, Sugar crops nes, Sugar nes, and Glucose and Dextrose) (FAO 2013). As both Canada and the United States list the commodity as important, it is possible that sugar maple should be considered, but the dominant presence of general processed sweetener categories cause sufficient uncertainty such that we did not count the species.
Pulses, Other; Cowpea (<i>Vigna unguiculata</i> (L.) Walp.), Chickpea (<i>Cicer arietinum</i> L.), Lentil (<i>Lens culinaris</i> Medik.), Faba bean (<i>Vicia faba</i> L.), Pigeonpea (<i>Cajanus cajan</i> (L.) Millsp.)	50	73	3	12	This general commodity encompasses Broad beans, Chickpeas, Cowpeas, Pigeonpeas, Lentils, Bambara beans, Vetches, and Lupins, along with a few non-specific commodities (FAO 2013). Given the importance of this commodity throughout the Mediterranean, West and South Asia and Africa, it is probable that at least Cowpeas (<i>Vigna unguiculata</i> (L.) Walp.), Chickpeas (<i>Cicer arietinum</i> L.), Broad beans (<i>Vicia faba</i> L.), Pigeonpeas (<i>Cajanus cajan</i> (L.) Millsp.), and Lentils (<i>Lens culinaris</i> Medik.) can be counted here. Prescott-Allen & Prescott-Allen (1990) listed each of these crops as important.
Beans (<i>Phaseolus vulgaris</i> L., <i>P. lunatus</i> L.)	35	49	3	16	We counted the two most cultivated bean species, Common bean (<i>Phaseolus vulgaris</i> L.) and Lima bean (<i>P. lunatus</i> L.) (Belay 2006).
Peas (<i>Pisum sativum</i> L.)	5	22	0	0	Commodities grouped included Soyabean Oil and Soyabeans.
Soybean (<i>Glycine max</i> (L.) Merr.)	87	24	113	16	Commodities grouped included Palm Oil, Palmkernel Oil, and Palmkernels. Given that the spread of countries in which the commodity is important includes both African and Neotropical countries, we listed both oil palm species (<i>Elaeis guineensis</i> Jacq., <i>E. oleifera</i> (Kunth) Cortés).
Palm oil (<i>Elaeis guineensis</i> Jacq., <i>E. oleifera</i> (Kunth) Cortés)	72	0	90	12	Commodities grouped included Groundnut Oil and Groundnuts (Shelled Eq).
Groundnut (<i>Arachis hypogaea</i> L.)	42	34	63	8	Commodities grouped included Coconut Oil and Coconuts - Incl Copra.
Coconuts (<i>Cocos nucifera</i> L.)	36	8	56	23	Commodities grouped included Sunflowerseed and Sunflowerseed Oil.
Sunflower (<i>Helianthus annuus</i> L.)	50	0	69	2	Commodities grouped included Oilcrops, Other and Oilcrops Oil, Other. The commodity encompasses Karite Nuts (Sheanuts), Castor oil, Tung Nut s, Jojoba, Safflower, Poppy, Melonseed, Tallowtree, Kapok, Linseed, Shillingia, Cocoa Butter and Hempseed, along with a number of non-specific oil commodities (FAO 2013). From the countries listed it was not feasible to identify particular crop species. We followed Prescott-Allen & Prescott-Allen (1990) in listing one crop- Shea nut (<i>Vitellaria paradoxa</i> C. F. Gaertn.), as it was listed as important in Ghana, Mali, and Nigeria in their data, and these countries also list the commodity as important in the current data. Melonseed, also listed by Prescott-Allen & Prescott-Allen (1990), was not counted here, but the species is counted in Fruits, Other.
Oilcrops, other; Shea nut (<i>Vitellaria paradoxa</i> C. F. Gaertn.)	29	2	74	1	

Commodity	Countries for Calories	Countries for Protein	Countries for Fat	Countries for Food Weight	Comments
Treenuts; Almond (<i>Prunus dulcis</i> (Mill.) D. A. Webb), Walnut (<i>Juglans regia</i> L.), Pistachio (<i>Pistacia vera</i> L.), Hazelnut (<i>Corylus avellana</i> L.), Brazil nut (<i>Bertholletia excelsa</i> Humb. & Bonpl.)	19	8	34	2	The commodity encompasses Brazil nuts, Cashew, Chestnuts, Almonds, Walnuts, Pistachios, Kolanuts, Hazelnuts, Arecanuts, Nuts nes, and Prepared Nuts (Exc. Groundnuts). Countries listing the commodity as important are spread worldwide, but concentrated in Europe, the Mediterranean, and Western Asia. We counted the two crops with the largest production worldwide- almond (<i>Prunus dulcis</i> (Mill.) D. A. Webb) and walnut (<i>Juglans regia</i> L.) (FAO 2010a). We also counted pistachio (<i>Pistacia vera</i> L.), hazelnut (<i>Corylus avellana</i> L.), and Brazil nut (<i>Bertholletia excelsa</i> Humb. & Bonpl.), which were also listed as important in Prescott-Allen & Prescott-Allen (1990). Taxonomy of almond is per GRIN Taxonomy (USDA, ARS, National Genetic Resources Program 2011).
Olives (<i>Olea europaea</i> L.)	22	0	29	5	Commodities grouped included Olive and Olive Oil.
Rape & mustard (<i>Brassica napus</i> L., <i>B. rapa</i> L.)	24	0	29	2	We counted two major oil species (<i>Brassica napus</i> L. and <i>B. rapa</i> L.), both common ingredients in canola oil. Other species of mustard may also be important in individual regions.
Sesame (<i>Sesamum indicum</i> L.)	9	6	20	1	Commodities grouped included Sesameseed and Sesameseed Oil.
Cottonseed oil (<i>Gossypium hirsutum</i> L., <i>G. herbaceum</i> L.)	7	0	23	0	We counted the common cotton species used for oil (<i>Gossypium hirsutum</i> L. and <i>G. herbaceum</i> L.), which may be mixed in oil products.
Vegetables, other- Cabbage and other brassicas (<i>Brassica oleracea</i> L.), Carrot (<i>Daucus carota</i> L.), Cucumber (<i>Cucumis sativus</i> L.), Pumpkin and squash (<i>Cucurbita maxima</i> Duchesne, <i>Cucurbita moschata</i> Duchesne, <i>Cucurbita pepo</i> L.), Lettuce (<i>Lactuca sativa</i> L.), Eggplant (<i>Solanum melongena</i> L.), Garlic (<i>Allium sativum</i> L.), Spinach (<i>Spinacia oleracea</i> L.), Artichoke (<i>Cynara cardunculus</i> L.)	75	90	1	149	The commodity encompasses Cabbages, Cauliflowers, Broccoli, and other brassicas, Artichokes, Asparagus, Lettuce and chicory, Spinach, Cassava leaves, Pumpkins, squash and gourds, Cucumbers and gherkins, Eggplants (aubergines), Chillies and peppers green, Onions (inc. shallots) green, Garlic, Leeks other alliaceous veg, Beans green, Peas green, String beans, Carrots and turnips, Okra, Sweet Corn, Mushrooms, Chicory roots, Carobs, Watermelons, Other melons (inc. cantaloupes), and Coffee Substitute, as well as a number of non-specific group commodities (FAO 2013). The commodity is extremely important worldwide, especially for weight, and it is likely that a large number of the crops contained within the general commodity individually contribute significantly in at least one country. We followed Prescott-Allen & Prescott-Allen (1990), counting 11 of the crop species now contained here. Chillies were included instead in Pimento; and watermelons instead in Fruits, Other.
Fruits, other- Watermelon (<i>Citrullus lanatus</i> (Thunb.) Matsum. & Nakai), Avocado (<i>Persea americana</i> Mill.), Mango (<i>Mangifera indica</i> L.), Melon (<i>Cucumis melo</i> L.), Papaya (<i>Carica papaya</i> L.), Pear (<i>Pyrus communis</i> L.), Peach (<i>Prunus persica</i> (L.) Batsh), Plum (<i>Prunus domestica</i> L.), Fig (<i>Ficus carica</i> L.), Strawberry (<i>Fragaria × ananassa</i> (Weston) Duchesne), Apricot (<i>Prunus armeniaca</i> L.), Cherry (<i>Prunus avium</i> (L.) L.), Currant (<i>Ribes nigrum</i> L. or <i>Ribes rubrum</i> L.)	38	3	7	127	The commodity encompasses Pears, Quinces, Apricots, Cherries, Peaches and nectarines, Plums and sloes, Strawberries, Raspberries, Gooseberries, Currants, Blueberries, Cranberries, Watermelons, Other melons (inc. cantaloupes), Figs, Mangoes, mangosteens, guavas, Avocados, Persimmons, Cashewapple, Kiwi fruit, and Papayas, as well as a number of non-specific group commodities (FAO 2013). The commodity is very important worldwide, especially for weight, and it is likely that a large number of the crops contained within the general commodity individually contribute significantly in at least one country. We followed Prescott-Allen & Prescott-Allen (1990), listing 13 of the crop species now contained here.
Bananas & plantains (<i>Musa acuminata</i> Colla, <i>Musa balbisiana</i> Colla)	45	20	3	78	Commodities grouped included Bananas and Plantains. As banana and plantain varieties are either of these species or a hybrid, we counted both species (<i>Musa acuminata</i> Colla and <i>Musa balbisiana</i> Colla) here.

Commodity	Countries for Calories	Countries for Protein	Countries for Fat	Countries for Food Weight	Comments
Tomatoes (<i>Solanum lycopersicum</i> Lam.)	2	11	0	77	The commodity includes Oranges, Orange juice single strength, Orange juice concentrated, Tangerines, mandarins, clem., and Tangerine Juice (FAO 2013). We counted only orange (<i>Citrus sinensis</i> (L.) Osbeck).
Oranges & Mandarines- Oranges (<i>Citrus sinensis</i> (L.) Osbeck)	4	2	0	82	
Grapes (<i>Vitis vinifera</i> L.)	31	0	0	48	
Apples (<i>Malus domestica</i> Borkh.)	1	0	0	54	Commodities grouped included Grapes and Wine. We counted only <i>Vitis vinifera</i> L., source of over 90% of grape and wine production worldwide (Winkler <i>et al.</i> 1974).
Onions (<i>Allium cepa</i> L.)	0	0	0	44	
Pineapples (<i>Ananas comosus</i> (L.) Merr.)	0	0	0	13	The commodity includes Onions dry. Green onions/shallots are included under the Vegetables, Other commodity (FAO 2013).
Dates (<i>Phoenix dactylifera</i> L.)	7	0	0	5	
Grapefruit (<i>Citrus paradisi</i> Macfad.)	2	1	0	8	
Lemons & limes (<i>Citrus limon</i> (L.) Burm. f., <i>C. aurantiifolia</i> (Christm.) Swingle)	0	0	0	8	The commodity includes Grapefruit (inc. pomebols), Juice of Grapefruit, and Grapefruit juice, concentrated (FAO 2013). We counted only grapefruit (<i>Citrus paradisi</i> Macfad.), as countries in which the commodity is important are generally from the West Indies. No South or Southeast Asian countries, where pomelo (<i>Citrus maxima</i> (Burm.) Merr.) is consumed, list the commodity as important.
Citrus, other	0	0	0	6	
Cocoa beans (<i>Theobroma cacao</i> L.)	11	1	19	0	We counted lemons (<i>Citrus limon</i> (L.) Burm. f.) and the major species of lime (<i>C. aurantiifolia</i> (Christm.) Swingle), a crop which also includes a number of minor species and hybrids.
Pimento-Chillies and peppers (<i>Capsicum annuum</i> L., <i>C. chinense</i> Jacq.), Pimento (<i>Pimenta dioica</i> (L.) Merr.)	2	1	4	0	
Coffee (<i>Coffea arabica</i> L., <i>C. canephora</i> Pierre ex A. Froehner)	0	4	0	2	The commodity includes Citrus fruit nes, Citrus juice single strength, and Citrus juice concentrated (FAO 2013). Given the spread of countries in which the commodity is important it is not possible to determine whether additional citrus species are represented here, or (more likely), the commodity is important due to juice of common citrus species.
Spices, other- Ginger (<i>Zingiber officinale</i> Roscoe), Nutmeg/mace (<i>Myristica fragrans</i> Houtt.), Cardamom (<i>Amomum subulatum</i> Roxb.)	2	1	2	0	

Commodity	Countries for Calories	Countries for Protein	Countries for Fat	Countries for Food Weight	Comments
Tea- Yerba Maté (<i>Ilex paraguariensis</i> A. St. -Hil.)	0	2	0	1	The commodity includes Tea (<i>Camellia sinensis</i> (L.) Kuntze) and Maté (<i>Ilex paraguariensis</i> A. St. -Hil.) (FAO 2013). As Paraguay and Uruguay, the main producers and consumers of Maté, are the two countries that list the commodity as important, it is very likely that they refer to Maté.
Beverages, alcoholic	35	0	0	10	The commodity is composed of Bever. Dist. Alc (FAO 2013). As it is not possible to determine crops, no species are listed.
Beverages, fermented	10	3	0	22	The commodity includes Wheat Fermented Beverage, Rice Fermented Beverages, Beer of Maize, Beer of Millet, Beer of Sorghum, and Cider Etc (FAO 2013). These species are represented in crop commodities above, and no species are counted here.
Miscellaneous	13	11	0	0	This plant general commodity is not further defined and therefore it is not possible to determine specific crops.
Animal Commodities total	152	152	152	152	The commodity includes all animal food commodities listed in FAOSTAT Food Balance Sheets, a sum of all commodities in Meat, Offals; Animal Fats; Eggs; Milk, excluding butter; Fish, Seafood; and Aquatic Products, other categories (FAO 2013).

For any particular variable (e.g. calories) for each country, the contributing crop commodities were listed in descending order of importance until 90% of food supply was reached. These commodities were then counted as significant to the food supply of the country, and the total number of countries finding each crop significant was summed to derive a quantitative measure of importance worldwide. Commodities are listed in order of importance defined as sum of country count across the four variables, within cereals, starchy roots, sugarcrops, pulses, oilcrops and nuts, fruits and vegetables, spices and stimulants, and alcoholic beverages categories. Taxonomy followed The Plant List (2010).

REFERENCES

- Adler, C.J., Dobney, K., Weyrich, L.S., Kaidonis, J., Walker, A.W., Haak, W., *et al.* (2013). Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial Revolutions. *Nat. Genetics* 45, 450-455.
- Alston, J.M., Beddow, J.M., Pardey, P.G. (2009). Agricultural research, productivity, and food prices in the long run. *Science* 325 (5945), 1209-1210.
- Belay, G. (2006). *Cereals and Pulses* (Wageningen: PROTA Foundation).
- Bonin, C.L., Tracy, B.F. (2012). Diversity influences forage yield and stability in perennial prairie plant mixtures. *Agr. Ecosyst. Environ.* 162, 1-7.
- Bonneuil, C., Goffaux, R., Bonnin, I., Montalent, P., Hamon, C., Balfourier, F., *et al.* (2012). A new integrative indicator to assess crop genetic diversity. *Ecol. Indic.* 23, 280-289.
- Burke, M.B., Lobell, D.B., Guarino, L. (2009). Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environ. Chang.* 19 (3), 317-325.
- Cabell, J.F., Oelofse, M. (2012). An indicator framework for assessing agroecosystem resilience. *Ecol. Soc.* 17 (1), 18.
- Cadotte, M.W., Dinnage, R., Tilman, D. (2012). Phylogenetic diversity promotes ecosystem stability. *Ecology* 93 (sp8), S223-S233.
- Cordell, D., Drangert, J.-O., White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Chang.* 19 (2), 292-305.
- Davis, A.P., Gole, T.W., Baena, S., Moat, J. (2012). The impact of climate change on indigenous arabica coffee (*Coffea arabica*): predicting future trends and identifying priorities. *PLoS One* 7 (11), e47981.
- Davis, A.S., Hill, J.D., Chase, C.A., Johanns, A.M., Liebman, M. (2012). Increasing cropping system diversity balances productivity, profitability and environmental health. *PLoS One* 7(10), e47149.
- Davis, D.R., Epp, M.D., Riordan, H.D. (2004). Changes in USDA food composition data for 43 garden crops, 1950 to 1999. *J. Am. Coll. Nutr.* 23 (6), 669-682.
- Dendy, D.A.V., Dobraszczyk, B.J. (2001). *Cereals and Cereal Products: Chemistry and Technology* (Gaithersburg: Aspen Publishers, Inc.).
- Doughty, J. (1979). Dangers of reducing the range of food choice in developing countries. *Ecol. Food Nutr.* 8, 275-283.
- Dwivedi, S., Sahrawat, K., Upadhyaya, H., Ortiz, R. (2013). Food, nutrition and agrobiodiversity under global climate change. *Adv. Agron.* 120, 1-128.
- Fader, M., Gerten, D., Krause, M., Lucht, W., Cramer, W. (2013). Spatial decoupling of agricultural production and consumption: quantifying dependences of countries on food imports due to domestic land and water constraints. *Environ. Res. Lett.* 8 (1), 014046.
- FAO (Food and Agriculture Organization of the United Nations). (1984). *Food Balance Sheets 1979-1981 Average* (Rome: Food and Agriculture Organization of the United Nations).
- FAO (Food and Agriculture Organization of the United Nations). (2001). *Food Balance Sheets: A handbook* (Rome: Food and Agriculture Organization of the United Nations).
- FAO (Food and Agriculture Organization of the United Nations). (2002). *The International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: Food and Agriculture Organization of the United Nations).
- FAO (Food and Agriculture Organization of the United Nations). (2010a). FAOSTAT, Production. <http://faostat3.fao.org>.
- FAO (Food and Agriculture Organization of the United Nations). (2010b). *Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: Food and Agriculture

- Organization of the United Nations).
- FAO (Food and Agriculture Organization of the United Nations). (2013). FAOSTAT, Methods and standards, classification of crops primary equivalent. <http://faostat3.fao.org/mes/classifications/E>.
- Foley, J.A., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., *et al.* (2005). Global consequences of land use. *Science* 309 (5734), 570-574.
- Fowler, C., Hodgkin, T. (2004). Plant genetic resources for food and agriculture: assessing global availability. *Annu. Rev. Env. Resour.* 29 (1), 143-179.
- Frei, M., Becker, K. (2004). Agro-biodiversity in subsistence-oriented farming systems in a Philippine upland region: nutritional considerations. *Biodivers. Conserv.* 13 (8), 1591-1610.
- Gamfeldt, L., Snäll, T., Bagchi, R., Jonsson, M., Gustafsson, L., Kjellander, P., *et al.* (2013). Higher levels of multiple ecosystem services are found in forests with more tree species. *Nat. Commun.* 4, 1340.
- Gepts, P. (2006). Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci.* 46 (5), 2278-2292.
- Graham, R.D., Welsh, R.M., Saunders, D.A., Ortiz-Monasterio, I., Bouis, H.E., Bonierbale, M., *et al.* (2007). Nutritious subsistence food systems. *Adv. Agron.* 92, 1-74.
- Hajjar, R., Jarvis, D.I., Gemmill-Herren, B. (2008). The utility of crop genetic diversity in maintaining ecosystem services. *Agr. Ecosyst. Environ.* 123 (4), 261-270.
- Hammer, K., Khoshbakht, K. (2005). Towards a “red list” for crop plant species. *Genet. Resour. Crop Ev.* 52 (3), 249-265.
- Harlan, J.R. (1975). *Crops and Man* (Madison: American Society of Agronomy).
- Hawkesworth, S., Dangour, A.D., Johnston, D., Lock, K., Poole, N., Rushton, J., *et al.* (2010). Feeding the world healthily: the challenge of measuring the effects of agriculture on health. *Philos. T. R. Soc. B.* 365 (1554), 3083–3097.
- Hooper, D.U., Adair, E.C., Cardinale, B.J., Byres, J.E.K., Hungate, B.A., Matulich, K.L., *et al.* (2012). A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* 486, 105-108.
- Hooper, D.U., Chapin III, F.S., Ewel, J.J., Hector, A., Inchausti, P., Lavore, S., *et al.* (2005). Effects of biodiversity on ecosystem functioning: a consensus of current knowledge. *Ecol. Monog.* 75, 3-35.
- International Institute for Tropical Agriculture. (2009). Yam. <http://www.iita.org/yam>.
- Jarvis, A., Ramirez-Villegas, J., Herrera-Campo, B.V., Navarro-Racines, C. (2012). Is cassava the answer to African climate change adaptation? *Trop. Plant Biol.* 5 (1), 9-29.
- Kastner, T., Rivas, M.J.I., Koch, W., Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. *Proc. Natl. Acad. Sci.* 109 (18), 6868-6872.
- Kearney, J. (2010). Food consumption trends and drivers. *Philos. T. R. Soc. B.* 365 (1554), 2793-2807.
- Kennedy, G., Islam, O., Eyzaguirre, P., Kennedy, S. (2005). Field testing of plant genetic diversity indicators for nutrition surveys: rice-based diet of rural Bangladesh as a model. *J. Food Compos. Anal.* 18 (4), 255-268.
- Kirwan, L., Luscher, A., Sebastia, M.T., Finn, J.A., Collins, R.P., Porceddu, C., *et al.* (2007). Evenness drives consistent diversity effects in intensive grassland systems across 28 European sites. *J. Ecol.* 95 (3), 530-539.
- Kremen, C., Miles, A. (2012). Ecosystem services in biologically diversified versus conventional farming systems: benefits, externalities, and trade-offs. *Ecol. Soc.* 17 (4), 40.
- Linares, O.F. (2002). African rice (*Oryza glaberrima*): history and future potential. *Proc. Natl. Acad. Sci.* 99 (25), 16360-16365.
- Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* 319 (5863), 607-610.

- Lozupone, C.A., Stombaugh, J.I., Gordon, J.I., Jansson, J.K., Knight, R. (2012). Diversity, stability and resilience of the human gut microbiota. *Nature* 489 (7415), 220-230.
- McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., *et al.* (2013). Agriculture: feeding the future. *Nature* 499(7456), 23-24.
- McCouch, S.R., McNally, K.L., Wang, W., Sackville Hamilton, R. (2012). Genomics of gene banks: A case study in rice. *Am. J. Bot.* 99 (2), 407-423.
- Mijatović, D., van Oudenhoven, F., Eyzaguirre, P., Hodgkin, T. (2013). The role of agricultural biodiversity in strengthening resilience to climate change: Towards an analytical framework. *Int. J. Agric. Sustain.* 11 (2), 95-107.
- National Research Council, (1972). *Genetic Vulnerability of Crops* (Washington D.C.: National Academy of Sciences).
- Naylor, R.L., Falcon, W.P., Goodman, R.M., Jahn, M.M., Sengooba, T., Tefera, H., *et al.* (2004). Biotechnology in the developing world: a case for increased investments in orphan crops. *Food Pol.* 29 (1), 15-44.
- Negin, J., Remans, R., Karuti, S., Fanzo, J.C. (2009). Integrating a broader notion of food security and gender empowerment into the African green revolution. *Food Sec.* 1 (3), 351-360.
- Padulosi, S., Dulloo, E. (2012). "Towards a viable system for monitoring agrobiodiversity on-farm: a proposed new approach for Red Listing of cultivated plant species," in: *On-farm Conservation of Neglected and Underutilized Species: Status, Trends and Novel Approaches to Cope with Climate Change*, 14–16 June, 2011, eds. S. Padulosi, N. Bergamini, T. Lawrence T. (Friedrichsdorf, Frankfurt: Proceedings of the International Conference), 171-199.
- Perrier, X., De Langhe, E., Donohue, M., Lentfer, C., Vrydaghs, L., Bakry, F., *et al.* (2011). Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proc. Natl. Acad. Sci.* 108 (28), 11311-11318.
- Picasso, V.D., Brummer, E.C., Liebman, M., Dixon, P.M., Wilsey, B.J. (2008). Crop species diversity affects productivity and weed suppression in perennial polycultures under two management strategies. *Crop Sci.* 48 (1), 331-342.
- Pingali, P. (2007). Westernization of Asian diets and the transformation of food systems: implications for research and policy. *Food Pol.* 32 (3), 281-298.
- Popkin, B.M. (2006). Technology, transport, globalization and the nutrition transition food policy. *Food Pol.* 31 (6), 554-569.
- Pradhan, P., Reusser, D.E., Kropp, J.P. (2013). Embodied greenhouse gas emissions in diets. *PLoS One* 8 (5), e62228.
- Prescott-Allen, R., Prescott-Allen, C. (1990). How many plants feed the world? *Conserv. Biol.* 4 (4), 365-374.
- Rachie, K.O. (1975). *The Millets: Importance, Utilization and Outlook. Research Report.* (Hyderabad: International Crops Research Institute for the Semi-Arid Tropics).
- Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8 (6), e66428.
- Remans, R., Flynn, D.F.B., DeClerk, G., Diru, W., Fanzo, J., Gaynor, K., *et al.* (2011). Assessing nutritional diversity of cropping systems in African villages. *PLoS One* 6 (6), e21235.
- Ruel, M.T. (2003). Operationalizing dietary diversity: a review of measurement issues and research priorities. *J. Nutr.* 133 (11), 3911S-3926S.
- Stehfest, E., Bouwman, L., van Vuuren, D.P., den Elzen, M.G.J., Eickhout, B., Kabat, P. (2009). Climate benefits of changing diet. *Climatic Change* 95 (1-2), 83-102.
- The Plant List. (2010). Version 1. <http://www.theplantlist.org/>.
- USDA, ARS, National Genetic Resources Program. (2011). Germplasm Resources Information

- Network (GRIN). <http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl>.
- van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., Visser, B. (2009). Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources* 8, 1-15.
- van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* 120 (6), 1241-1252.
- Wheeler, T., von Braun, J. (2013). Climate change impacts on global food security. *Science* 341 (6145), 508-513.
- Winkler, A.J., Cook, J.A., Kliewer, W.M., Lider, L.A. (1974). *General Viticulture* (Berkeley: University of California Press).
- Zeid, M., Echenique, V., Diaz, M., Pessino, S., Sorrells, M. (2011). "Chapter 8: Eragrostis," in *Wild Crop Relatives: Genomic and Breeding Resources, Millets and Grasses*, ed. C. Kole (Berlin Heidelberg: Springer-Verlag).
- Zhang, Y., Chen, H.Y.H., Reich, P.B. (2012). Forest productivity increases with evenness, species richness and trait variation: a global meta-analysis. *J. Ecol.* 100 (3), 742-749.
- Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., *et al.* (2000). Genetic diversity and disease control in rice. *Nature* 406 (6797), 718-722.

CHAPTER 3

An inventory of crop wild relatives of the United States

Colin K. Khoury^{1,2}, Stephanie Greene³, John Wiersema⁴, Nigel Maxted⁵, Andy Jarvis^{1,6}, and
Paul C. Struik²

¹ International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia.

² Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

³ United States Department of Agriculture, Agricultural Research Service, Plant Germplasm Introduction and Testing, 24106 North Bunn Rd., Prosser, Washington, 99350, USA.

⁴ United States Department of Agriculture, Agricultural Research Service, National Germplasm Research Laboratory, Building 003, BARC-West, Beltsville, Maryland 20705-2350, USA.

⁵ School of Biosciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

⁶ CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia.

Published in *Crop Science* 53 (4), 1496. doi: 10.2135/cropsci2012.10.0585

Abstract

The use of crop wild relatives (CWR) in breeding is likely to continue to intensify as utilization techniques improve and crop adaptation to climate change becomes more pressing. Significant gaps remain in the conservation of these genetic resources. As a first step toward a national strategy for the conservation of CWR, we present an inventory of taxa occurring in the United States, with suggested prioritization of species based upon potential value in crop improvement. We listed 4,600 taxa from 985 genera and 194 plant families, including CWR of potential value via breeding as well as wild species of direct use for food, forage, medicine, herb, ornamental, and/or environmental restoration purposes. US CWR are related to a broad range of important food, forage and feed, medicinal, ornamental, and industrial crops. Some potentially valuable species are threatened in the wild, including relatives of sunflower (*Helianthus annuus* L.), walnut (*Juglans regia* L.), pepo squash (*Cucurbita pepo* L.), wild rice (*Zizania* L.), raspberry (*Rubus idaeus* L.), and plum (*Prunus salicina* Lindl.), and few accessions of such taxa are currently conserved *ex situ*. We prioritize 821 taxa from 69 genera primarily related to major food crops, particularly the approximately 285 native taxa from 30 genera that are most closely related to such crops. Both the urgent collection for *ex situ* conservation and the management of such taxa in protected areas are warranted, necessitating partnerships between concerned organizations, aligned with regional and global initiatives to conserve and provide access to CWR diversity.

Keywords: Crop wild relatives, Plant genetic resources, Wild utilized species, National inventory

INTRODUCTION

Nearly 40 years ago Jack Harlan outlined the major factors explaining the extent of use of wild relatives of crops (CWR) in plant breeding. His list included the degree of domestication of the crop, the perceived genetic vulnerability of the crop, the availability of CWR for utilization, the degree of difficulty in using CWR in breeding, and the economic conditions and disposition of breeders toward their use (Harlan 1976).

Utilization of CWR has steadily increased over the past decades, providing improved pest and disease resistance, tolerance to abiotic stresses, increased yield, novel cytoplasm, and quality traits to banana (*Musa acuminata* Colla), barley (*Hordeum vulgare* L.), bean (*Phaseolus vulgaris* L.), cassava (*Manihot esculenta* Crantz), chickpea (*Cicer arietinum* L.), corn (*Zea mays* L.), lettuce (*Lactuca sativa* L.), oat (*Avena sativa* L.), millet [*Pennisetum glaucum* (L.) R. Br.], potato (*Solanum tuberosum* L.), rice (*Oryza sativa* L.), sugarcane (*Saccharum officinarum* L.), sunflower, tomato (*Solanum lycopersicum* L.), and wheat (*Triticum aestivum* L.), among others (Iltis 1988, Xiao *et al.* 1996, Gur & Zamir 2004, Hajjar & Hodgkin 2007, McCouch *et al.* 2007, Maxted *et al.* 2012a). Advancements in breeding, particularly through novel molecular approaches, have increased the efficiency of the use of wild germplasm substantially (Tanksley & McCouch 1997, Zamir 2001, Ford-Lloyd *et al.* 2011, Volk & Richards 2011). Research for adaptation to future climates is likely to increase the exploitation of the variation represented in CWR (Ortiz *et al.* 2008, Guarino & Lobell 2011).

Despite substantial efforts over these years, the most significant bottleneck in utilization identified by Harlan - the availability of CWR for research and breeding - continues to be of concern, with substantial genetic variation yet to be conserved, and financial and political constraints still to be resolved (Fowler & Hodgkin 2004, FAO 2010).

An estimated one out of five plant species is threatened worldwide by habitat loss or modification, agricultural modernization, pollution, over-exploitation, invasive species, and/or climate change (Brummitt & Bachman 2010), and nearly 30% of the native flora of the United States is of conservation concern (CPC 2012). CWR are not exempt from these pressures (Wilkes 2007, Jarvis *et al.* 2008, Bilz *et al.* 2011, Ureta *et al.* 2011).

The urgent collection and subsequent storage in *ex situ* facilities, where these genetic resources

can be made available for research and breeding is thus warranted. The complementary protection of CWR *in situ* is necessary to support the ongoing evolution of CWR populations (Maxted *et al.* 1997, Meilleur & Hodgkin 2004, Heywood 2008, Maxted & Kell 2009) in both wild areas and traditional agricultural systems (Rawal 1975, GSPC 2002, Zizumbo-Villarreal *et al.* 2005).

It is becoming increasingly feasible to formulate comprehensive strategies for the conservation of CWR diversity due to advancements in understanding the taxonomic relationships between crops and their wild relatives (Andersson & de Vicente 2010, Wiersema *et al.* 2012), improved availability of data on the distribution of these taxa (e.g. FNA 2008a, GBIF 2012) and increased power of distribution modelling and conservation analysis (Hijmans & Spooner 2001, Jarvis *et al.* 2005, Ramírez-Villegas *et al.* 2010, Parra-Quijano *et al.* 2011).

The starting point for CWR conservation planning typically involves the creation of a checklist of included CWR taxon names, to which ancillary data (e.g. eco-geographic information, conservation status, utilization potential, etc.) is added to generate an inventory for the target area of research (Maxted *et al.* 2008). National inventories of CWR have been published for a growing list of nations, particularly in Europe (Magos Brehm *et al.* 2007, Maxted *et al.* 2012b), and targeted subsequent conservation efforts have been made in over 40 countries worldwide (Meilleur & Hodgkin 2004). On the global level, a specialist group is active in listing CWR of conservation concern (International Union for Conservation of Nature, Species Survival Commission 2008), an initiative is underway to document, collect, conserve *ex situ*, and pre-breed the CWR of major food and forage crops (Guarino and Lobell 2011), and progress is being made in planning for an integrated system of genetic reserves for the CWR of highest priority worldwide (Maxted & Kell 2009).

As a primary step in the process toward a national strategy for the conservation of CWR, we report on an inventory of the CWR flora occurring in the US, and a prioritization of these taxa based upon their potential value in agricultural crop research.

The United States Context

More than 20,000 species of plants, or about 7% of the world's flora, are native or naturalized in North America north of Mexico (FNA 2008b), but the region has not been considered a major

center of crop plant diversity (Vavilov 1926). Those indigenous species that were domesticated before European contact in Eastern North America include pepo squash (*Cucurbita pepo* L.), sunflower (*Helianthus annuus* L.), marsh-elder (*Iva annua* L.), and chenopod (*Chenopodium berlandieri* Moq.) (Smith 2006), and in the arid southwestern US include Sonoran panic (*Panicum hirticaule* J. Presl var. *hirticaule*) (Nabhan 1985) and devil's-claw (*Proboscidea parviflora* (Wooton) Wooton & Standl. subsp. *parviflora*) (Bretting & Nabhan 1986). Blueberry (*Vaccinium* section *Cyanococcus*) and cranberry (*Vaccinium* section *Oxycoccus*) (Ballington 2001), blackberry (*Rubus fruticosus* L., *sensu lato* and hybrids) (Finn, 2001), and pecan [*Carya illinoensis* (Wangenh.) K. Koch] (Flack 1970) may be counted as more recent domestications. The number of CWR native to the US may thus at first glance be estimated to be fairly small.

Three factors increase the number of potentially valuable CWR significantly. The relatives of a complex of Mesoamerican crop species, including corn, a number of bean (*Phaseolus* L.) and squash species, chili pepper (*Capsicum* L.), American cotton (*Gossypium hirsutum* L.), and tobacco (*Nicotiana rustica* L.) (Nabhan 1985, Zizumbo-Villarreal & Colunga-GarcíaMarín 2010) are distributed in the southern regions of the US (USA-ARS National Genetic Resources Program 2012). Second, a number of crops domesticated in other temperate regions of the world are congeneric with species occurring in the US, for example strawberry (*Fragaria* L.) (Hummer *et al.* 2011) and hops (*Humulus* L.) (Peredo *et al.* 2010). Finally, approximately 4,000 plant species have been introduced to the region since the Colombian Exchange (Guo *et al.* 2009), including weedy relatives of crop plants.

Several well-documented examples of use of native CWR in breeding exist. North American wild grape (*Vitis* L.) germplasm proved critical in providing resistance to phylloxera (*Phylloxera vitifoliae* Fitch) as a rootstock in European grape (*Vitis vinifera* L.) production in the late 1800s, and these stocks continue to provide the basis for protection worldwide (Gale 2003). Genes for resistance to a range of diseases and pests, including rust (*Puccinia helianthi* Schwein.), downy mildew [*Plasmopara halstedii* (Farl.) Berl. & De Toni], powdery mildew [*Golovinomyces cichoracearum* (DC.) V.P. Heluta], broomrape (*Orobanche cumana* Wallr.), sclerotinia head and stalk rot [*Sclerotinia sclerotiorum* (Lib.) de Bary], and sunflower moth (*Homoeosoma electellum* Hulst), have been identified in native sunflowers and successfully transferred into cultivars (Seiler & Gulya 2004).

Several US government entities support activities focused on CWR conservation. The US Department of Agriculture, Agricultural Research Service (ARS), National Plant Germplasm System (NPGS) published *in situ* conservation guidelines for US CWR (Plant Germplasm Operations Committee 1999), and recently formed a Subcommittee on CWR within the Plant Germplasm Operations Committee (Plant Germplasm Operations Committee 2010). The NPGS Germplasm Resources Information Network (GRIN) Taxonomy Section is preparing a database of crop gene pools listing CWR based on an evaluation of breeding and crossability studies (Wiersema *et al.* 2012).

Over the past decade the NPGS has supported 61 explorations for the US CWR of food, forage, woody landscape and ornamental crop plants (K. Williams, personal communication 2012). Genetic reserves for the wild relatives of grape (Pavek *et al.* 2001), chili pepper (Nabhan 1990), and cranberry (K. Hummer, personal communication 2010) have been established. Explorations regarding possible locations and feasibility of protected areas for CWR of pecan, potato, sweet pea (*Lathyrus* L.), and edible alliums (*Allium* L.) were also completed (Plant Germplasm Operations Committee 1999).

The US Forest Service (USFS) actively maintains a number of CWR populations in National Forests (USFS 2010) and CWR are informally conserved across the US on these and other public lands. The Bureau of Land Management, in partnership with the Millennium Seed Bank of the Royal Botanic Gardens, Kew and local organizations across the country, has collected CWR accessions within the “Seeds of Success” Program (Bureau of Land Management 2012). Building upon a partnership between ARS and USFS to collaborate on the establishment of *in situ* reserves for US CWR, the agencies are in the process of developing a coordinated strategy for CWR management (L. Stritch, personal communication 2012).

MATERIALS AND METHODS

National Inventory of Crop Wild Relatives

In order to broadly cover the wild plant species occurring in the US that have potential value in crop research, we compiled an Inventory of CWR that may be utilized in crop breeding as well as wild utilized species (WUS) directly used for food, forage, medicine, herb, ornamental, and/or environmental restoration purposes. Very few WUS are the central focus of plant breeding programs, although some taxa may be semi-domesticated, and WUS may have a high

potential for crop development.

For listed CWR we aimed to include the full range of taxa with the potential to contribute to crop improvement, including both those species where gene exchange with the crop is relatively straightforward, and more distant relatives requiring advanced techniques in order to produce viable hybrid progeny. Our starting point for defining CWR followed Maxted *et al.* (2006, p. 2680): “A crop wild relative is a wild plant taxon that has an indirect use derived from its relatively close genetic relationship to a crop; this relationship is defined in terms of the CWR belonging to Gene Pools (GPs) 1 or 2, or taxon groups (TGs) 1 to 4 of the crop.” The definition classifies taxa based on whether they occur in Harlan and de Wet’s (1971) crop GP 1 or 2, encompassing closely related taxa that are relatively easy to cross using conventional methods in breeding programs. If data from interspecific hybridization or genetic relatedness studies are unavailable, Maxted *et al.* (2006) proposed a classification system based on taxonomic groups equating to rank in relation to the crop species. Data on gene pool and taxon group concepts for available crops was gathered from GRIN taxonomy (Wiersema *et al.* 2012) and from the ‘Harlan and de Wet Crop Wild Relative Inventory’ (Vincent *et al.* 2013).

A growing number of crops have benefitted from traits introgressed from distant gene pools (Ballington 2001, Frese *et al.* 2001, Bradshaw *et al.* 2006, Mallikarjuna *et al.* 2006, Abberton 2007, Rygulla *et al.* 2007, Chuda & Adamus 2009, Mii 2009), and as breeding techniques improve, taxa from such gene pools are increasingly likely to be of interest to crop improvement programs. Such species are additionally useful for taxonomic and evolutionary research. We therefore broadened our CWR definition to include species in the tertiary gene pool. In some crops these may include taxa from related genera [e.g., *Tripsacum* L. for maize, *Aegilops* L. and *Amblyopyrum* Eig for wheat (Wiersema *et al.* 2012)].

CWR and WUS taxa occurring in the US were compiled from the GRIN *World Economic Plants* database, based on Wiersema and León (1999), completed volumes of the *Flora of North America* (FNA 2008a), McGuffin (2000), and the Native Seed Network Database (Native Seed Network 2010). Both native and introduced taxa were included. In addition to listing taxa to the infraspecific level, information on origin status, number of accessions (available plus unavailable) in GRIN, noxious weed status, associated crop, crop gene pool and associated crop use was obtained. Taxa with multiple uses were listed first by their primary use and thereafter

by subsequent uses. For taxa with uses both as CWR and WUS, utilization as a genetic resource was prioritized over direct uses. Additional data on occurrence and weed status was gathered from the PLANTS Database (USDA NRCS 2010). The threat status of taxa was recorded from NatureServe (NatureServe 2009) and the IUCN Red List of Threatened Species (IUCN 2012).

Taxonomic verification was performed via the ‘Taxonomic Name Resolution Service’ (Boyle *et al.* 2013) and GRIN taxonomy (USDA-ARS National Genetic Resources Program 2012), the latter of which served as the final authority. The Inventory was reviewed by NPGS curators, members of the NPGS Crop Germplasm Committees, and USDA Agricultural Research Service (ARS) crop experts, who submitted revisions and proposed additional taxa.

Prioritization of the Crop Wild Relatives of Agricultural Crops

When using an inclusive definition for crops together with a broad definition of CWR, national and regional studies have resulted in the majority of flora being listed as CWR. Approximately 80% of the species in the European and Mediterranean floras were listed as CWR in an inventory for that region (Kell *et al.* 2008), and 77% of the flora of Portugal similarly listed as CWR or WUS (Magos Brehm *et al.* 2007). Given the extent of potentially useful plant taxa in the US and general resource constraints in conservation and research funding, we further prioritized taxa within the Inventory in order to focus subsequent conservation efforts on species with the greatest potential impact on crop research.

We first compiled and prioritized crop species based on their contribution to global agricultural production and food security, with the assumption that important crops are the focus of the most active breeding programs with experience in the utilization of exotic germplasm. The crop list was collated from the Food and Agriculture Organization of the United Nations statistical database (FAOSTAT) production and food supply data (FAO 2011), Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGR) (FAO 2002), Appendix 2 (“Important Food Crops”) of Groombridge & Jenkins (2002), and Prescott-Allen & Prescott-Allen 1990). These sources emphasize food crops, but some include fiber, forage, and industrial crops.

Listed crops were further prioritized based upon the number of sources and importance attributed within the sources: 1) Major crops (Priority 1) were assigned to crops listed in more

than one source, among the specific crop commodities listed in FAOSTAT and in Prescott-Allen & Prescott-Allen (1990), and all crops in Annex 1 of the ITPGR and in Groombridge & Jenkins (2002); 2) Minor and non-food crops (Priority 2) were assigned to crops listed in only one source, plus crops recorded in FAOSTAT general commodities, as well as Annex 1 forages and the “*Brassica* complex” crops other than those in the genus *Brassica* L. itself. The resulting compilation of the world’s major crops included 242 crops and 268 genera (101 crops and 119 genera in Priority 1, and 141 crops and 149 genera in Priority 2) (Supplementary Table 1). The list included all agricultural crops recorded in FAOSTAT as important to production or to food supply in the US (FAO 2011).

The National Inventory was compared to the world’s major crops list and crosschecked with GRIN taxonomy to derive a list of CWR taxa occurring in the US that are within the gene pools of priority crops. The resulting list of priority CWR was reviewed by NPGS curators, members of the NPGS Crop Germplasm Committees, and ARS crop experts.

Priority 1 CWR taxa were further categorized based upon perceived value and ease of utility in breeding programs. Native plant species were assigned a higher priority, as they have a long history of adaptation in contrast to naturalized species, which may have limited variation due to the founder effect (Amsellem *et al.* 2001). Closely related (defined here as within GP 1-2 or TG 1-3) native taxa, plus any additional taxa recorded in the literature or identified by researchers as potentially useful in crop breeding, were assigned the highest priority (Priority 1A). Distantly related and/or non-native taxa that were not specifically identified by the research community as a target for utilization were listed as Priority 1B. The few gene pools (notably blackberry and raspberry in *Rubus* L.) for which relatedness information was not available were categorized based upon occurrence status.

RESULTS AND DISCUSSION

The National Inventory contains 4,596 taxa, representing 3,912 species from 985 genera and 194 plant families. CWR in the inventory are represented by 2,495 taxa representing 1,905 species from 160 genera and 56 families. WUS are represented by 2,101 taxa from 2,007 species from 833 genera and 182 families. Major families are listed in Table 1. The Inventory is available online at: <http://www.ars-grin.gov/misc/tax/> (accessed 1 Oct. 2012). Future plans are to fully integrate these data into GRIN so that detailed information is available for each

Table 1. Major families of US crop wild relatives (CWR) and wild utilized species (WUS).

<i>CWR (genetic resource of)</i>			<i>WUS (direct use for)</i>		
Families	No. of genera	No. of taxa	Families	No. of genera	No. of taxa
Fabaceae	19	693	Asteraceae	97	216
Poaceae	31	448	Poaceae	71	156
Asteraceae	12	182	Rosaceae	29	135
Rosaceae	6	163	Fabaceae	52	106
Amaranthaceae	6	137	Ericaceae	23	79
Brassicaceae	11	67	Pinaceae	6	71
Grossulariaceae	1	67	Cyperaceae	6	55
Solanaceae	4	63	Salicaceae	2	48
Cyperaceae	1	48	Ranunculaceae	14	45
Amaryllidaceae	1	47	Boraginaceae	17	43
Other (46)	68	580	Other (172)	516	1147

taxon and the Inventory can be queried by taxonomy, priority level, and geographic distribution.

CWR identified in the US are primarily related to food crops (Table 2). These include genetic resources for globally important crops such as strawberry, sunflower, sweet potato [*Ipomoea batatas* (L.) Lam.], bean, bean, stone fruits (*Prunus* L.), and grape, as well as regionally important crops such as pecan, yerba maté (*Ilex paraguariensis* A. St.-Hil.), quinoa (*Chenopodium quinoa* Willd.), and cranberry. Forage and feed CWR are also well represented, particularly legumes (*Trifolium* L., *Lupinus* L., *Lotus* L., and *Astragalus* L.) and grasses (*Agrostis* L., *Bromus* L., *Festuca* L., and *Poa* L.). Significant genetic resources of medicinal crops include Echinacea (*Echinacea* Moench), tobacco (*Nicotiana* L.), St. John's-wort (*Hypericum* L.), poppy (*Papaver* L.), and artemisia (*Artemisia* L.). CWR of ornamental crops include *Rosa* L., *Coreopsis* L., *Lilium* L., *Phlox* L., *Rudbeckia* L. and *Penstemon* Schmidel. CWR of material and industrial crops include relatives of flax (*Linum* L.), cotton (*Gossypium* L.), and jatropha (*Jatropha* L.).

The WUS species listed as distributed in the US are primarily utilized for ornamental, restoration, and medicinal purposes. A number of food species of cultural and economic significance are also identified, such as wild rice (*Zizania* L.), sugar maple (*Acer saccharum* Marshall), and pawpaw (*Asimina triloba* (L.) Dunal).

Non-native species comprise 12.3% of the Inventory (14.7% of CWR and 9.6% of WUS), and 212 taxa (4.6% of total) are Federal and/or State listed noxious weeds. Genetic resource priorities for listed taxa should take into account weed regulations and conservation priorities

Table 2. Uses of US crop wild relatives (CWR) and wild utilized species (WUS).

Use	Major families (and No. of taxa)	No. of families	No. of genera	No. of taxa
<i>CWR (genetic resource of)</i>				
Food	Poaceae (304), Fabaceae (168), Rosaceae (136), Amaranthaceae (95), Asteraceae (90), Grossulariaceae (67), Brassicaceae (61), Solanaceae (54), Cyperaceae (48), Amaryllidaceae (47), Convolvulaceae (40), Ericaceae (39), Asparagaceae (35)	47	103	1472
Forage and feed	Fabaceae (521), Poaceae (142), Amaranthaceae (42)	5	29	709
Medicinal	Asteraceae (63), Hypericaceae (44), Euphorbiaceae (17)	7	10	150
Ornamental	Plantaginaceae (39), Rosaceae (27), Asteraceae (19)	5	6	99
Material and industrial	Linaceae (21), Malvaceae (13), Asteraceae (8)	10	15	57
Herb	Lamiaceae (5)	1	2	5
Soil conservation	Fabaceae (3)	1	1	3
<i>WUS (direct use for)</i>				
Use	Major families and number of taxa	# of families	# of genera	# of taxa
Ornamental	Ericaceae (65), Asteraceae (49), Fabaceae (43), Rosaceae (41), Salicaceae (30), Oleaceae (22), Poaceae (20), Papaveraceae (17), Ranunculaceae (16)	149	440	812
Restoration	Asteraceae (151), Poaceae (53), Cyperaceae (48), Fabaceae (31), Boraginaceae (29), Ranunculaceae (26), Rhamnaceae (26), Apiaceae (23), Onagraceae (22), Liliaceae (22), Polygonaceae (18), Rosaceae (17)	83	336	755
Medicinal	Asteraceae (12), Lamiaceae (11), Fabaceae (9)	82	139	180
Food	Rosaceae (72), Poaceae (6), Sapindaceae (5)	17	26	112
Forage and feed	Poaceae (72), Fabaceae (11), Araceae (5)	9	61	99
Forestry	Pinaceae (49), Cupressaceae (7), Betulaceae (7)	16	28	87
Material and industrial	Fagaceae (9), Cupressaceae (6)	26	37	52
Soil conservation	Fabaceae (3)	1	3	3
Turf	Poaceae (1)	1	1	1

for the species affected by these invasive plants.

Several introduced CWR taxa were identified as containing genetic resources of interest to

breeders, including relatives of beet (*Beta vulgaris* L.) (L. Panella, personal communication 2011), lettuce (B. Hellier, personal communication 2011), and clover (W. Williams, personal communication 1997). Recent alfalfa (*Medicago sativa* L.) breeding efforts have used cold-adapted naturalized populations of *Medicago sativa* L. subsp. *falcata* (L.) Arcang from South Dakota to develop rangeland varieties adapted to the Intermountain West (Peel *et al.* 2009). These examples demonstrate the potential value of novel variation in naturalized species (Bossdorf *et al.* 2005), which should not be neglected in inventories of useful plant species.

The threat status of 3,512 (76.4%) taxa in the Inventory has been recorded in NatureServe. Eight (0.2%) taxa were assessed as known or presumed extinct in the wild; 115 (2.5%) as globally critically imperiled and 111 (2.4%) as imperiled; 337 (7.3%) vulnerable; 798 (17.4%) apparently secure; and 2,143 (46.6%) globally secure. Of the included taxa, the International Union for Conservation of Nature Red List of Threatened Species assessed 16 taxa as extinct, endangered, or vulnerable (IUCN 2012). Sixty-two taxa are listed as endangered under the US Endangered Species Act (Endangered Species Act of 1973, 16 U.S.C. Sec 1531), 10 taxa as threatened, and 11 taxa as candidates for listing (NatureServe 2009). Among the taxonomic groups with the largest number of threatened taxa are members of the family Fabaceae, particularly within the genera *Astragalus*, *Lotus*, *Lupinus*, and *Trifolium* (see Supplementary Table 2 for a full listing of extinct, imperiled, endangered, and threatened taxa).

Threatened species with known or high potential value in crop breeding include the wild walnut *Juglans hindsii* (Jeps.) R. E. Sm., which is used as a primary rootstock for English walnut (*Juglans regia* L.) worldwide and is critically imperiled in its native California habitat (Phillips & Meilleur 1998), and close relatives of sunflower, squash, cotton, gooseberry (*Ribes uva-crispa* L.), raspberry, onion (*Allium cepa* L.), wild rice, and plum (Table 3). The conservation of these genetic resources should be prioritized urgently.

Over 96,000 gene bank accessions of 2,800 taxa listed in the Inventory are recorded in GRIN, but a large proportion of this material is cultivated germplasm conspecific with wild taxa such as American cotton (*G. hirsutum*) and chili pepper (*C. annuum*). Germplasm of Inventory taxa listed as wild total 48,780 accessions, and that listed as both wild and from the US total 20,739 accessions from 2,135 taxa. These accessions are distributed unevenly within the Inventory, with 51.8% of accessions comprised of 14 genera (*Fraxinus* L., *Helianthus* L., *Pinus* L., *Avena*

Table 3. Threatened US crop wild relatives of major crops.

Taxon	Priority category ¹	US ESA ²	Nature Serve ³	No. of accessions ⁴
<i>Allium munzii</i> (Ownbey & Aase ex Traub) McNeal	P1B	LE	G1	0
<i>Allium obtusum</i> Lemmon var. <i>conspicuum</i> Mortola & McNeal	P1B		T2 to 3	0
<i>Allium scilloides</i> Douglas ex S. Watson	P1B		G2 to 3	0
<i>Cucurbita okeechobeensis</i> (Small) L. H. Bailey	P1A	LE	G1	0
<i>Fragaria chiloensis</i> (L.) Mill. subsp. <i>sandwicensis</i> (Decne.) Staudt	P1A		T2	2
<i>Gossypium tomentosum</i> Nutt. ex Seem.	P1A		G2 to 3	0
<i>Helianthus carnosus</i> Small	P1B		G1 to 2	2
<i>Helianthus niveus</i> (Benth.) Brandegees subsp. <i>tephrodes</i> (A. Gray) Heiser	P1A		G2*	10
<i>Helianthus nuttallii</i> Torr. & A. Gray subsp. <i>parishii</i> (A. Gray) Heiser	P1B		GX*	0
<i>Helianthus paradoxus</i> Heiser	P1A		G2	1
<i>Helianthus smithii</i> Heiser	P1B		G2	6
<i>Helianthus verticillatus</i> Small	P1B		G1	2
<i>Hordeum arizonicum</i> Covas	P1B		G2 to 4	0
<i>Ipomoea microdactyla</i> Griseb.	P1B		G2	1
<i>Juglans hindsii</i> (Jeps.) R. E. Sm.	P1A		G1	16
<i>Lathyrus grimesii</i> Barneby	P1B		G2	3
<i>Lathyrus holochlorus</i> (Piper) C. L. Hitchc.	P1B		G2	1
<i>Leymus pacificus</i> (Gould) D. R. Dewey	P1B		G2 to 3	0
<i>Manihot walkerae</i> Croizat	P1B	LE	G1	0
<i>Phaseolus texensis</i> A. Delgado & W. R. Carr	P1B		G2	0
<i>Prunus eremophila</i> Prigge	P1B		G1	0
<i>Prunus murrayana</i> E. J. Palmer	P1A		GX	0
<i>Ribes binominatum</i> A. Heller	P1A		G2 to 3	3
<i>Ribes echinellum</i> (Coville) Rehder	P1B	LT	G1	3
<i>Ribes erythrocarpum</i> Coville & Leiberg	P1B		G2	2
<i>Rubus aliceae</i> L. H. Bailey	P1A		GX	0
<i>Rubus hawaiiensis</i> A. Gray	P1A		G2 to 3	13
<i>Rubus macraei</i> A. Gray	P1A		G2	1
<i>Solanum incompletum</i> Dunal	P1B	LE	G1	0
<i>Solanum nelsonii</i> Dunal	P1B	C	G2	0
<i>Solanum sandwicense</i> Hook. & Arn.	P1B	LE	G1	0
<i>Solanum wallacei</i> (A. Gray) Parish	P1B		G2	0
<i>Tripsacum floridanum</i> Porter ex Vasey	P1A		G2	0
<i>Vanilla mexicana</i> Mill.	P1A		G2 to 4	0
<i>Vicia menziesii</i> Spreng.	P1B	LE	G1	0
<i>Vicia ocalensis</i> R. K. Godfrey & Kral	P1B		G1	1
<i>Zizania texana</i> Hitchc.	P1A	LE	G1	0

¹P1A = native taxa closely related to important crop plants; P1B = non-native and/or distantly related to important crop plants. ²Taxa listed as endangered (LE), threatened (LT), or as a candidate for listing (C) under the US Endangered Species Act of 1973, 16 U.S.C. Sec1531) (US ESA), and/or listed as known or presumed extinct in the wild (GX), globally critically imperiled (G1), and globally imperiled (G2) in NatureServe (NatureServe 2009). Note (G3) is categorized as globally vulnerable, and (G4) as apparently secure; and T denotes global listing at the infraspecific level. *denotes threat assessment at the species level. ⁴Number of accessions denotes NPGS germplasm listed as wild and collected in the US.

L., *Elymus* L., *Vaccinium* L., *Rubus* L., *Vitis* L., *Fragaria* L., *Lupinus* L., *Achnatherum* P. Beauv., *Ribes* L., *Solanum* L., and *Trifolium* L.). Of the 232 taxa listed as endangered, threatened, or as a candidate for listing under the Endangered Species Act (Endangered Species Act of 1973, 16 U.S.C. Sec 1531) as well as taxa listed as known or presumed extinct in the wild, globally critically imperiled, and imperiled in NatureServe (2009), only 157 accessions listed as wild and collected in the US are conserved in the NPGS.

Priority Crop Wild Relatives of Agricultural Crops

Priority species occurring in the US total 2,256 taxa within 176 genera. These include 821 taxa from 69 genera related to 63 major agricultural crops (Priority 1 gene pools) (Table 4), and 1,435 taxa from 107 genera of minor food crops, forages, and other crops (Priority 2) (Supplementary Table 3). Within Priority 1, 285 closely related, native taxa from 30 genera are listed 1A, and 536 distantly related and/or non-native taxa within 57 genera in 1B.

A number of iconic US edible WUS were given priority for conservation considerations. Within Priority 1, these include sugar maple, wild rice, and American chestnut (*Castanea dentata* (Marshall) Borkh.), plants that have held important stature in traditional regional diets. In addition, 148 food, medicinal, and ornamental WUS from 22 genera were assigned to Priority 2.

At least 17 major crops have benefited from traits contributed by 55 Priority 1 CWR taxa (Table 5). As this count is limited to published references, it is likely an underestimate of the taxa occurring in the US that have been successfully utilized in breeding programs.

The NPGS conserves 8,195 accessions of wild native Priority 1 taxa (3,952 Priority 1A and 4,243 Priority 1B), and 4,020 accessions of Priority 2 taxa. Of Priority 1 CWR, 366 (44.6%) taxa are completely absent from *ex situ* collections and another 307 (37.4%) are represented by less than 10 germplasm accessions.

Far from possessing few genetic resources, the United States contains a wealth of native and introduced plants related to a broad range of crops. Significant gaps in the *ex situ* collections of these taxa remain to be filled, and a number of potentially valuable species are threatened in the wild. Meanwhile, new populations of some species are still being discovered (Kraft *et al.*

Table 4. Priority US crop wild relatives and wild utilized species.

Genus	Associated crop name	Priority (and no. of taxa) ¹	Genus	Associated crop name	Priority (and no. of taxa) ¹
<i>Acer</i>	Sugar maple	P1A (6)	<i>Juglans</i>	Walnut	P1A (7) and P1B (2)
<i>Aegilops</i>	Wheat	P1B (5)	<i>Lactuca</i>	Lettuce	P1A (10) and P1B (1)
<i>Agropyron</i>	Wheat	P1B (2)	<i>Lathyrus</i>	Grasspea	P1B (31)
<i>Allium</i>	Onion, garlic, and leek	P1A (4) and P1B (43)	<i>Leymus</i>	Wheat	P1B (17)
<i>Artocarpus</i>	Breadfruit and jackfruit	P1B (1)	<i>Malus</i>	Apple	P1B (4)
<i>Asparagus</i>	Asparagus	P1B (3)	<i>Manihot</i>	Cassava	P1B (4)
<i>Avena</i>	Oat	P1B (3)	<i>Medicago</i>	Alfalfa	P1B (10)
<i>Beta</i>	Beet, sugar beet and chard	P1A (4)	<i>Nicotiana</i>	Tobacco	P1B (9)
<i>Brassica</i>	Cabbage, rapeseed, etc.	P1B (5)	<i>Olea</i>	Olive	P1B (1)
<i>Capsicum</i>	Chili pepper and sweet pepper	P1A (2)	<i>Pennisetum</i>	Pearl millet	P1B (10)
<i>Carica</i>	Papaya	P1B (1)	<i>Persea</i>	Avocado	P1B (3)
<i>Carthamus</i>	Safflower	P1B (1)	<i>Phaseolus</i>	Bean	P1A (4) and P1B (15)
<i>Carya</i>	Pecan	P1A (9) and P1B (4)	<i>Piper</i>	Pepper	P1B (1)
<i>Castanea</i>	Chestnut	P1A (3) and P1B (2)	<i>Pistacia</i>	Pistachio	P1A (1)
<i>Cinnamomum</i>	Cinnamon	P1B (1)	<i>Prunus</i>	Stone fruits	P1A (17) and P1B (26)
<i>Cocos</i>	Coconut	P1B (1)	<i>Psathyrostachys</i>	Wheat	P1B (2)
<i>Colocasia</i>	Taro	P1B (1)	<i>Pseudoroegneria</i>	Wheat	P1B (1)
<i>Corylus</i>	Hazelnut	P1A (3)	<i>Psidium</i>	Guava	P1A (1) and P1B (1)
<i>Cucumis</i>	Melon	P1B (4)	<i>Pyrus</i>	Pear	P1B (1)
<i>Cucurbita</i>	Pumpkin and squash	P1A (8) and P1B (2)	<i>Ribes</i>	Currant and gooseberry	P1A (27) and P1B (40)
<i>Cynara</i>	Artichoke	P1B (3)	<i>Rubus</i>	Raspberry and blackberry	P1A (58) and P1B (10)
<i>Daucus</i>	Carrot	P1B (2)	<i>Saccharum</i>	Sugar cane	P1B (9)
<i>Dioscorea</i>	Yam	P1B (3)	<i>Solanum</i>	Potato and tomato	P1A (1) and P1B (38)
<i>Diospyros</i>	Persimmon	P1A (2)	<i>Sorghum</i>	Sorghum	P1B (4)
<i>Diplotaxis</i>	Cabbage, rapeseed, etc.	P1B (2)	<i>Syzygium</i>	Clove	P1B (2)
<i>Elymus</i>	Wheat	P1B (43)	<i>Thinopyrum</i>	Wheat	P1B (2)
<i>Ficus</i>	Fig	P1A (1) and P1B (3)	<i>Tripsacum</i>	Maize	P1A (4)
<i>Foeniculum</i>	Fennel	P1B (1)	<i>Vaccinium</i>	Blueberry and Cranberry	P1A (23) and P1B (16)
<i>Fragaria</i>	Strawberry	P1A (11) and P1B (10)	<i>Vanilla</i>	Vanilla	P1A (2)
<i>Gossypium</i>	Cotton	P1A (3)	<i>Vernicia</i>	Tung nut	P1B (1)
<i>Helianthus</i>	Sunflower	P1A (23) and P1B (49)	<i>Vicia</i>	Fava bean and vetch	P1B (14)
<i>Hordeum</i>	Barley	P1B (18)	<i>Vigna</i>	Cowpea, bambara groundnut, etc.	P1B (2)
<i>Ilex</i>	Maté	P1A (6) and P1B (15)	<i>Vitis</i>	Grape	P1A (29)
<i>Illicium</i>	Star-anise	P1A (1)	<i>Zizania</i>	Wild rice	P1A (6)
<i>Ipomoea</i>	Sweet potato	P1A (9) and P1B (31)			

¹P1A = native taxa closely related to important crop plants; P1B = non-native and/or distantly related to important crop plants. Origin status from the Germplasm Resources Information Network (USDA-ARS National Genetic Resources Program 2012). Contributing gene pool and taxon group concepts from Wiersema *et al.* (2012), and the 'Harlan and de Wet Crop Wild Relative Inventory' (Vincent *et al.* 2013).

Table 5. Confirmed use of priority crop wild relatives.

Taxon	Trait ¹
<i>Aegilops cylindrica</i> Host	Salt tolerance ^{a,b}
<i>Aegilops geniculata</i> Roth	Hessian fly resistance ^c
<i>Aegilops tauschii</i> Coss.	Rust resistance, wheat soil-borne mosaic virus, and wheat spindle-streak mosaic virus ^d ; drought tolerance ^e ; yellow rust and leaf rust resistance ^f ; glutenins improvement ^g ; agronomic traits and yield improvement ^h ; hessian fly resistance ⁱ ; karnal bunt ^j ; water-logging tolerance ^k ; and sprouting suppression ^l
<i>Allium fistulosum</i> L.	Disease resistance ^m
<i>Avena sterilis</i> L.	Crown rust resistance ⁿ and yield improvement ^o
<i>Corylus americana</i> Marshall	Eastern filbert blight resistance ^p
<i>Helianthus anomalus</i> S. F. Blake	Fertility restoration ^q
<i>Helianthus argophyllus</i> Torr. & A. Gray	Downy mildew resistance ^{r,s} , disease resistance ^t , and fertility restoration ^q
<i>Helianthus bolanderi</i> A. Gray	Genetic stock ^u and fertility restoration ^q
<i>Helianthus debilis</i> Nutt.	Powdery mildew resistance ^v and fertility restoration ^q
<i>Helianthus deserticola</i> Heiser	Downy mildew resistance ^w
<i>Helianthus divaricatus</i> L.	Broomrape resistance ^x
<i>Helianthus giganteus</i> L.	Fertility restoration ^y and cytoplasmic male sterility ^z
<i>Helianthus grosseserratus</i> M. Martens	Broomrape resistance ^x
<i>Helianthus hirsutus</i> Raf.	Fertility restoration aa ^y
<i>Helianthus maximilianii</i> Schrad.	Broomrape resistance ^x and cytoplasmic male sterility ^z
<i>Helianthus neglectus</i> Heiser	Fertility restoration ^q
<i>Helianthus paradoxus</i> Heiser	Salt tolerance ^{ab} and fertility restoration ^q
<i>Helianthus pauciflorus</i> Nutt.	Cytoplasmic male sterility ^{ac} and sclerotinia resistance ^{ad}
<i>Helianthus petiolaris</i> Nutt.	Verticillium resistance ^{ae} , disease resistance ^t , cytoplasmic male sterility ^{af} , sunflower moth resistance ^{ag} , and fertility restoration ^q
<i>Helianthus praecox</i> Engelm. & A. Gray	Downy mildew, rust, verticillium wilt and broomrape resistance ^{ah} ; fertility restoration ^q ; and downy mildew resistance ^w
<i>Helianthus resinosus</i> Small	Fertility restoration ^{aa}
<i>Helianthus strumosus</i> L.	Fertility restoration ^y
<i>Helianthus tuberosus</i> L.	Broomrape resistance ^{ai} , sunflower moth resistance ^{ag} , and fertility restoration ^y
<i>Hordeum bulbosum</i> L.	Powdery mildew resistance ^{aj} , mosaic virus resistance ^{ak} , septoria resistance ^{al} , and leaf rust resistance ^{am}
<i>Ipomoea trifida</i> (Kunth) G. Don	Root knot nematode and root lesion nematode resistance ^{an}
<i>Juglans californica</i> S. Watson	Rootstock ^{ao}
<i>Juglans hindsii</i> (Jepss) R. E. Sm.	Rootstock ^{ap,ao}
<i>Juglans major</i> (Torr.) A. Heller and <i>Juglans microcarpa</i> Berland	Rootstock for alkaline soil ^{ap}
<i>Juglans nigra</i> L.	Anthrachnose resistance ^{ap} and rootstock ^{ao}
<i>Lactuca serriola</i> L.	Downy mildew resistance ^{aq}

Taxon	Trait ¹
<i>Lactuca virosa</i> L.	Leaf aphid resistance ^{ar}
<i>Malus fusca</i> (Raf.) C. K. Schneid.	Rootstock ^{ao}
<i>Medicago sativa</i> L. subsp. <i>falcata</i> (L.) Arcang.	Winter hardiness ^{as}
<i>Pennisetum purpureum</i> Schumach.	Pest resistance, vigor and yield ^{at}
<i>Prunus andersonii</i> A. Gray, <i>Prunus pumila</i> L., <i>Prunus pumila</i> L. var. <i>besseyi</i> (L. H. Bailey) Gleason, and <i>Prunus rivularis</i> Scheele	Rootstock ^{ao}
<i>Pyrus calleryana</i> Decne	Rootstock ^{ao}
<i>Ribes nigrum</i> L.	Pest and disease resistance ^{au}
<i>Ribes uva-crispa</i> L.	Gall mite resistance ^{av}
<i>Solanum stoloniferum</i> Schltdl. & Bouche	Late blight resistance ^{aw} and potato Y virus resistance ^{ax}
<i>Tripsacum dactyloides</i> (L.) L.	Corn leaf blight resistance ^{ay} and yield improvement and top firing resistance ^{af}
<i>Vitis acerifolia</i> Raf., <i>Vitis aestivalis</i> Michx., <i>Vitis cinerea</i> (Engelm.) Engelm. ex Millardet, <i>Vitis cinerea</i> var. <i>helleri</i> (L. H. Bailey) M. O. Moore, <i>Vitis monticola</i> Buckley, <i>Vitis mustangensis</i> Buckley, and <i>Vitis vulpina</i> L.	Rootstock ^{ao}
<i>Vitis labrusca</i> L.	Cold tolerance ^{az}
<i>Vitis riparia</i> Michx. and <i>Vitis rupestris</i> Scheele	<i>Phylloxera vitifoliae</i> resistance ^{af} and rootstock ^{ao}

¹Published trait listing adapted from the 'Harlan and de Wet Crop Wild Relative Inventory' (Vincent *et al.* 2013). ^aFarooq *et al.* 1995, ^bWang *et al.* 2003, ^cEl Khilfi *et al.* 2004, ^dCox *et al.* 1995, ^eGororo *et al.* 2002, ^fMa *et al.* 1995, ^gPena *et al.* 1995 ^hPestsova *et al.* 2006, ⁱSuszkiw 2005, ^jVillareal *et al.* 1996, ^kVillareal *et al.* 2001, ^lXiu-Jin *et al.* 1997. ^mKhrustaleva and Kik 1998, ⁿHoffman *et al.* 2006, ^oTakeda & Frey 1976, ^pThompson *et al.* 1996, ^qSeiler 1991a, ^rHulke *et al.* 2010, ^sMiller & Gulya 1988, ^tJan *et al.* 2004, ^uJan 1992, ^vJan & Chandler 1988, ^wSeiler 1991b, ^xJan *et al.* 2002, ^ySeiler 2000, ^zWhelan & Dedio 1980, ^{aa}Seiler 1991c, ^{ab}Lexer *et al.* 2004, ^{ac}Jan *et al.* 2006, ^{ad}Miller & Gulya 1999, ^{ae}Hoes *et al.* 1973, ^{af}Prescott-Allen & Prescott-Allen 1986, ^{ag}Rogers *et al.* 1984, ^{ah}Hajjar & Hodgkin 2007, ^{ai}Putt 1978, ^{aj}Pickering & Johnston 2005, ^{ak}Ruge-Wehling *et al.* 2006, ^{al}Toubia-Rahme *et al.* 2003, ^{am}Zhang *et al.* 2001, ^{an}Sakamoto 1976, ^{ao}USDA-ARS National Genetic Resources Program 2012, ^{ap}McGranahan & Leslie 2009, ^{aq}Hooftman *et al.* 2007, ^{ar}Eenink *et al.* 1982, ^{as}Barnes *et al.* 1977, ^{at}Hanna 1997, ^{au}Barney & Hummer 2005, ^{av}Brennan 2008, ^{aw}Bradshaw *et al.* 2006, ^{ax}Ross 1979, ^{ay}Goodman *et al.* 1987, ^{az}Reisch & Pratt 1996.

2012). Crops that are nationally as well as globally important to food security could benefit significantly from the long-term conservation and exploitation of these taxa. Following the prioritization of such taxa based upon their potential use value, planning for conservation will be facilitated through an analysis of the range of distribution of these taxa, and the subsequent identification of hotspots of richness of CWR in the US as well as geographic and taxonomic gaps in germplasm collections and *in situ* conservation.

The focus on the gene pools of major agricultural crops during prioritization within the Inventory resulted in a number of minor or locally important crops and WUS, forages and other non-food crops holding secondary priority (Supplementary Table 3). Many of these taxa are economically important and their native US genetic resources may have substantial use value. The development and collation of information both on the utilization of these taxa in breeding programs as well as the value of their associated crops will contribute significantly to their potential for prioritization and subsequent conservation.

Given the considerable development pressures on wild plants in the US (Stein *et al.* 2000) and projected increasing impacts from climate change (Loarie *et al.* 2009), both the urgent collection for *ex situ* conservation and the management of taxa in conservation areas are warranted. In order to achieve these goals for the diversity of prioritized taxa, broad partnerships and networks between the federal, state, tribal and non-governmental organizations pursuing conservation activities are needed. Because many of the taxa are distributed across national borders and the genetic resources of such species are potentially valuable globally, such efforts should be aligned with neighboring national strategies and with regional and global initiatives to conserve and provide access to CWR diversity.

SUPPLEMENTARY INFORMATION

Supplementary Table 1: Prioritization of agricultural crops worldwide.

Priority 1 Crops			
Genus	Crop name	Genus	Crop name
<i>Abelmoschus</i>	Okra	<i>Lablab</i>	Lablab, hyacinth bean
<i>Actinidia</i>	Kiwi	<i>Lactuca</i>	Lettuce
<i>Allium</i>	Onion, garlic, leek	<i>Lathyrus</i>	Grass pea
<i>Anacardium</i>	Cashew	<i>Lens</i>	Lentil
<i>Ananas</i>	Pineapple	<i>Linum</i>	Flax, linseed
<i>Arachis</i>	Peanut	<i>Malus</i>	Apple
<i>Artocarpus</i>	Breadfruit, jackfruit	<i>Mangifera</i>	Mango
<i>Asparagus</i>	Asparagus	<i>Manihot</i>	Cassava
<i>Avena</i>	Oat	<i>Medicago</i>	Alfalfa
<i>Bertholletia</i>	Brazil nut	<i>Musa, Ensete</i>	Banana, plantain
<i>Beta, Patellifolia</i>	Beet, sugar beet, chard	<i>Olea</i>	Olive
<i>Brassica, Diplotaxis</i>	Cabbage, rapeseed, etc.	<i>Oryza</i>	Rice
<i>Cajanus</i>	Pigeonpea	<i>Papaver</i>	Poppy
<i>Camellia</i>	Tea	<i>Pennisetum</i>	Pearl millet
<i>Capsicum</i>	Chili pepper, sweet pepper	<i>Persea</i>	Avocado
<i>Carica</i>	Papaya, babaco	<i>Phaseolus</i>	Bean
<i>Carthamus</i>	Safflower	<i>Phoenix</i>	Date
<i>Ceratonia</i>	Carob	<i>Pimenta</i>	Pimento
<i>Cicer</i>	Chickpea	<i>Pimpinella</i>	Anise
<i>Cicorium</i>	Chicory	<i>Piper</i>	Pepper
<i>Cinnamomum</i>	Cinnamon	<i>Pistacia</i>	Pistachio
<i>Citrullus</i>	Watermelon	<i>Pisum, Vavilovia</i>	Pea
<i>Citrus, Fortunella*, Poncirus*</i>	Orange, grapefruit, lemon, etc.	<i>Prunus</i>	Stone fruits
<i>Cocos</i>	Coconut	<i>Psidium</i>	Guava
<i>Coffea</i>	Coffee	<i>Pyrus</i>	Pear
<i>Colocasia</i>	Taro	<i>Ribes</i>	Currant, gooseberry
<i>Coriandrum</i>	Coriander	<i>Ricinus</i>	Castor oil
<i>Corylus</i>	Hazelnut	<i>Rubus</i>	Raspberry, blackberry
<i>Cucumis</i>	Melon, cucumber	<i>Saccharum</i>	Sugar cane
<i>Cucurbita</i>	Pumpkin, squash	<i>Secale</i>	Rye
<i>Cydonia</i>	Quince	<i>Sesamum</i>	Sesame
<i>Cynara</i>	Artichoke	<i>Sinapis</i>	Mustard seed
<i>Daucus, Tornabenea</i>	Carrot	<i>Solanum, Lycopersicon</i>	Potato, tomato, eggplant
<i>Dioscorea</i>	Yam	<i>Sorghum</i>	Sorghum
<i>Diospyros</i>	Persimmon	<i>Spinacia</i>	Spinach
<i>Elaeis</i>	Oil palm	<i>Syzygium</i>	Clove
<i>Elettaria</i>	Cardamom	<i>Theobroma</i>	Cocoa bean
<i>Eleusine</i>	Finger millet	<i>Trifolium</i>	Clover
<i>Fagopyrum</i>	Buckwheat	<i>Triticosecale</i>	Triticale

Chapter 3

<i>Ficus</i>	Fig	<i>Triticum, Aegilops, Amblyopyrum, Dasypyrum, Elymus, Leymus, Psathyrostachys, Pseudoroegneria, Thinopyrum</i>	Wheat
<i>Foeniculum</i>	Fennel	<i>Vaccinium</i>	Blueberry, cranberry
<i>Fragaria</i>	Strawberry	<i>Vanilla</i>	Vanilla
<i>Garcinia</i>	Mangosteen	<i>Vernicia</i>	Tung nut
<i>Glycine</i>	Soybean	<i>Vicia</i>	Fava bean, vetch
<i>Gossypium</i>	Cotton	<i>Vigna</i>	Cowpea, bambara groundnut, etc.
<i>Helianthus</i>	Sunflower	<i>Vitellaria</i>	Karite nut, shea nut
<i>Hordeum</i>	Barley	<i>Vitis</i>	Grape
<i>Ilex</i>	Maté	<i>Xanthosoma</i>	Yautia, cocoyam
<i>Illicium</i>	Badian	<i>Zea, Tripsacum</i>	Maize, popcorn
<i>Ipomoea</i>	Sweet potato	<i>Zingiber</i>	Ginger
<i>Juglans</i>	Walnut		
Priority 2 Crops			
Genus	Crop name	Genus	Crop name
<i>Abroma</i>	Indian flax	<i>Lepidium</i>	Cress
<i>Abutilon</i>	China jute	<i>Lespedeza</i>	Lespedeza
<i>Achras</i>	Sapodilla, chicle gum	<i>Licania</i>	Licania
<i>Aframomum</i>	Aframomum cardamom	<i>Lolium</i>	Lolium
<i>Agrostis</i>	Agrostis	<i>Lotus</i>	Lotus
<i>Aleurites</i>	Aleurites moluccana	<i>Lupinus</i>	Lupin
<i>Alopecurus</i>	Alopecurus	<i>Lygeum</i>	Alfa, esparto
<i>Amaranthus</i>	Quihuicha, Inca wheat	<i>Macadamia</i>	Macadamia nut
<i>Andropogon</i>	Andropogon	<i>Majorana</i>	Marjoram
<i>Anethum</i>	Dill seed	<i>Mammea</i>	Mammee
<i>Annona</i>	Cherimoya, custard apple	<i>Manilkara</i>	Balata
<i>Anthriscus</i>	Chervil	<i>Maranta</i>	Arrowroot
<i>Apium</i>	Celery	<i>Melilotus</i>	Melilotus
<i>Arbutus</i>	Strawberry tree	<i>Mentha</i>	Peppermint
<i>Areca</i>	Areca nut	<i>Mespilus</i>	Medlar
<i>Arenga</i>	Sugar palm	<i>Metroxylon</i>	Sago palm
<i>Armoracia, Barbarea, Camelina, Crambe, Eruca, Isatis, Raphanobrassica, Rorippa</i>	Brassica complex	<i>Morus</i>	Mulberry, loganberry
<i>Arracacoa</i>	Arracacha	<i>Myristica</i>	Nutmeg, mace
<i>Arrhenatherum</i>	Arrhenatherum	<i>Myrtus</i>	Myrtleberry
<i>Artemisia</i>	Tarragon	<i>Nasturtium</i>	Watercress
<i>Asimina</i>	Pawpaw	<i>Neoglaziovia</i>	Caroa
<i>Astrocaryum</i>	Astrocaryum	<i>Nephelium</i>	Litchi, longan, rambutan
<i>Atriplex</i>	Atriplex	<i>Nicotiana</i>	Tobacco
<i>Averrhoa</i>	Carambola	<i>Onobrychis</i>	Onobrychis
<i>Bambusa</i>	Bamboo	<i>Opuntia</i>	Prickly pear
<i>Bassia</i>	Bassia	<i>Ornithopus</i>	Ornithopus

An inventory of crop wild relatives of the United States

<i>Boehmeria</i>	Ramie	<i>Oxalis</i>	Oca
<i>Calocarpum</i>	Sapote	<i>Pachyrhizus</i>	Jicama, yam bean
<i>Canarium</i>	Pili nut, Java almond, Chinese olives	<i>Palachium</i>	Gutta-percha
<i>Canavalia</i>	jack bean	<i>Panicum</i>	Little millet, proso millet
<i>Cannabis</i>	Hemp	<i>Parthenium</i>	Guayule
<i>Capparis</i>	Caper	<i>Paspalum</i>	Kodo millet
<i>Carapa</i>	Carapa	<i>Passiflora</i>	Passion fruit
<i>Caryocar</i>	Butter nut	<i>Pastinaca</i>	Parsnip
<i>Castanea</i>	Chestnut	<i>Perilla</i>	Perilla
<i>Ceiba</i>	Kapok fruit	<i>Petroselinum</i>	Parsley
<i>Chenopodium</i>	Quinoa	<i>Phalaris</i>	Canary seed
<i>Chrysanthemum</i>	Pyrethrum	<i>Phleum</i>	Phleum
<i>Chrysophyllum</i>	Star apple	<i>Phormium</i>	New Zealand flax
<i>Cochlearia</i>	Horseradish	<i>Poa</i>	Poa
<i>Coix</i>	Adlay, Job's tears	<i>Pongamia</i>	Pongamia
<i>Cola</i>	Kola nut	<i>Prosopis</i>	Prosopis
<i>Corchorus</i>	Jute	<i>Psophocarpus</i>	Winged bean
<i>Coronilla</i>	Coronilla	<i>Pueraria</i>	Pueraria
<i>Crataegus</i>	Azarole	<i>Punica</i>	Pomegranate
<i>Crocus</i>	Saffron	<i>Raphanus</i>	Radish
<i>Crotalaria</i>	Sunn hemp	<i>Rheum</i>	Rhubarb
<i>Croton</i>	Croton	<i>Rosa</i>	Rose hips
<i>Curcuma</i>	Turmeric	<i>Rumex</i>	Sorrel
<i>Cyamopsis</i>	Guar bean	<i>Salsola</i>	Salsola
<i>Cyperus</i>	Chufa	<i>Sambucus</i>	Elderberry
<i>Dactylis</i>	Dactylis	<i>Samuela</i>	Palma ixtle
<i>Dieva</i>	Jelutong	<i>Sansevieria</i>	Bowstring hemp
<i>Digitaria</i>	Fonio	<i>Satureja</i>	Savory
<i>Durio</i>	Durian	<i>Scorzonera</i>	Scorzonera
<i>Echinochloa</i>	Barnyard Millet, Japanese millet	<i>Setaria</i>	Foxtail millet
<i>Eragrostis</i>	Teff	<i>Shorea</i>	Shorea
<i>Eriobotrya</i>	Loquat	<i>Sorbus</i>	Rowanberry, service-apple
<i>Fagus</i>	Beech nut	<i>Spondias</i>	Hog plum, mombin
<i>Feijoa</i>	Feijoa	<i>Stillingia</i>	Stillingia
<i>Festuca</i>	Festuca	<i>Stipa</i>	Alfa, esparto
<i>Furcraea</i>	Fuque fibre, Mauritius hemp	<i>Stizolobium</i>	Velvet bean
<i>Gaylussacia</i>	Huckleberry, dangleberry	<i>Tamarindus</i>	Tamarind
<i>Guizotia</i>	Guizotia	<i>Thymus</i>	Thyme
<i>Hedysarum</i>	Hedysarum	<i>Tragopogon</i>	Oyster plant
<i>Hevea</i>	Rubber	<i>Triadica</i>	Tallowtree Seed
<i>Hibiscus</i>	Kenaf, meshta, rosella hemp	<i>Trigonella</i>	Fenugreek seed
<i>Humulus</i>	Hops	<i>Tropaeolum</i>	Mashua
<i>Jatropha</i>	Jatropha	<i>Ullucus</i>	Ullucu
<i>Laurus</i>	Bay leaves	<i>Urena</i>	Congo jute, malva, paka
<i>Lecythis</i>	Paradise nut	<i>Zizyphus</i>	Jujube

Priority 1 crops are generally major food crops; Priority 2 crops are minor food crops, forages, and other crops. *denotes genera used for rootstock. Prioritization formed from an analysis of crops listed in FAO (2002), FAO (2011), Groombridge & Jenkins (2002), and Prescott-Allen & Prescott-Allen (1990).

Chapter 3

Supplementary Table 2: Threatened US crop wild relatives and wild utilized species.

Taxon	Priority category	US ESA	NatureServe	IUCN	# of accessions
<i>Abies fraseri</i>			G2	VU	0
<i>Abutilon eremitopetalum</i>	P2	LE	G1	CR	0
<i>Abutilon menziesii</i>	P2	LE	G1	CR	0
<i>Abutilon parishii</i>	P2		G2		0
<i>Abutilon sandwicense</i>	P2	LE	G1	CR	0
<i>Agave murpheyi</i>			G2		0
<i>Agave schottii</i> var. <i>treleasei</i>			T1		0
<i>Agave shawii</i>			G2-3		0
<i>Agrostis hendersonii</i>	P2		G1		0
<i>Allium munzii</i>	P1B	LE	G1		0
<i>Allium obtusum</i> var. <i>conspicuum</i>	P1B		T2-3		0
<i>Allium scilloides</i>	P1B		2-3		0
<i>Alopecurus aequalis</i> var. <i>sonomensis</i>	P2	LE	T1		0
<i>Amaranthus brownii</i>	P2	LE	G1	CR	0
<i>Amaranthus pumilus</i>	P2	LT	G2		6
<i>Artemisia porteri</i>	P2		G2		0
<i>Asimina tetramera</i>	P2	LE	G1	EN	5
<i>Astragalus aequalis</i>			G2		2
<i>Astragalus agnicidus</i>			G2		0
<i>Astragalus albens</i>		LE	G1		0
<i>Astragalus ampullarioides</i>		LE	G1		4
<i>Astragalus applegatei</i>		LE	G1		0
<i>Astragalus bibullatus</i>		LE	G1		0
<i>Astragalus brauntonii</i>		LE	G2		0
<i>Astragalus calycosus</i> var. <i>monophyllidius</i>			T2		0
<i>Astragalus clarianus</i>		LE	G1		0
<i>Astragalus cottonii</i>			G2		0
<i>Astragalus cremnophylax</i>			G1		0
<i>Astragalus cremnophylax</i> var. <i>cremnophylax</i>		LE	T1		0
<i>Astragalus cremnophylax</i> var. <i>myriorrhaphis</i>			T1		0
<i>Astragalus cusickii</i> var. <i>packardiae</i>		C	T2		1
<i>Astragalus desereticus</i>		LT	G1		0
<i>Astragalus diversifolius</i>			G2		0
<i>Astragalus geyeri</i> var. <i>triquetrus</i>			T2-3		0
<i>Astragalus holmgreniorum</i>		LE	G1		0
<i>Astragalus humillimus</i>		LE	G1		0
<i>Astragalus hypoxylus</i>			G1		0
<i>Astragalus jaegerianus</i>		LE	G1		0
<i>Astragalus kentrophyta</i> var. <i>danaus</i>			T2-3		0
<i>Astragalus lentiginosus</i> var. <i>coachellae</i>		LE	T2		1
<i>Astragalus lentiginosus</i> var. <i>micans</i>			T1		1
<i>Astragalus lentiginosus</i> var. <i>piscinensis</i>		LT	T1		1
<i>Astragalus limnocharis</i>			G2		0
<i>Astragalus limnocharis</i> var. <i>limnocharis</i>			T1		0
<i>Astragalus limnocharis</i> var. <i>montii</i>		LT	T1		0
<i>Astragalus magdalenae</i> var. <i>peirsonii</i>		LT	T2		0
<i>Astragalus microcymbus</i>		C	G1		0
<i>Astragalus mulfordiae</i>			G2		0
<i>Astragalus nevinii</i>			G2		0
<i>Astragalus osterhoutii</i>		LE	G1		0

An inventory of crop wild relatives of the United States

Taxon	Priority category	US ESA	NatureServe	IUCN	# of accessions
<i>Astragalus phoenix</i>		LT	G2		0
<i>Astragalus preussii</i> var. <i>laxiflorus</i>			T2		0
<i>Astragalus pycnostachyus</i>			G2		0
<i>Astragalus pycnostachyus</i> var. <i>lanosissimus</i>		LE	T1		0
<i>Astragalus pycnostachyus</i> var. <i>pycnostachyus</i>			T2		0
<i>Astragalus rafaelensis</i>			G2-3		1
<i>Astragalus rattanii</i> var. <i>jepsonianus</i>			T2		0
<i>Astragalus ravenii</i>			G1		0
<i>Astragalus robbinsii</i> var. <i>jesupii</i>		LE	T1		0
<i>Astragalus robbinsii</i> var. <i>robbinsii</i>			TX		0
<i>Astragalus schmolliae</i>		C	G1		0
<i>Astragalus sinuatus</i>			G1		0
<i>Astragalus tener</i>			G1		0
<i>Astragalus tener</i> var. <i>tener</i>			T1		0
<i>Astragalus tener</i> var. <i>titi</i>		LE	T1		0
<i>Astragalus traskiae</i>			G2		0
<i>Astragalus tricarinatus</i>		LE	G1		0
<i>Astragalus trichopodus</i> var. <i>trichopodus</i>			T2-3		1
<i>Astrophytum asterias</i>			G1	VU	0
<i>Atriplex canescens</i> var. <i>gigantea</i>	P2		T1		0
<i>Atriplex coronata</i> var. <i>notatior</i>	P2	LE	T1		0
<i>Atriplex joaquiniana</i>	P2		G2		0
<i>Betula uber</i>		LT	G1		1
<i>Canavalia galeata</i>	P2		G2		0
<i>Canavalia molokaiensis</i>	P2	LE	G1	CR	0
<i>Canavalia napaliensis</i>	P2	LE	G1	CR	0
<i>Capparis sandwichiana</i>	P2		G2	VU	0
<i>Ceanothus cyaneus</i>			G2		0
<i>Ceanothus impressus</i>			G2-3		0
<i>Centrosema arenicola</i>			G2		0
<i>Chenopodium foggii</i>	P2		G2-3		0
<i>Chenopodium incanum</i> var. <i>occidentale</i>	P2		T2-4		1
<i>Cneoridium dumosum</i>			G2-3		3
<i>Crataegus beata</i>	P2		G2-4		0
<i>Crataegus harbisonii</i>	P2		G1		1
<i>Crotalaria avonensis</i>	P2	LE	G1		0
<i>Croton alabamensis</i> var. <i>texensis</i>	P2		T2		0
<i>Cucurbita okeechobeensis</i>	P1A	LE	G1		0
<i>Cuphea aspera</i>	P2		G2		1
<i>Cupressus macrocarpa</i>			G1		0
<i>Cyperus fauriei</i>	P2	LE	G1		0
<i>Cyperus pennatiformis</i>	P2	LE	G1		0
<i>Cyperus pennatiformis</i> var. <i>bryanii</i>	P2		T1		0
<i>Cyperus pennatiformis</i> var. <i>pennatiformis</i>	P2		TX		0
<i>Cyperus rockii</i>	P2		GX		0
<i>Cyperus trachysanthos</i>	P2	LE	G1		0
<i>Desmodium humifusum</i>			G1-2		0
<i>Digitaria floridana</i>	P2		G1		0
<i>Digitaria pauciflora</i>	P2	C	G1		0
<i>Echinacea angustifolia</i> var. <i>strigosa</i>	P2		TX		2
<i>Echinacea laevigata</i>	P2	LE	G2-3		9

Chapter 3

Taxon	Priority category	US ESA	NatureServe	IUCN	# of accessions
<i>Echinacea paradoxa</i>	P2		G2		1
<i>Echinacea paradoxa</i> var. <i>neglecta</i>	P2		T1		4
<i>Echinacea paradoxa</i> var. <i>paradoxa</i>	P2		T2		5
<i>Echinacea tennesseensis</i>	P2		G2		4
<i>Eragrostis fosbergii</i>	P2	LE	G1		0
<i>Ericameria pinifolia</i>			G2-3		0
<i>Eriogonum cinereum</i>			G2		0
<i>Eriogonum crocatum</i>			G2		0
<i>Eriogonum giganteum</i>			G2		0
<i>Eugenia koolauensis</i>	P2	LE	G1	EN	0
<i>Festuca hawaiiensis</i>	P2	C	G1		0
<i>Festuca ligulata</i>	P2	C	G1		0
<i>Fragaria chiloensis</i> subsp. <i>sandwicensis</i>	P1A		T2		2
<i>Franklinia alatomaha</i>			GX	EW	0
<i>Fraxinus papillosa</i>			G2		0
<i>Fremontodendron mexicanum</i>			G2		0
<i>Gossypium tomentosum</i>	P1A		G2-3		0
<i>Guaiacum sanctum</i>			G2	EN	0
<i>Hazardia cana</i>			G2		0
<i>Hedysarum boreale</i> var. <i>gremiale</i>	P2		T1		0
<i>Helianthus carnosus</i>	P1B		G1-2		2
<i>Helianthus niveus</i> subsp. <i>tephrodes</i>	P1A		G2*		10
<i>Helianthus nuttallii</i> subsp. <i>parishii</i>	P1B		GX*		0
<i>Helianthus paradoxus</i>	P1A		G2		1
<i>Helianthus smithii</i>	P1B		G2		6
<i>Helianthus verticillatus</i>	P1B		G1		2
<i>Hibiscus brackenridgei</i>	P2	LE	G1		0
<i>Hibiscus clayi</i>	P2	LE	G1	CR	0
<i>Hibiscus dasycalyx</i>	P2	C	G1		0
<i>Hibiscus waimeae</i>	P2		G2		0
<i>Hordeum arizonicum</i>	P1B		G2-4		0
<i>Hypericum cumulicola</i>		LE	G2		0
<i>Hypericum lissophloeus</i>			G2		1
<i>Ipomoea microdactyla</i>	P1B		G2		1
<i>Juglans hindsii</i>	P1A		G1		16
<i>Lathyrus grimesii</i>	P1B		G2		3
<i>Lathyrus holochlorus</i>	P1B		G2		1
<i>Lepidium arbuscula</i>	P2	LE	G1		0
<i>Lepidium barnebyanum</i>	P2	LE	G1		0
<i>Lepidium crenatum</i>	P2		G2		1
<i>Lepidium jaredii</i>	P2		G1		0
<i>Lepidium papilliferum</i>	P2	LT	G2		0
<i>Lespedeza leptostachya</i>	P2	LT	G3		0
<i>Leymus pacificus</i>	P1B		G2-3		0
<i>Lindera melissifolia</i>		LE	G2-3		0
<i>Linum carteri</i>	P2		G2		0
<i>Linum carteri</i> var. <i>carteri</i>	P2	C	T1		0
<i>Linum carteri</i> var. <i>smallii</i>	P2		T2		0
<i>Linum lewisii</i> var. <i>alpicola</i>	P2		T2-4		0
<i>Lotus argophyllus</i> var. <i>adsurgens</i>	P2		T1		0
<i>Lotus argyraeus</i> var. <i>multicaulis</i>	P2		T1		0

An inventory of crop wild relatives of the United States

Taxon	Priority category	US ESA	NatureServe	IUCN	# of accessions
<i>Lotus argyraeus</i> var. <i>notitius</i>	P2		T1		0
<i>Lotus crassifolius</i> var. <i>otayensis</i>	P2		T1		0
<i>Lotus dendroideus</i> var. <i>traskiae</i>	P2	LE	T2		0
<i>Lotus nuttallianus</i>	P2		G1		0
<i>Lotus oblongifolius</i> var. <i>cupreus</i>	P2		T2		0
<i>Lotus procumbens</i> var. <i>jepsonii</i>	P2		T1-2		0
<i>Lotus rubriflorus</i>	P2		G1		0
<i>Lupinus angustifolius</i>	P2		G1-5		0
<i>Lupinus antoninus</i>	P2		G1		0
<i>Lupinus arboreus</i>	P2		G1-5		7
<i>Lupinus citrinus</i>	P2		G2		1
<i>Lupinus nipomensis</i>	P2	LE	G1		0
<i>Lupinus rivularis</i>	P2		G2-4		6
<i>Lupinus tidestromii</i>	P2	LE	G2		0
<i>Lupinus westianus</i> var. <i>aridorum</i>	P2	LE	T1		0
<i>Lyonothamnus floribundus</i> subsp. <i>aspleniifolius</i>			T2		0
<i>Malva assurgentiflora</i>			G2		0
<i>Manihot walkerae</i>	P1B	LE	G1		0
<i>Mespilus canescens</i>	P2		G1		1
<i>Opuntia basilaris</i> var. <i>longiareolata</i>	P2		T2		0
<i>Opuntia basilaris</i> var. <i>treleasei</i>	P2	LE	T2		0
<i>Opuntia engelmannii</i> var. <i>flexospina</i>	P2		T1		0
<i>Opuntia engelmannii</i> var. <i>linguiformis</i>	P2		T1		0
<i>Panicum fauriei</i> var. <i>carteri</i>	P2	LE	T1		0
<i>Panicum niihauense</i>	P2	LE	G1		0
<i>Paxistima canbyi</i>			G2		0
<i>Penstemon barrettiae</i>	P2		G2		0
<i>Penstemon haydenii</i>	P2	LE	G1-2		1
<i>Phaseolus texensis</i>	P1B		G2		0
<i>Pinus radiata</i>			G1	LR/ cd VU	0
<i>Pinus torreyana</i>			G1		0
<i>Poa atropurpurea</i>	P2	LE	G2		0
<i>Poa mannii</i>	P2	LE	G1		0
<i>Poa napensis</i>	P2	LE	G1		0
<i>Poa sandvicensis</i>	P2	LE	G1		0
<i>Poa siphonoglossa</i>	P2	LE	G1		0
<i>Portulaca molokiniensis</i>	P2		G1		
<i>Portulaca sclerocarpa</i>	P2	LE	G2		
<i>Portulaca umbraticola</i> subsp. <i>coronata</i>	P2		T2		
<i>Prunus eremophila</i>	P1B		G1		0
<i>Prunus murrayana</i>	P1A		GX		0
<i>Quercus dumosa</i>			G1-2	EN	0
<i>Ribes binominatum</i>	P1A		G2-3		3
<i>Ribes echinellum</i>	P1B	LT	G1		3
<i>Ribes erythrocarpum</i>	P1B		G2		2
<i>Robinia hispida</i> var. <i>fertilis</i>			T1		
<i>Robinia viscosa</i> var. <i>hartwigii</i>			T2		
<i>Rorippa subumbellata</i>	P2	C	G1		0
<i>Roystonea regia</i>			G2-3		0
<i>Rubus aliceae</i>	P1A		GX		0

Chapter 3

Taxon	Priority category	US ESA	NatureServe	IUCN	# of accessions
<i>Rubus hawaiiensis</i>	P1A		G2-3		13
<i>Rubus macraei</i>	P1A		G2		1
<i>Rumex giganteus</i>	P2		G2-3		0
<i>Santalum ellipticum</i>			G2-3		0
<i>Setaria arizonica</i>	P2		G2-4		0
<i>Solanum incompletum</i>	P1B	LE	G1		0
<i>Solanum nelsonii</i>	P1B	C	G2		0
<i>Solanum sandwicense</i>	P1B	LE	G1		0
<i>Solanum wallacei</i>	P1B		G2		0
<i>Spiraea virginiana</i>		LT	G2		
<i>Suaeda californica</i>			G1		0
<i>Trifolium amoenum</i>	P2	LE	G1		1
<i>Trifolium andinum</i> var. <i>podocephalum</i>	P2		T1		0
<i>Trifolium barnebyi</i>	P2		G1-2		0
<i>Trifolium buckwestiorum</i>	P2		G1		1
<i>Trifolium calcaricum</i>	P2		G1		0
<i>Trifolium douglasii</i>	P2		G2		3
<i>Trifolium jokerstii</i>	P2		G1		1
<i>Trifolium leibergii</i>	P2		G2		1
<i>Trifolium neurophyllum</i>	P2		G2		0
<i>Trifolium owyheense</i>	P2		G2		2
<i>Trifolium stoloniferum</i>	P2	LE	G3		3
<i>Trifolium thompsonii</i>	P2		G2		4
<i>Trifolium trichocalyx</i>	P2	LE	G1		1
<i>Tripsacum floridanum</i>	P1A		G2		0
<i>Vanilla mexicana</i>	P1A		G2-4		0
<i>Vicia menziesii</i>	P1B	LE	G1		0
<i>Vicia ocalensis</i>	P1B		G1		1
<i>Wikstroemia uva-ursi</i>			G2		0
<i>Zizania texana</i>	P1A	LE	G1		0

Taxa listed as endangered (LE), threatened (LT), or as a candidate for listing (C) under the US Endangered Species Act (US ESA), and/or listed as known or presumed extinct in the wild (GX), globally critically imperiled (G1), and globally imperiled (G2) in NatureServe (NatureServe, 2012). Note (G3) is categorized as globally vulnerable, (G4) as apparently secure, and (G5) as globally secure; and T denotes global listing at the infraspecific level. *denotes threat assessment at the species level. IUCN Red Listing categories include extinct in the wild (EW), critically endangered (CR), endangered (EN), vulnerable (VU), and lower risk/conservation dependent (LR/cd) (IUCN 2012). P1A = native taxa closely related to important crop plants; P1B = non-native and/or distantly related to important crop plants. # of accessions denotes NPGS germplasm listed as wild and collected in the U.S.

Supplementary Table 3: Additional priority US crop wild relatives and wild utilized species.

Genus	Associated crop name	# of taxa	Genus	Associated crop name	# of taxa
<i>Abutilon</i>	Chinese bell flower,	8			
	China jute				
<i>Actaea</i>	Black cohosh	1	<i>Licania</i>	Oiticica	1
<i>Agave</i>	Agave	4	<i>Lilium</i>	Lily	5
<i>Agrostis</i>	Bentgrass	15	<i>Limnanthes</i>	Meadowfoam	1
<i>Alopecurus</i>	Foxtail grass	4	<i>Linum</i>	Flax	21
<i>Amaranthus</i>	Amaranth	40	<i>Lolium</i>	Annual ryegrass	3
<i>Andropogon</i>	Andropogon	13	<i>Lotus</i>	Lotus	77
<i>Annona</i>	Cherimoya, custard apple	1	<i>Lupinus</i>	Lupine	95
<i>Apios</i>	American groundnut, potato bean	1	<i>Manilkara</i>	Balata	1
<i>Apium</i>	Celery	1	<i>Melilotus</i>	Sweet clover	3
<i>Arbutus</i>	Strawberry tree	3	<i>Mentha</i>	Mint	4
<i>Armoracia</i>	Horseradish	1	<i>Mespilus</i>	Medlar	1
<i>Aronia</i>	Chokeberry	3	<i>Morus</i>	Mulberry, loganberry	2
<i>Arrhenatherum</i>	Oat-grass	2	<i>Nasturtium</i>	Watercress	4
<i>Artemisia</i>	Artemisia, wormwood, tarragon	50	<i>Oplopanax</i>	Devil's club	1
<i>Asimina</i>	Pawpaw	9	<i>Opuntia</i>	Prickly pear	31
<i>Atriplex</i>	Saltbush	37	<i>Oxalis</i>	Oca	8
<i>Bassia</i>	Bassia	1	<i>Panax</i>	Ginseng	1
<i>Boehmeria</i>	Ramie	1	<i>Panicum</i>	Little millet, proso millet	37
<i>Bromus</i>	Brome	35	<i>Papaver</i>	Poppy	14
<i>Camelina</i>	Camelina	1	<i>Parthenium</i>	Guayule	7
<i>Canavalia</i>	Jack bean	6	<i>Paspalum</i>	Kodo millet, ditch millet	42
<i>Capparis</i>	Caper	2	<i>Passiflora</i>	Passion fruit	13
<i>Chenopodium</i>	Quinoa	51	<i>Pastinaca</i>	Parsnip	1
<i>Chrysanthemum</i>	Pyrethrum	1	<i>Penstemon</i>	Penstemon	39
<i>Chrysophyllum</i>	Star apple	2	<i>Phalaris</i>	Canary seed, Phalaris	6
<i>Cochlearia</i>	Scurvy-grass	1	<i>Phleum</i>	Phleum	2
<i>Coix</i>	Adlay, Job's tears	1	<i>Phlox</i>	Phlox	9
<i>Corchorus</i>	Corchorus	2	<i>Physalis</i>	Groundcherry, tomatillo	13
<i>Coreopsis</i>	Coreopsis	8	<i>Physaria</i>	Physaria	4
<i>Crataegus</i>	Hawthorn azarole	70	<i>(Lesquerella)</i>		
<i>Crotalaria</i>	Sunn hemp	6	<i>Pinus</i>	Pine nut	4
			<i>Poa</i>	Kentucky blue grass	42
<i>Croton</i>	Croton	15	<i>Portulaca</i>	Portulaca	10
<i>Cuphea</i>	Cuphea	5	<i>Prosopis</i>	Mesquite	9
<i>Cyperus</i>	Chufa	48	<i>Pueraria</i>	Pueraria	3
<i>Dactylis</i>	Dactylis	1	<i>Raphanus</i>	Radish	1
<i>Digitaria</i>	Fonio	20	<i>Rhododendron</i>	Rhododendron, azalea	30

Chapter 3

Genus	Associated crop name	# of taxa	Genus	Associated crop name	# of taxa
<i>Echinacea</i>	Echinacea	13	<i>Rorippa</i>	Cress	9
<i>Echinochloa</i>	Barnyard millet, Japanese millet	15	<i>Rosa</i>	Rose	27
<i>Eragrostis</i>	Teff	27	<i>Rudbeckia</i>	Rudbeckia	11
<i>Eruca</i>	Rocket	2	<i>Ruellia</i>	Ruellia	2
<i>Eugenia</i>	Pitanga, Suriname Cherry	3	<i>Rumex</i>	Sorrel	19
<i>Fagus</i>	Beech nut	2	<i>Salsola</i>	Salsola	4
<i>Festuca</i>	Festuca	36	<i>Sambucus</i>	Elderberry	11
<i>Gaylussacia</i>	Huckleberry, dangleberry	8	<i>Satureja</i>	Savory	1
<i>Glycyrrhiza</i>	Liquorice	1	<i>Scorzonera</i>	Scorzonera	1
<i>Hedysarum</i>	Hedysarum	7	<i>Setaria</i>	Foxtail millet	27
<i>Hibiscus</i>	Kenaf, meshta, rosella hemp	18	<i>Simmondsia</i>	Jobba	1
<i>Humulus</i>	Hop	6	<i>Sorbus</i>	Rowanberry, service-apple	11
<i>Hydrastis</i>	Goldenseal	1	<i>Stillingia</i>	Stillingia	2
<i>Hypericum</i>	St John's wort	1	<i>Thlaspi</i>	Field penny-cress	1
<i>Jatropha</i>	Jatropha	4	<i>Tragopogon</i>	Oyster plant	4
<i>Lepidium</i>	Cress	37	<i>Triadica</i>	Tallowtree Seed	1
<i>Lespedeza</i>	Lespedeza	11	<i>Trifolium</i>	Clover	96

REFERENCES

- Aberton, M. (2007). Interspecific hybridization in the genus *Trifolium*. *Plant Breed.* 126, 337-342.
- Amsellem, L., Noyer, J.L., Le Bourgeois, T., Hossaert-McKey, M. (2001). Comparison of genetic diversity of the invasive weed *Rubus alceifolius* Poir. (Rosaceae) in its native range and in areas of introduction, using amplified fragment length polymorphism (AFLP) markers. *Mol. Ecol.* 9, 443-455.
- Andersson, M.S., de Vicente, M.C. (2010). *Gene Flow Between Crops and their Wild Relatives* (Baltimore: The Johns Hopkins University Press).
- Ballington, J.R. (2001). Collection, utilization and preservation of genetic resources of *Vaccinium*. *Hortic. Sci.* 36, 213-220.
- Barnes, D.K., Bingham, E.T., Murphy, R.P., Hunt, O.J., Beard, D.F., Skrdla, W.H., et al. (1977). *Alfalfa Germplasm in the United States: Genetic Vulnerability, Use, Improvement and Maintenance*. Agricultural Research Service Tech. Bull. No. 1571 (Hyattsville: USDA).
- Barney, D., Hummer, K. (2005). *Currants, Gooseberries and Jostaberries - A Guide for Growers, Marketers and Researchers in North America* (Binghamton: Haworth Press).
- Bilz, M., Kell, S.P., Maxted, N., Lansdown, R.V. (2011). *European Red List of Vascular Plants* (Luxembourg: Publications Office of the European Union).
- Bossdorf, O., Auge, H., Lafuma, L., Rogers, W.E., Siemann, E., Prati, D. (2005). Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144, 1-11.
- Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J.A., Mozzherin, D., Rees, T., et al. (2013). The taxonomic name resolution service: An online tool for automated standardization of plant names. *BMC Bioinform.* 14, 16.
- Bradshaw, J.E., Bryan, G.J., Ramsay, G. (2006). Genetic resources (including wild and cultivated *Solanum* species) and progress in their utilization in potato breeding. *Potato Res.* 49, 49-65.
- Brennan, R. (2008). "Currants and gooseberries," in: *Temperate Fruit Crop Breeding: Germplasm to Genomics*, ed. J. Hancock (New York: Springer Science and Business Media B.V.), 177-196.
- Bretting, P.K., Nabhan, G.P. (1986). Ethnobotany of devil's claw (*Proboscidea parviflora* ssp. *parviflora*: Martyniaceae) in the Greater Southwest. *J. California and Great Basin Anthropology* 8, 226-237.
- Brummitt, N., Bachman, S. (2010). *Plants Under Pressure: A Global Assessment. The First Report of the IUCN Sampled Red List Index for Plants* (London: Natural History Museum, UK).
- Bureau of Land Management. (2012). Seeds of success. U.S. Department of the Interior, Bureau of Land Management, Washington, DC. <http://www.nps.gov/plants/sos/> (accessed 1 Aug. 2012).
- Center for Plant Conservation (CPC). (2012). Center for Plant Conservation: Conserving and restoring America's native plants. Center for Plant Conservation, St. Louis, MO. <http://www.centerforplantconservation.org/> (accessed 1 Aug. 2012).
- Chuda, A., Adamus, A. (2009). Aspects of interspecific hybridization within edible Alliaceae. *Acta Physiologiae Plantarum* 31, 223-227.
- Cox, T.S., Sears, R.G., Bequette, R.K., Martin, T.J. (1995). Germplasm enhancement in winter wheat x *Triticum tauschii* backcross populations. *Crop Sci.* 35, 913-919.
- Eenink, A.H., Groenwold, R., Dieleman, F.L. (1982). Resistance of lettuce (*Lactuca*) to the leaf aphid *Nasonovia ribisnigri*: transfer of resistance from *L. virosa* to *L. sativa* by interspecific crosses and selection of resistant breeding lines. *Euphytica* 31, 291-300.
- El Khlifi, O.K., Sharma, H., Benlhabib, O. (2004). Transfer of Hessian fly resistance through interspecific crosses between cultivated wheat and four *Aegilops* species. *Plant Gen. Resour. Newsletter* 138, 43-49.

- FAO (Food and Agriculture Organization of the United Nations). (2002). *International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2010). *Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2011). FAOSTAT (Rome: FAO). <http://faostat.fao.org/> (accessed 1 May 2011).
- Farooq, S., Asghar, M., Iqbal, N., Asian, E., Arif, M., Shah, T.M. (1995). Production of salt tolerant wheat germplasm through crossing cultivated wheat with *Aegilops cylindrica*, I. Field evaluation of salt tolerant germplasm. *Cereal Res. Commun.* 23, 275-282.
- Finn, C. (2001). Trailing blackberries: from clear-cuts to your table. *Hortic. Sci.* 36, 236-238.
- Flack, J.R. (1970). *The spread and domestication of the pecan (Carya illinoensis) in the United States* (Madison: University of Wisconsin Press).
- Flora of North America Association (FNA). (2008a). *Flora of North America* (Cambridge: Flora of North America Association, Harvard University Herbaria) <http://floranorthamerica.org/> (accessed 1 Jan. 2011).
- Flora of North America Association (FNA). (2008b). *Flora of North America – Introduction*. (Cambridge: Flora of North America Association, Harvard University Herbaria). <http://floranorthamerica.org/introduction> (accessed 1 Jan. 2011).
- Ford-Lloyd, B., Schmidt, M., Armstrong, S.J., Barazani, O., Engels, J., Hadas, R., *et al.* (2011). Crop wild relatives- undervalued, underutilized and under threat? *BioScience* 61, 559-565.
- Fowler, C., Hodgkin, T. (2004). Plant genetic resources for food and agriculture: assessing global availability. *Annu. Rev. Env. Resour.* 29, 10.1-10.37.
- Frese, L., Desprez, B., Ziegler, D. (2001). "Potential of genetic resources and breeding strategies for base-broadening in *Beta*," in: *Broadening the Genetic Base of Crop Production*, eds. H.D. Cooper, C. Spillane, T. Hodgkin (Wallingford: CABI), 295-309.
- Gale, G. (2003). "Saving the vine from *Phylloxera*: a never-ending battle," in: *Wine: A Scientific Exploration*, eds. M. Sandler, R. Pinder (London & New York: Taylor & Francis).
- GBIF (Global Biodiversity Information Facility). (2012). *GBIF Data Portal* (Copenhagen: GBIF). <http://data.gbif.org/welcome.htm> (accessed 1 Jan. 2011).
- Global Strategy for Plant Conservation (GSPC). (2002). *Global Strategy for Plant Conservation (GSPC)*, Convention on Biological Diversity (Montreal: Secretariat of the Convention on Biological Diversity).
- Goodman, R.M., Hauptli, H., Crossway, A., Knauf, V.C. (1987). Gene transfer in crop improvement. *Science* 236, 48-54.
- Gororo, N.N., Eagles, H.A., Eastwood, R.F., Nicolas, M.E., Flood, R.G. (2002). Use of *Triticum tauschii* to improve yield of wheat in low-yielding environments. *Euphytica* 123, 241-254.
- GRIN (Germplasm Resources Information Network), USDA, ARS, National Genetic Resources Program. (2012). National Germplasm Resources Laboratory, Beltsville, Maryland. <http://www.ars-grin.gov/~sbmljw/cgi-bin/index.pl?language=en> (Accessed January-August 2012)
- GRIN (Germplasm Resources Information Network), USDA, ARS, National Genetic Resources Program. (2012). *World Economic Plants Database*. <http://www.ars-grin.gov/cgi-bin/npgs/html/wep.pl> (accessed Jan. 2011)
- Groombridge, B., Jenkins, M.D. (2002). *World Atlas of Biodiversity*. Prepared by the UNEP World Conservation Monitoring Centre (Berkeley: University of California Press).
- Guarino, L., Lobell, D.B. (2011). A walk on the wild side. *Nat. Clim. Chang.* 1 (8), 374-375.
- Guo, Q., Falcone, J., Brownsmith, J. (2009). "Building the database for introduced plants in the United States," in: *Proceedings 20th U.S. Department of Agriculture Interagency Research Forum on*

- Invasive Species 2009*, Gen. Tech. Rep. NRS-P-51, 2009 January 13-16, eds. K.A. McManus. K.W. Gottschalk (Annapolis: U.S. Department of Agriculture, Forest Service, Northern Research Station), 73.
- Gur, A., Zamir, D. (2004). Unused natural variation can lift yield barriers in plant breeding. *PLoS Biol.* 2(10), e245.
- Hajjar, R., Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1-13.
- Hanna, W.W. (1997). Influence of cytoplasm from a wild grassy subspecies on dry matter yields in pearl millet. *Crop Sci.* 37, 614-616.
- Harlan, J.R. (1976). Genetic resources in wild relatives of crops. *Crop Sci.* 16, 329-333.
- Harlan, J.R., de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20, 509-517.
- Heywood, V.H. (2008). Challenges of *in situ* conservation of crop wild relatives. *Turk. J. Bot.* 32, 421-432.
- Hijmans, R.J., Spooner, D. (2001). Geographic distribution of wild potato species. *Am. J. Bot.* 88, 2101-2112.
- Hoes, J.A., Putt, E.D., Enns, H. (1973). Resistance to *Verticillium* wilt in collections of wild *Helianthus* in North America. *Phytopathology* 63, 1517-1520.
- Hoffman D.L, Chong, J., Jackson, E.W., Obert, D.E. (2006). Characterization and mapping of a crown rust resistance gene complex (Pc58) in TAM O-301. *Crop Sci.* 46, 2630-2635.
- Hooftman, D.A.P., De Jong, M.J., Oostermeijer, G.B., den Nijs, H.J.C.M. (2007). TI: Modelling the long-term consequences of crop x wild relative hybridization: a case study using four generations of hybrids. *J. Appl. Ecol.* 44(5), 1035-1045.
- Hulke, B.S., Miller, J.F., Gulya, T.J., Vick, B.A. (2010). Registration of the oilseed sunflower genetic stocks HA 458, HA 459, and HA 460 possessing genes for resistance to downy mildew. *J. Plant Registrations* 4, 1-5.
- Hummer, K.E., Bassil, N., Njuguna, W. (2011). "Fragaria," in: *Wild Crop Relatives: Genomic and Breeding Resources*, ed. C. Kole (Berlin: Springer).
- Iltis, H.H. (1988). "Serendipity in the exploration of biodiversity. what good are weedy tomatoes?," in: *Biodiversity*, ed. E.O. Wilson (Washington D.C.: National Academy Press), 98-105.
- International Union for Conservation of Nature (IUCN), Species Survival Commission. (2008). Crop wild relative specialist group. Bioersivity International, Rome, Italy. <http://www.cwrsg.org/index.asp> (accessed 1 Jan. 2012).
- International Union for Conservation of Nature (IUCN). (2012). *The IUCN Red List of Threatened Species*. Version 2012.1 (Cambridge: International Union for Conservation of Nature and Natural Resources). <http://www.iucnredlist.org/> (accessed 1 Oct. 2012).
- Jan C.C, Fernandez-Martinez, J.M., Ruso, J., Muñoz-Ruz, J. (2002). Registration of four sunflower germplasms with resistance to *Orobanche cumana* race F. *Crop Sci.* 42, 2217-2218.
- Jan, C.C. (1992). Registration of sunflower amphiploid germplasm line, ANN-BOL-AMP1. *Crop Sci.* 32, 1513.
- Jan, C.C., Chandler, J.M. (1988). Registration of powdery mildew resistant sunflower germplasm pool, PM 1. *Crop Sci.* 28, 1040.
- Jan, C.C., Miller, J.F., Seiler, G.J., Fick, G.N. (2006). Registration of one cytoplasmic male sterile and two fertility restoration sunflower genetic stocks. *Crop Sci.* 46, 1835.
- Jan, C.C., Quresh, Z., Gulya, T.J. (2004). Registration of seven rust resistant sunflower germplasms. *Crop Sci.* 44, 1887-1888.
- Jarvis, A., Lane, A., Hijmans, R.J. (2008). The effect of climate change on crop wild relatives. *Agr.*

- Ecosyst. Environ.* 126, 13-23.
- Jarvis, A., Williams, K., Williams, D.E., Guarino, L., Caballero, P.J., Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet. Resour. Crop Ev.* 52, 671-682.
- Kell, S.P., Knüpffer, H., Jury, S.L., Ford-Lloyd, B.V., Maxted, N. (2008). "Crops and wild relatives of the Euro-Mediterranean region: making and using a conservation catalogue," in: *Crop Wild Relative Conservation and Use*, eds. N. Maxted, B.V. Ford-Lloyd, S.P. Kell, J.M. Iriondo, M.E. Dulloo, J. Turok (Wallingford: CAB International), 69-109.
- Khrustaleva, L., Kik, C. (1998). Cytogenetical studies in the bridge cross *Allium cepa* x (*A. fistulosum* x *A. roylei*). *Theor. Appl. Genet.* 96, 8-14.
- Kraft, K.H., Luna-Ruiz, J.J., Gepts, P. (2012). A new collection of wild populations of *Capsicum* in Mexico and the southern United States. *Genet. Resour. Crop Ev.* 60, 225-232.
- Lexer, C., Lai, Z., Rieseberg, L.H. (2004). Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *New Phytol.* 161, 225-233.
- Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B., Ackerly, D.D. (2009). The velocity of climate change. *Nature* 462 (7276), 1052-1055.
- Ma, H., Singh, R.P., Mujeeb-Kazi, A. (1995). Resistance to stripe rust in *O. rufipogon*, *T. tauschii* and their synthetic hexaploids. *Euphytica* 82, 117-124.
- Magos Brehm, J., Maxted, N., Ford-Lloyd, B.V., Martins-Loução, M.A. (2007). National inventories of crop wild relatives and wild harvested plants: case-study for Portugal. *Genet. Resour. Crop Ev.* 55, 779-796.
- Mallikarjuna, N., Deepak, J., Prabhakar, R. (2006). Introgression of *Cajanus platycarpus* genome into cultivated pigeonpea, *C. cajan*. *Euphytica* 149, 161-167.
- Maxted, N., Akparov, Z.I., Aronsson, M., Asdal, Å., Avagyan, A., et al. (2012b). "Current and future threats and opportunities facing European crop wild relative and landrace diversity," in: *Agrobiodiversity Conservation: Securing the Diversity of Crop Wild Relatives and Landraces*, eds. N. Maxted, M.E. Dulloo, B.V. Ford-Lloyd, L. Frese, J.M. Iriondo, et al. (Wallingford: CAB International), 333-354.
- Maxted, N., Dulloo, E., Ford-Lloyd, B.V., Iriondo, J.M., Jarvis, A. (2008). Gap analysis: a tool for complementary genetic conservation assessment. *Divers. Distrib.* 14, 1018-1030.
- Maxted N., Ford-Lloyd, B.V., Hawkes, J.G. (1997). "Complementary conservation strategies," in: *Plant Genetic Conservation: The In situ Approach*, eds. N. Maxted, B.V. Ford-Lloyd, J.G. Hawkes (London: Chapman and Hall), 20-55.
- Maxted, N., Ford-Lloyd, B.V., Jury, S.L., Kell, S.P., Scholten, M.A. (2006). Towards a definition of a crop wild relative. *Biodivers. Conserv.* 15, 2673-2685.
- Maxted, N., Kell, S.P. (2009). *Establishment of a Global Network for the In Situ Conservation of Crop Wild Relatives: Status and Needs* (Rome: FAO Commission on Genetic Resources for Food and Agriculture).
- Maxted, N., Kell, S.P., Ford-Lloyd, B.V., Dulloo, E., Toledo, Á. (2012a). Toward the systematic conservation of global crop wild relative diversity. *Crop Sci.* 52, 774-785.
- McCouch, S., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septiningsih, E., et al. (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait enhancing alleles for *O. sativa*. *Euphytica* 154, 317-339.
- McGranahan, G., Leslie, C. (2009). "Breeding walnuts (*Juglans regia*)," in: *Breeding Plantation Tree Crops: Temperate Species*, eds. S.M. Jain, P.M. Priyadarshan (New York: Springer), 249-273.
- McGuffin, M., American Herbal Products Association. (2000). *Herbs of Commerce. 2nd ed* (Silver

- Spring: American Herbal Products Association).
- Meilleur, B.A., Hodgkin, T. (2004). *In situ* conservation of crop wild relatives: status and trends. *Biodivers. Conserv.* 13, 663-684.
- Mii, M. (2009). Breeding ornamental plants through interspecific hybridization using advanced techniques with a special focus on *Dianthus*, *Primula*, *Cosmos*, and *Kalanchoe*. *Acta Hortic.* 836, 63-72.
- Miller, J. F., Gulya, T.J. (1988). Registration of six downy mildew resistant sunflower germplasm lines. *Crop Sci.* 28, 1040-1041.
- Miller, J.F., Gulya, T.J. (1999). Registration of eight Sclerotinia-tolerant sunflower germplasm lines. *Crop Sci.* 39, 301-302.
- Nabhan, G.P. (1985). Native crop diversity in Aridoamerica: conservation of regional gene pools. *Econ. Bot.* 39, 387-399.
- Nabhan, G.P. (1990). Conservationist and forest service join forces to save wild chiles. *Diversity Magazine* 6, 47-48.
- Native Seed Network. (2010). *Native Seed Network Database* (Corvallis: Native Seed Network). <http://www.nativeseednetwork.org> (accessed 1 Jan. 2011).
- NatureServe. (2009). *NatureServe Explorer: An Online Encyclopedia of Life. Version 7.1* (Arlington: NatureServe). <http://www.natureserve.org/explorer/> (accessed 1 Apr. 2010).
- Ortiz, R., Sayre, K.D., Govaerts, B., Gupta, R., Subbarao, G.V., Ban, T., *et al.* (2008). Climate change: can wheat beat the heat? *Agr. Ecosyst. Environ.* 126, 46-58.
- Parra-Quijano, M., Iriondo, J.M., Torres, E. (2011). Improving representativeness of genebank collections through species distribution models, gap analysis and ecogeographical maps. *Biodivers. Conserv.* 21, 79-96.
- Pavek, D.S., Lamboy, W.F., Garvey, E.J. (2001). *In situ* conservation of America's wild grapes. *Hortic. Sci.* 36, 232-235.
- Peel, M.D., Asay, K.H., Waldron, B.L., Jensen, K.B., Robins, J.G., Mott, I.W. (2009). 'Don', a diploid falcata alfalfa for western U.S. rangelands. *J. Plant Registrations* 3, 115-118.
- Pena, R.J., Zarco-Hernandez, J., Mujeeb-Kazi, A. (1995). Glutenin subunit compositions and breadmaking quality characteristics of synthetic hexaploid wheats derived from *O. rufipogon* x *Triticum tauschii* (Coss.) Schmal crosses. *J. Cereal Sci.* 21, 15-23.
- Peredo, E.L., Ángeles Revilla, M., Reed, B.M., Javornik, B., Cires, E., Fernández Prieto, J.A., *et al.* (2010). The influence of European and American wild germplasm in hop (*Humulus lupulus* L.) cultivars. *Genet. Resour. Crop Ev.* 57, 575-586.
- Pestsova, E.G., Borner, A., Roder, M.S. (2006). Development and QTL assessment of *Triticum aestivum*-*Aegilops tauschii* introgression lines. *Theor. Appl. Genet.* 112, 634-647.
- Phillips, O.L., Meilleur, B.A. (1998). Usefulness and economic potential of the rare plants of the United States: a statistical survey. *Economic Bot.* 52(1), 57-67.
- Pickering, R., Johnston, P.A. (2005). Recent progress in barley improvement using wild species of *Hordeum*. *Cytogenet. Genome Res.* 109, 344-349.
- Plant Germplasm Operations Committee. (1999). *The American Wild Relatives of Crops: In Situ Conservation Guidelines* (Beltsville: USDA National Plant Germplasm System, *In situ* subcommittee).
- Plant Germplasm Operations Committee. (2010). *Minutes of the Annual Meeting, Geneva, NY. 28-29 July 2010* (Beltsville: USDA National Germplasm Resources Laboratory).
- Prescott-Allen, R., Prescott-Allen, C. (1986). *The First Resource: Wild Species in the North American Economy* (New Haven: Yale University Press).
- Prescott-Allen, R., Prescott-Allen, C. (1990). How many plants feed the world? *Conserv. Biol.* 4, 365-

374.

- Putt, E.D. (1978). "History and present world status," in: *Sunflower Science and Technology*, ed. J.P. Carter (Madison: American Society of Agronomy), 1-29.
- Ramírez-Villegas J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. *PLoS One* 5(10), e1349.
- Rawal, K.M. (1975). Natural hybridization among wild, weedy and cultivated *Vigna unguiculata* (L.) Walp. *Euphytica* 24, 699-707.
- Reisch, B., Pratt, C. (1996). "Grapes," in: *Fruit Breeding Volume II: Vine and Small Fruit Crops*, eds. J. Janick, J. Moore (New York: John Wiley & Sons), 297-369.
- Rogers, C.E., Thompson, T.E., Seiler, G.J. (1984). Registration of three *Helianthus* germplasms for resistance to the sunflower moth. *Crop Sci.* 24, 212-213.
- Ross, H. (1979). "Wild species and primitive cultivars as ancestors of potato varieties," in: *Proceedings of the Conference Broadening the Genetic Base of Crops*, 3-7 July 1978, eds. A.C. Zeven, A.M. van Harten (Wageningen, Centre for Agricultural Publishing and Documentation), 237-245.
- Ruge-Wehling, B., Linz, A., Habeku, A., Wehling, P. (2006). Mapping of RYML6Hb, the second soilborne virus resistance gene introgressed from *Hordeum bulbosum*. *Theor. Appl. Genet.* 113, 867-673.
- Rygulla, W., Snowdon, R., Eynck, C., Koopmann, B., Von Tiedemann, A., Lühs, W., et al. (2007). Broadening the genetic basis of *Verticillium longisporum* resistance in *Brassica napus* by interspecific hybridization. *Phytopathology* 97, 1391-1396.
- Sakamoto, S. (1976). Breeding of a new sweet potato variety, Minamiyutaka, by the use of wild relatives. *Japanese Agric. Res. Quest.* 10(4), 183-186.
- Seiler, G.J. (1991a). Registration of 15 interspecific sunflower germplasm lines derived from wild annual species. *Crop Sci.* 31, 1389-1390.
- Seiler, G.J. (1991b). Registration of 13 downy mildew tolerant interspecific sunflower germplasm lines derived from wild annual species. *Crop Sci.* 31, 1714-1716.
- Seiler, G.J. (1991c). Registration of six interspecific sunflower germplasm lines derived from wild perennial species. *Crop Sci.* 31, 1097-1098.
- Seiler, G.J. (2000). Registration of 10 interspecific germplasms derived from wild perennial sunflower. *Crop Sci.* 40, 587-588.
- Seiler, G.J., Gulya, T.J. (2004). "Exploration for wild *Helianthus* species in North America: challenges and opportunities in the search for global treasures," in: *International Sunflower Conference Proceedings*, v.1, Fargo, ND. 29 Aug. – 2 Sept. 2004 (Paris: International Sunflower Association), 43-68.
- Smith, B.D. (2006). Eastern North America as an independent center of plant domestication. *Proc. Natl. Acad. Sci.* 103, 12223-12228.
- Stein, B.A., Kutner, L.S., Adams, J.S. (2000). *Precious Heritage: The Status of Biodiversity in the United States* (New York: Oxford University Press).
- Suszkiv, J. (2005). *Hessian Fly Resistant Wheat Germplasm Available* (Washington D.C.: USDA-ARS). <http://www.ars.usda.gov/is/pr/2005/050211.htm> (accessed 1 Aug. 2012).
- Takeda, K., Frey, K.J. (1976). Contributions of vegetative growth rate and harvest index to grain yield of progenies from *Avena sativa* X *A. sterilis* Crosses. *Crop Sci.* 16, 817-821.
- Tanksley, S.D., McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063-1066.
- Thompson, M., Lagerstedt, H., Mehlenbacher, S. (1996). "Hazelnuts," in: *Fruit Breeding, Volume III: Nuts*, eds. J. Janick, J. Moore (New York: John Wiley & Sons), 125-184.

- Toubia-Rahme, H., Johnston, P.A., Pickering, R.A., Steffenson, B.J. (2003). Inheritance and chromosomal location of *Septoria passerinii* resistance introgressed from *Hordeum bulbosum* into *Hordeum vulgare*. *Plant Breeding* 122, 405-409.
- Ureta, C., Martinez-Meyer, E., Perales, H.R., Alvarez-Buylla, E.R. (2011). Projecting the effects of climate change on the distribution of maize races and their wild relatives in Mexico. *Glob. Change Biol.* 18, 1073-1082.
- USDA-ARS National Genetic Resources Program. (2011). *World Economic Plants: A Standard Reference* (Beltsville: National Germplasm Resources Laboratory). <http://www.ars-grin.gov/cgi-bin/npgs/html/wep.pl> (accessed 1 Jan. 2011).
- USDA-ARS National Genetic Resources Program. (2012). *GRIN (Germplasm Resources Information Network) Taxonomy for Plants* (Beltsville: National Germplasm Resources Laboratory). <http://www.ars-grin.gov/~sbmljw/cgi-bin/index.pl?language=en> (accessed 1 Jan. 2012).
- USDA-Natural Resources Conservation Service (USDA-NRCS). (2010). *The PLANTS Database* (Baton Rouge: National Plant Data Center). <http://plants.usda.gov> (accessed 23 Apr. 2010).
- US Forest Service (USFS). (2010). *Celebrating Wildflowers – Ethnobotany – Wild Crop Progenitors* (Washington D.C.: USFS). <http://www.fs.fed.us/wildflowers/ethnobotany/progenitors.shtml> (accessed 1 Aug. 2012).
- Vavilov, N.I. (1926). Studies on the origin of cultivated plants. *Bulletin of Applied Botany, Genetics and Plant Breeding* 16, 1-248.
- Villareal, R.L., Mujeeb-Kazi, A., Fuentes-Davila, G., Rajaram, S. (1996). Registration of four synthetic hexaploid wheat germplasm lines derived from *O. rufipogon* x *T. tauschii* crosses and resistance to kernal bunt. *Crop Sci.* 36, 218.
- Villareal, R.L., Sayre, K., Banuelos, O., Mujeeb-Kazi, A. (2001). Registration of four synthetic hexaploid wheat (*O. rufipogon/Aegilops tauschii*) germplasm lines tolerant to waterlogging. *Crop Sci.* 41, 274.
- Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., *et al.* (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265-75.
- Volk, G.M., Richards, C.M. (2011). Integration of georeferencing, habitat, sampling, and genetic data for documentation of wild plant genetic resources. *Hortic. Sci.* 46 (11), 1446-1449.
- Wang, W., Vinocur, B., Altaian, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1-14.
- Whelan, E.D.P., Dedio, W. (1980). Registration of sunflower germplasm composite crosses CMG-1, CMG-2, and CMG-3. *Crop Sci.* 20, 832.
- Wiersema, J.H., León, B. (1999). *World Economic Plants: A Standard Reference* (Boca Raton: CRC Press LLC).
- Wiersema, J.H., León, B., Garvey, E.J. (2012). Identifying wild relatives of subtropical and temperate fruit and nut crops. *Acta Hortic.* 948, 285-288.
- Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. A modest proposal for teosinte evolution and conservation *in situ*: The Balsas, Guerrero, Mexico. *Maydica* 52, 49-58.
- Xiao J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D. (1996). Genes from wild rice improve yield. *Nature* 384, 223-224.
- Xiu-Jin, L., Deng-Cai, L., Zhi-Rong, W. (1997). Inheritance in synthetic hexaploid wheat ‘RSP’ of sprouting tolerance derived from *Aegilops tauschii* Coss. *Euphytica* 95, 321-323.
- Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Genet.* 2, 983-989.
- Zhang, L., Pickering, R.A., Murray, B.G. (2001). *Hordeum vulgare* x *H. bulbosum* tetraploid hybrid

Chapter 3

- provides useful agronomic introgression lines for breeders. *New Zeal. J. Crop Hort.* 29, 239-246.
- Zizumbo-Villarreal, D., Colunga-GarcíaMarín, P. (2010). Origin of agriculture and plant domestication in west Mesoamerica. *Genet. Resour. Crop Ev.* 57, 813-825.
- Zizumbo-Villarreal, D., Colunga-GarciaMarin, P., Payro de la Cruz, E., Delgado-Valerio, P., Gepts, P. (2005). Population structure and evolutionary dynamics of wild-weedy-domesticated complexes of common bean in a Mesoamerican region. *Crop Sci.* 45, 1073-1083.

CHAPTER 4

A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans

Julián Ramírez-Villegas¹, Colin Khoury², Andy Jarvis^{1,3,4}, Daniel G. Debouck⁵, and
Luigi Guarino²

¹ Decision and Policy Analysis Program, International Center for Tropical Agriculture (CIAT), Cali, Colombia.

² Global Crop Diversity Trust, Rome, Italy.

³ Adaptation to Progressive Climate Change, Challenge Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark.

⁴ Bioversity International, Regional Office for the Americas, Cali, Colombia.

⁵ Genetic Resources Unit, International Center for Tropical Agriculture (CIAT), Cali, Colombia.

Published in *PLoS One* 5 (10), e13497. doi:10.1371/journal.pone.0013497

Abstract

The wild relatives of crops represent a major source of valuable traits for crop improvement. These resources are threatened by habitat destruction, land use changes, and other factors, requiring their urgent collection and long-term availability for research and breeding from *ex situ* collections. We propose a method to identify gaps in *ex situ* collections (i.e., gap analysis) of crop wild relatives as a means to guide efficient and effective collecting activities. The methodology prioritizes among taxa based on a combination of sampling, geographic, and environmental gaps. We apply the gap analysis methodology to wild taxa of the *Phaseolus* genepool. Of 85 taxa, 48 (56.5%) are assigned high priority for collecting due to lack of, or under-representation, in genebanks, 17 taxa are given medium priority for collecting, 15 low priority, and 5 species are assessed as adequately represented in *ex situ* collections. Gap "hotspots", representing priority target areas for collecting, are concentrated in central Mexico, although the narrow endemic nature of a suite of priority species adds a number of specific additional regions to spatial collecting priorities. Results of the gap analysis method mostly align very well with expert opinion of gaps in *ex situ* collections, with only a few exceptions. A more detailed prioritization of taxa and geographic areas for collection can be achieved by including in the analysis predictive threat factors, such as climate change or habitat destruction, or by adding additional prioritization filters, such as the degree of relatedness to cultivated species (i.e., ease of use in crop breeding). Furthermore, results for multiple crop genepools may be overlaid, which would allow a global analysis of gaps in *ex situ* collections of the world's plant genetic resources.

Keywords: Crop wild relative, *Ex situ* conservation, Plant Genetic Resources, Germplasm exploration, Tropical legumes

INTRODUCTION

Crop wild relatives (CWR) are wild plant species sharing relatively recent common ancestry with cultivated plants. CWR typically possess wide diversity, much of it not found in the crop, and this diversity may be introgressed into the crop by plant breeders, with the ease of transfer of genes generally dependent on the degree of relatedness between the wild species and the domesticate (Harlan & de Wet 1971, Singh 2001). Wild relatives have provided to crops traits such as pest and disease resistance, tolerance to abiotic stresses, increased yield, male sterility, and quality, increasing the value and sustainability of banana, barley, beans, cassava, chickpea, lettuce, maize, oats, pearl millet, potatoes, rice, sugar cane, sunflower, tomato, and wheat production, among others. In the past 20 years, there has been a steady increase in the rate of release of cultivars containing genes from CWR, and their contribution should only increase as the development of molecular technologies makes identification and utilization of diverse germplasm more efficient (Prescott-Allen & Prescott-Allen 1986, Tanksley & McCouch 1997, Singh 2001, Hajjar & Hodgkin 2007).

Plant breeders obtain CWR material from genebanks. However, major gaps in the genetic diversity of important crop gene pools remain to be filled in *ex situ* germplasm collections. These gaps are particularly evident for non-cereal crops (e.g., legumes, roots and tubers, vegetables), and for wild and weedy forms (FAO 1997, Maxted & Kell 2009, Khoury *et al.* 2010). Maxted and Kell (2009) estimated that 94% of European CWR species are completely missing from *ex situ* collections. At the same time, habitat destruction, invasive species, urbanization, and the shift from traditional to industrial agricultural practices, among other factors, continue to threaten PGR, and climate change is projected to impose further pressures on both wild and agricultural ecosystems (Meilleur & Hodgkin 2004, Thuiller *et al.* 2005, Brooks *et al.* 2006, Challinor *et al.* 2007, Intergovernmental Panel on Climate Change 2007, Hawkins *et al.* 2008, Burke *et al.* 2009).

Clearly, much collecting of CWR diversity is still required. Unfortunately international efforts in collecting plant genetic resources in general have been in decline in recent decades (FAO 2009b). The recent coming into force of the International Treaty on Plant Genetic Resources for Food and Agriculture is, however, expected to provide impetus for the development of an integrated, effective, efficient, global approach to conserving PGR. The development of strategic planning approaches will be necessary to prioritize PGR for collecting as part of such

a rational global system.

Gap analysis refers to a systematic method of analyzing the degree of conservation of taxa, in order to identify those locations, taxa, and particular traits (adaptations) un- or under- secured in conservation systems (Maxted *et al.* 2008). Nabhan (1990) identified four ways by which gap analysis techniques may lead to better collecting and conservation: targeting localities where sets of species absent from existing collections can be obtained with least effort and cost; determining which areas are ‘under-collected’ or ‘over-collected’ for germplasm relative to the known distribution of a taxon; locating which regions have the greatest or most dissimilar species richness compared with other regions; and outlining the ecological amplitudes of each species so that a wider representation of the ecotypes or genetically adapted populations of each can be sampled.

Geographic Information Systems (GIS) technologies have enabled a better understanding of species distributions and of the representativeness of germplasm collections, and have contributed to conservation planning of wild species, CWR, and domesticates (Nabhan 1990, Jones *et al.* 1997, Hijmans & Spooner 2001, Hijmans *et al.* 2002, Jarvis *et al.* 2002, Jarvis *et al.* 2003, Lobo Burle *et al.* 2003, Maxted *et al.* 2004, Jarvis *et al.* 2005, Graham & Hijmans 2006, Parthasarathy *et al.* 2006, Maxted *et al.* 2008). Pioneering the use of these tools in conservation, Jones *et al.* (1997) successfully predicted the location of populations of wild common bean (*Phaseolus vulgaris*), based on climatic suitability. Significant developments have occurred in recent years in the application of GIS to PGR conservation planning, including the development and validation of various approaches to niche modeling, new analysis tools and extensions, and better access to geographic information, results and approaches (Hijmans *et al.* 2001).

We propose here a gap analysis method designed to inform planning of germplasm collecting for *ex situ* conservation, based upon available information resources, using GIS. The distributions of *ex situ* collections are compared to GIS-modeled taxon distributions based on both herbarium and genebank data. The gross total number of germplasm accessions, as well as the distribution (geographical and environmental) of those accessions, are compared against modeled distributions in order to identify gaps in *ex situ* conservation coverage. These results form the basis for a prioritization of taxa across the genepool for collecting, and the

identification of the highest priority locations (i.e., diverse and under-represented areas) for the most efficient and effective collecting, in order to further enhance *ex situ* holdings. Our model genepool is *Phaseolus*.

The genus *Phaseolus* originated in the tropics and subtropics of the New World, and contains up to 81 species and 34 infraspecific taxa (Gepts 2001, Singh 2001, Freytag & Debouck 2002, Debouck 2009), having undergone a series of revisions, notably in association with members of *Vigna*, which have included splitting some species into new genera (e.g., *Strophostyles*, *Dysolobium*, *Macroptilium*, *Minkelersia* and *Alepidocalyx*) (Maxted *et al.* 2004). The main centers of diversity for the genus are in wide Mesoamerica (from southern USA, Mexico, and Central America down to Panama), the northern Andean region (Colombia to northern Peru), and the central Andes (northern Peru, Bolivia to northwest Argentina). Of these, the Mesoamerican centre is the richest in species (Delgado-Salinas 1985, Debouck 2000, Freytag & Debouck 2002, Debouck 2009).

Phaseolus has five domesticated species, each a result of an independent domestication process: *P. vulgaris* L.- common bean; *P. lunatus* L.- lima bean; *P. coccineus* L.- runner bean; *P. acutifolius* A. Gray - tepary bean; and *P. dumosus* Macfady - year bean. The genus has been cultivated for over 7000 years, and each of the cultivated species has distinct ecological adaptations (Debouck & Smartt 1995). Common bean is the world's most important legume for food production and security, and represents 50% of the grain legumes consumed worldwide, reaching primary importance in the staple diet of over 500 million people, especially for its protein content (Gepts 2001, McClean *et al.* 2004). Common bean is now grown on over 27 million hectares globally, producing over 20 million tons (FAO 2009a).

Diversity in *Phaseolus* in relation to the cultivated species is organized into genepools based on phylogenetic relationships (Smartt 1981, Singh & Jauhar 2005). The primary genepool of cultivated species includes both cultivars and wild populations, hybrids of which are generally fully fertile with no major reproductive barriers. *P. vulgaris* also allows a measure of interspecific hybridization with species in its secondary genepool. *P. lunatus* and *P. acutifolius* appear less capable of gene exchange with related species (Debouck 1999).

Like many important food crops, cultivars of common bean have a narrow genetic base,

attributable to the genetic bottleneck accompanying the domestication process, stringent quality requirements in the market, limited past use of exotic germplasm in breeding, and conservative breeding programs for the crop (Singh 2001). Interspecific and wide intraspecific crossing have been useful strategies for crop improvement, but given the still limited genetic base, more along these lines is needed. Useful alleles for many agronomic traits deficient in common bean cultivars, including resistance to storage insects, leafhoppers, ascochyta blight, common bacterial blight, white mold, bean common mosaic virus, rust, drought, and soil fertility problems, as well as early maturity, adaptation to higher latitudes, upright plant type, pod quality, and seed yield have been identified in wild common bean and species in the secondary and tertiary genepools, and utilized in breeding programs (Kornegay & Cardona 1991, Delgado-Salinas *et al.* 1999, Singh 2001, Acosta-Gallegos *et al.* 2007). Wild common bean has also contributed high protein digestibility (Shelley-Dessert & Bliss 1991) and nodulation (Kipe-Nolt *et al.* 1992) traits. Despite the increasing utilization of CWR in common bean breeding, Singh (2001) estimated as much as 90% of the genetic variability available in the primary genepool and related species as under- or not utilized. Widening of genetic diversity in the other *Phaseolus* crop species may also prove important. The domestication of tepary bean involved a severe genetic bottleneck event, leading to a particularly low level of genetic diversity in the crop (Schinkel & Gepts 1988, Garvin & Weeken 1994, Munoz *et al.* 2006).

Close to 250 *ex situ* germplasm collections of *Phaseolus*, holding approximately 260,000 accessions, have been established worldwide (FAO 2009b). The vast majority of these accessions are of common bean, with much smaller collections of the other cultivated species, and a small percentage of wild species. The largest collections of CWR of *Phaseolus* are held in the international collection managed by the Consultative Group on International Agricultural Research (CGIAR), with close to 2000 accessions (The Consultative Group on International Agricultural Research [CGIAR], System-wide Information network for Genetic Resources [SINGER] 2009) and in the United States National Genetic Resources Program, with close to 500 accessions (USDA, ARS, National Genetic Resources Program [NGRP] 2009).

MATERIALS AND METHODS

An eight-step gap analysis process is presented, which attempts to evaluate conservation deficiencies at three different levels: (1) taxonomic, (2) geographic and (3) environmental. The aim is to define the extent to which current genebank holdings represent total genetic diversity

within a genepool. We apply the protocol to all the wild members of the genus *Phaseolus*.

Based upon the average of overall taxonomic, geographic, and environmental coverage factors, the method produces a table outlining the high, medium and low priority species for collecting. From this table, potential collecting areas for high priority species may be highlighted, and overlapping high priority regions for the collection of multiple taxa identified. In detail, the method is as follows:

Determination of target taxa, delineation of target area and harvesting of occurrence data

This involves five steps:

1. Identification of the target cultivated species.
2. Taxonomical review of all CWR related to the cultivated species, and analysis of relatedness to the domesticated species using the concept established by Maxted *et al.* (2006).
3. Creation of a database containing as many records as possible both of genebank accessions and herbarium specimens, along with (when available) their respective passport data, specifically the names of the places of collection and coordinates (i.e., latitude and longitude). Samples listed as weedy or cultivated are not included in the database.
4. Cross-check, verification, and correction of geographic references (coordinates) through thorough review of data and use of verification tools such as BioGeomancer (www.biogeomancer.org) (Guralnick *et al.* 2006), Google Earth, and high detail physical maps of localities, and strict selection only of verified geo-referenced samples for distribution modeling, as the quality of location data strongly affects the performance of niche modeling techniques (Graham *et al.* 2008).
5. Determine target area for the gap analysis: based upon the native (wild) distribution of the target taxa. Depending on the genepool, the area can range from a small region within a country to the entire world.

Determination of sampling deficiencies at the taxon level

A gross representativeness of genebank accessions for each taxon is calculated using the ‘sampling representativeness score’ (SRS, Equation 1), comparing total germplasm accessions to herbarium records.

$$SRS = \frac{GS}{GS + HS} * 10 \quad (\text{Equation 1})$$

SRS is calculated as the number of germplasm samples (GS) divided by the total number of samples, i.e., the sum of germplasm plus herbarium samples (HS), regardless of whether samples contain location data. SRS therefore permits a general estimation of adequacy of germplasm holdings of each taxon based upon all available data. In the case that a taxon has no genebank samples, it is listed as a “high priority species” for collecting by setting the FPS (see step 7 below) to 0.

In the rare case that for a particular taxon there is obviously deficient herbarium sample data in comparison to germplasm records, the analysis should eliminate SRS as an input for that taxon, as its inclusion would overestimate adequacy of conservation. Mapping of herbarium samples and genebank accessions can be performed (e.g., using DIVA-GIS (version 7.1.70) (Hijmans *et al.* 2001, Hijmans *et al.* 2005b) in order to provide a general geographic assessment of the available data.

Create potential distribution models for taxa

Potential distributions of taxa are calculated using the maximum entropy (Maxent) model (Phillips *et al.* 2006), with a set of bioclimatic variables and species presence data as inputs. We do not consider the total number of samples with coordinates but the number of different populations represented by those samples (unique locations) (Hernandez *et al.* 2006, Phillips *et al.* 2006, Loiselle *et al.* 2008, Costa *et al.* 2010). We use Maxent due to its precision and confidence when predicting species distributions (Dormann 2006, Elith *et al.* 2006, Hernandez *et al.* 2006, Hijmans & Graham 2006, Phillips *et al.* 2006, Loiselle *et al.* 2008, Phillips & Dudik 2008, Costa *et al.* 2010). Default features are used in Maxent, in which complexity of the models (represented by the number of terms and the type of interactions between environmental variables) depend upon the number of input data points (Phillips *et al.* 2006, Phillips & Dudik 2008). Background points for model training equal 10,000 random points over the distributional range of the genepool in order to avoid overfitting (Phillips 2008, VanDerWal *et al.* 2009).

As the Maxent distribution is generally broader than the real distribution of the species, the modeled distribution is further refined by selecting only known native areas and high

probability zones, which generally are defined as the most climatically suitable for the taxon, thus avoiding over-estimation of the realized niche (VanDerWal *et al.* 2009, Smolik *et al.* 2010). The potential distribution is limited to the native area reported in the literature and then thresholded using the ROC (receiver operating characteristic) curve plot-based approach (point on the ROC curve [sensitivity vs. 1-specificity] which has the shortest distance to the top-left corner [0,1] in the ROC plot) (Liu *et al.* 2005, Phillips *et al.* 2006, Phillips & Dudik 2008). We use this threshold as it provides a decent omission rate, is taxon-specific and shows better performance than other thresholds when predicting potential presence (Liu *et al.* 2005). We call this thresholded modeled distribution the “potential distribution coverage”.

Based on the above, for each taxon, we report three model performance metrics: (1) the 25-fold average area under the ROC curve (AUC) (Liu *et al.* 2005, Elith *et al.* 2006, Phillips *et al.* 2006, Phillips 2008) of test data (ATAUC), (2) the standard deviation of the test AUC of the 25 different folds (STAUC), and (3) the proportion of the potential distribution coverage with standard deviation above 0.15 (ASD15). Maxent models with ATAUC above 0.7 (Smolik *et al.* 2010), STAUC below 0.15, and ASD15 below 10% can be considered “accurate and stable” and are thus used in further calculations. We use three measures of model accuracy as the use of AUC alone might mislead the interpretation given the sensitivity of this measure to spatial autocorrelation (Lobo *et al.* 2008, Veloz 2009).

For those taxa for which the Maxent model training fails or is inaccurate or unstable, we assign a priority to the taxa using the following criteria:

1. As with step (2), taxa with no genebank samples are listed as “high priority species” for collecting by setting the FPS (see step 5 below) to 0.
2. Taxa with genebank samples but no herbarium samples with verified location data are listed as “high priority species” for collecting, as more data are needed in order to perform the analysis. Taxa with such paucity of herbarium records are likely to also have limited germplasm conserved, and are therefore very likely to be “high priority species”. However, these taxa might differ from taxa in (3a) since they already have at least one genebank accession, which certainly permits some type of analyses (e.g., genetic diversity). These taxa are thus differentiated from taxa in (3a) by a flag in the final priorities table (see results).
3. Taxa with genebank samples and one or more herbarium samples with verified location

data are assessed using the area of the convex hull around all known populations (unique locations) of the taxon in lieu of potential distribution coverage. We use the convex hull since, particularly for taxa with very limited occurrence data, it provides a polygon resembling the type of area produced by the Maxent distribution model.

At this point, the potential distribution coverage for all taxa (for which a niche model is possible) may be mapped together in order to display the distribution of the genus, and a richness map along with an uncertainty map (i.e., maximum standard deviation of probabilities among the species that are present in each pixel) for the genepool may be calculated from the results.

Geographic coverage assessment

The adequacy of geographic coverage of genebank accessions is calculated as a ‘geographic representativeness score’ (GRS, Equation 2), assessed by comparing the taxon potential distribution coverage with the genebank samples geographic coverage, modeled using the ‘circular area statistic with a 50 km radius’ (CA50) value (Hijmans *et al.* 2001).

$$GRS = \frac{GCG}{PDC} * 10 \quad (\text{Equation 2})$$

GRS is thus the geographic coverage of germplasm collections (*GCG*) divided by the potential distribution coverage of the taxon under analysis (*PDC*). The higher the GRS, the higher the representativeness of genebank collections in relation to the potential distribution of the taxon.

Determination of environmental gaps

The adequacy of environmental coverage of genebank accessions is calculated as an ‘environmental representativeness score’ (ERS, Equation 3), assessed by comparing the germplasm samples in relation to the full environmental range of the modeled taxon distribution. The same set of climatic layers used for developing the potential distribution coverage are standardized to have an average of zero and a standard deviation of 1 in order to perform a principal components analysis. The first two of these spatially explicit components (which normally account for more than 70% of the spatial variability) are reclassified into twenty equal classes.

$$ERS = \sum_{i=1}^2 w_i \frac{EC_i}{PED_i} * 10 \quad (\text{Equation 3})$$

For these two principal components ($i = 2$), ERS is calculated as the environmental coverage (i.e., number of different classes) of germplasm collections (EC) divided by the potential environmental coverage of the taxon under analysis (PED), times the weight (w) of the principal component (weights of the two components are re-scaled so that the sum of their weights is 1). If the total variation explained by the first two components is too small (i.e., less than 70%), additional components can be included in the analysis, and should be weighted accordingly.

Rarity of each species based on environmental variables determination

All records for the gene pool (i.e., GS + HS for all taxa combined) are plotted against a specific environmental variable or linear combination of variables (i.e., vector or principal component) to identify taxa with records falling in rare environmental classes (i.e., extremes of the distribution). We assume that the frequency of the data presents a normal distribution and ‘environmentally rare’ taxa are those located in sites where extreme environmental conditions are found (tails of the distribution - 5th [NS_{P5}] and 95th [NS_{P95}] percentiles). A numeric value (environmentally rare taxa score, ERTS, Equation 4) is calculated for each taxon as the number of populations in rare environments divided by the total number of populations of that taxon.

$$ERTS = \frac{NS_{P5} + NS_{P95}}{GS + HS} * 10 \quad (\text{Equation 4})$$

As this step of the gap analysis should be conducted only when there is sufficient data for all the taxa under analysis in order to avoid bias in the results (an abundant number of populations so that a histogram can be calculated), usually it will not be included in the overall assessment. We suggest that this step can be usefully included for the assessment of a specific subset of well-sampled species.

Numeric assessment to determine the priority of collecting for *ex situ* conservation

All level-specific representativeness scores (SRS, GRS, ERS, and if possible ERTS) are averaged with equal weight to obtain a final score of prioritization of species. The ‘final priority score’ (FPS), is then used to classify taxa according to the following ranges: (1) as high priority

species if the FPS is between 0 and 3, (2) as medium priority species if the FPS is between 3.01 and 5, (3) as low priority species if the FPS is between 5.01 and 7.5, and (4) as well conserved species (no need for further collection) if the FPS is between 7.51 and 10. All taxa flagged as high priority in steps (2) and (3) are included in the list of high priority taxa to be further collected.

Prioritization of geographic areas for collecting germplasm

The potential collection zones for each high priority species are identified separately and then combined to highlight those zones where gaps for multiple species overlap (“collection gap richness”). This is done through the following steps:

1. Identify un-collected zones for each taxon by comparing the potential distribution coverage with the current geographic coverage of germplasm collections (CA50). Areas where the taxon is potentially present but already sampled are dismissed at this stage; the remaining areas are highlighted as uncollected.
2. Four products treating all mappable high priority taxa are finally produced: (1) individual maps showing potential collecting zones of all high priority taxa, (2) a map of collection gap richness: the number of different taxa that can be collected in each 2.5 arc-minutes (~5 km at the Equator) grid cell, (3) a map showing the maximum standard deviation of high priority taxa (derived from the 25-fold Maxent model training procedure) in each pixel, and (4) a map of the maximum distance of each pixel to the nearest accession (this calculation is done taxon-by-taxon and then aggregated into a single map output, by calculating the maximum of all ‘high priority taxa’.

Testing the gap analysis methodology

The methodology relies on available data and utilizes modeling tools, and is therefore vulnerable to the quantity and quality of input data and the limitations of the modeling applied. In order to test the quality of the results, we have compared them to expert opinion, as following:

1. Identify one or more experts on the target taxa (i.e., genepool)
2. Query the selected expert(s) to provide
 - a. A ranking of taxa for importance for conservation: To achieve this, the list of taxon names under analysis is sent to the expert(s), who is asked to provide a rating from 1 to 10 for each taxon (where 1 corresponds to a very high priority [i.e., an incomplete collection], and 10 corresponds to the lowest priority [i.e., a complete collection]),

without having seen the results of the gap analysis. The expert is requested to rate taxa strictly on the basis of adequacy of *ex situ* holdings for the taxon.

- b. The expert is then shown the results of the analysis and is asked to give general comments on the validity of the taxa and geographic prioritizations.
3. Compare the expert and method-based prioritization of each taxon using the relative difference (RD) between the expert priority score (EPS) and the gap analysis FPS, with respect to the total maximum possible difference (Equation 5).

$$RD = \frac{FPS - EPS}{10} * 100 \quad (\text{Equation 5})$$

RD is calculated for each taxon and the number of taxa with very similar ratings ($-30\% < RD < 30\%$), the number of taxa somewhat similar ratings ($-50\% < RD < 50\%$), and the number of taxa with very different ratings ($RD < -70\%$ and $RD > 70\%$) are then counted. We also plot the FPS and the EPS in a scattergram and calculate both the Spearman correlation coefficient and the P-value of the Spearman correlation coefficient. With these metrics, we aim to provide a general evaluation of the gap analysis method in identifying high priority taxa in comparison to best available expert knowledge.

RESULTS

Determination of target taxa, delineation of the target area and harvesting of occurrence data

We conducted a literature review for the *Phaseolus* genus (Freitag & Debouck 2002, Salcedo *et al.* 2006, Delgado-Salinas & Carr 2007, Debouck 2009, Salcedo *et al.* 2009), checked against genepool experts (Debouck) and created a complete list of taxonomically verified species. We used the concept established by Maxted *et al.* (2006), including with equal weight all taxa belonging to taxon groups 1 to 4 of the genepool.

According to a recent revision of the *Phaseolus* genepool (Debouck 2009), there are 81 species and 34 infra-specific taxa, totaling 115 taxa within the genepool. With various species synonyms and historical revisions (Maxted *et al.* 2004, Freitag & Debouck 2002, Debouck 2009), specimen identification and data availability issues persist. Although taxonomically verified herbarium specimens provided the bulk of the data used in the analysis, we also rely

on the specimen identification performed by the individual holding institutions. Based on the recent history of *Phaseolus* taxonomy, we made the following changes to the determination of specimens used in the data: Any variant within *P. polymorphus* Wats. was considered as *P. polymorphus*, and the same was done for *P. coccineus* L. and *P. leptostachyus* Benth. (Debouck 2000). The variants *P. polystachyus* subsp. *smilacifolius* (Pollard) Freytag and *P. polystachyus* subsp. *sinuatus* (Nuttall ex Torrey & Gray) Freytag were considered as separated species (*P. smilacifolius* and *P. sinuatus*, respectively), and the species *P. pyramidalis* Freytag and *P. palmeri* Piper were merged into *P. grayanus* Woot. & Standl. The only infraspecific taxa that were considered were those of wild teparies (*P. acutifolius*) and those of *P. maculatus*, for which there was not enough evidence for merging into single species. For taxa with ongoing taxonomic uncertainty (e.g., *P. neglectus* Hermann), we followed Debouck (2009) and CIAT's Genetic Resources Unit genebank practice. After these modifications, a total of 85 taxa were finally listed, including 81 species and 4 infraspecific taxa.

We gathered data from all known available sources, including primary datasets accessed directly from herbaria and genebanks, as well as online global databases, such as the Global Biodiversity Information Facility (GBIF, www.gbif.org), the System-wide Information Network for Genetic Resources (SINGER, www.singer.cgiar.org) database held by the CGIAR, and the United States Department of Agriculture (USDA) Germplasm Resources Information Network (GRIN, www.ars-grin.gov) database (Table 1).

Data were available for all taxa, including the 81 species, 2 subspecies and 2 varieties. The entire dataset was carefully geographically verified and corrected using BioGeomancer, and, when possible, new geographic references (coordinates) were added to the passport data. The final dataset contained 11,442 records, of which 6,926 (60.5%) had coordinates or enough location data to obtain coordinates, and 4,516 (39.5%) samples had no location data or coordinates.

The analysis was based on the native range for the genus throughout the Americas (northeastern United States to northern Argentina, including the Caribbean and the Galapagos Islands) (Freytag & Debouck 2002, USDA, ARS, National Genetic Resources Program [NGRP] 2009). Records outside the boundaries of the Americas, as well as those listed as weedy or cultivated, were deleted and a final dataset was produced for analysis. The average total number of samples

Table 1. List of institutions from which data was harvested.

Institution	Number of records with coordinates	Number of records without coordinates
<i>Genebank accessions</i>		
Bioversity International	7	51
CIAT-Genetic Resources Unit (via SINGER)	2278	250
German National Resource Centre for Biological Material (DSMZ)	0	2
International Livestock Research Institute (ILRI)	0	271
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)	0	21
National Vegetable Germplasm Bank, Mexico (BANGEV)	7	0
Native Seeds/SEARCH (NSS)	37	1
Plant Breeding and Acclimatization Institute (IHAR)	0	17
US National Plant Germplasm System (NPGS-GRIN)	1081	771
Sub-total	3410	1384
<i>Herbarium samples</i>		
A Database System for Systematics and Taxonomy (SysTax)	2	49
Arizona State University Vascular Plant Herbarium	829	172
Bernice Pauahi Bishop Museum	0	1
Botanic Garden and Botanical Museum Berlin-Dahlem	0	1
Cahiers de Phaseologie (DGD)	1486	182
Canadian Biodiversity Information Facility	0	1
Colorado State University Herbarium (CSU)	33	4
Comision nacional para el conocimiento y uso de la biodiversidad (CONABIO)	1049	360
Dutch national node of the Global Biodiversity Information Facility (NLBIF)	0	25
Fairchild Tropical Botanic Garden Virtual Herbarium	2	18
GBIF-Spain	0	5
GBIF-Sweden	0	6
Harvard University Herbaria	2	86
Herbarium of the University of Aarhus	8	0
Instituto de Biología, Universidad Nacional de México, (IBUNAM)	0	2
Instituto de Ciencias Naturales	22	68
Instituto Nacional de Biodiversidad (Costa Rica)	78	0
Integrated Taxonomic Information System (ITIS)	8	0
Louisiana State University Herbarium	0	9
Missouri Botanical Garden	713	621
Museo Nacional de Costa Rica	100	45
Muséum national d'histoire naturelle et Réseau des Herbiers de France	4	0
National Botanic Garden of Belgium (NBGB)	70	20
National Museum of Natural History	28	64
NatureServe	0	134
NavNat, GE, FR	2	0
New Mexico Biodiversity Collections Consortium	0	112
New York Botanical Garden (NYBG)	7	4
Royal Botanic Gardens, Kew	1	2
The Deaver Herbarium, Northern Arizona University	8	0
University of Alabama Biodiversity and Systematics	6	0
University of California, Davis	0	7
University of Connecticut	1	0
University of Kansas Biodiversity Research Center	1	3
USDA Plants	402	65
Utah Valley State College (UVSC)	1	3
Sub-total	4863	2069
Total (genebank accessions and herbarium samples)	8273	3453

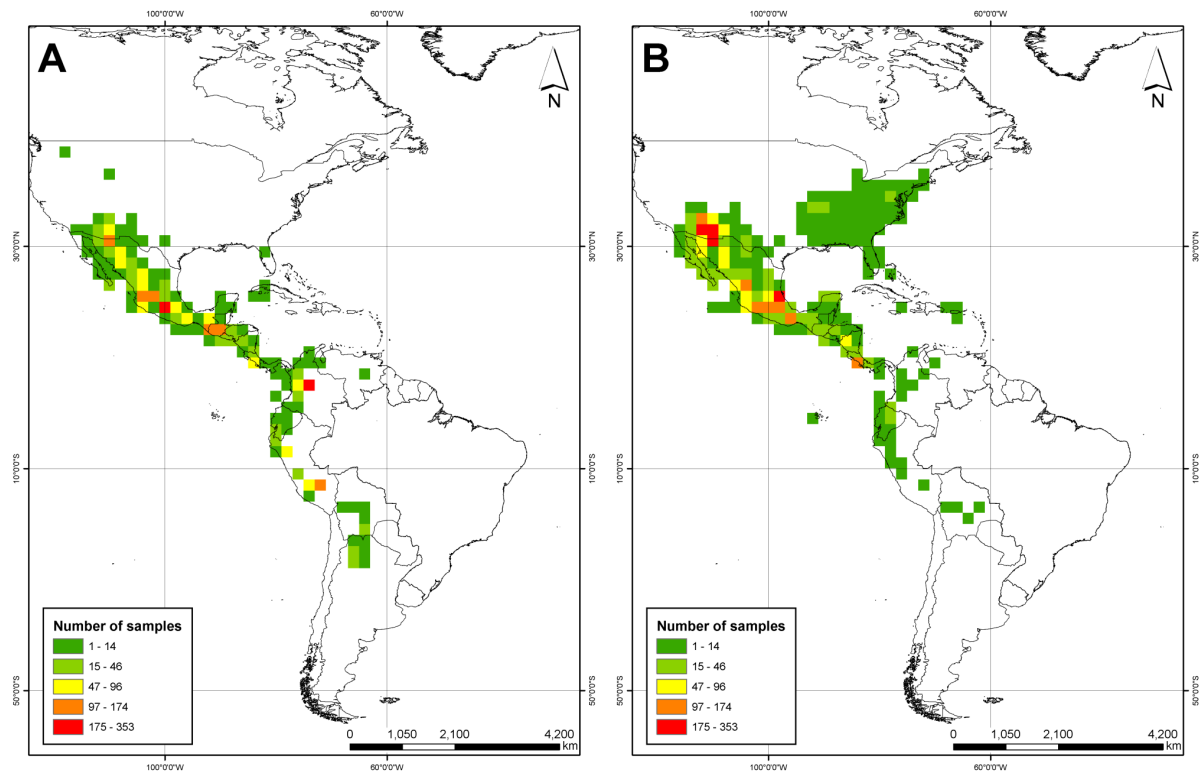


Figure 1. Density of sampling (sampling richness) for (A) genebank assessments and (B) herbarium records for *Phaseolus*.

per taxon was 144.8, but data was unevenly distributed. Samples were predominantly concentrated in wild progenitors of domesticated species (i.e., *P. acutifolius*, *P. coccineus*, *P. dumosus*, *P. lunatus*, *P. vulgaris*), comprising about 55% of the total records.

Germplasm collections of the *Phaseolus* genepool are not distributed equally in relation to total herbarium collections (Figure 1). The number of genebank accessions in a 200 km cell ranged from 1 to 273, while that of herbarium collections ranged from 1 to 373. Observable differences in the two maps (gaps) are present in the eastern United States, Costa Rica, Nicaragua, and in the north of Mexico and along its border with United States. Most of the areas in central Mexico are however well sampled and it is possible that species occurring in those areas are adequately conserved. This was also observed in some areas in South America (particularly in the Colombian, Ecuadorian and Peruvian Andes), where a greater proportion of genebank accessions have been collected, potentially indicating a better coverage of taxa in genebanks for populations from these regions.

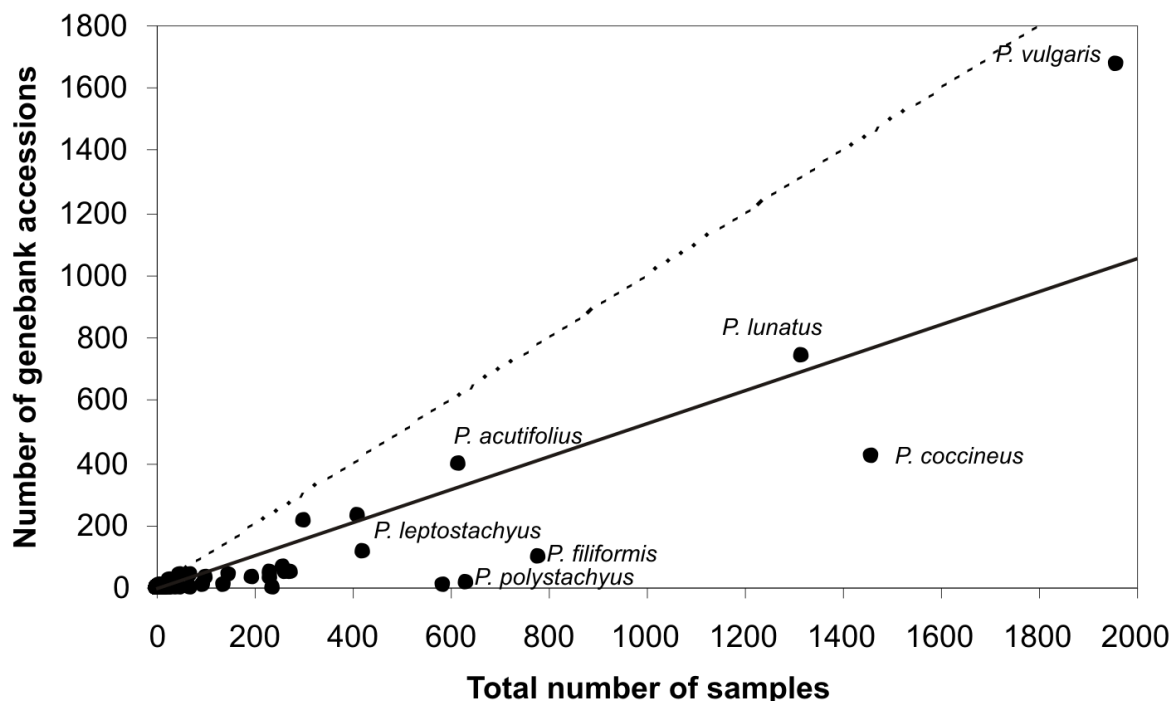


Figure 2. Number of genebank accessions versus all samples (genebank accessions plus herbarium specimen records).

Sampling deficiencies at the taxa level

Of 85 taxa, 35 (41.2%) had no germplasm accessions, 26 taxa (30.6%) had 1-9 accessions, and 24 taxa (28.2%) had 10 or more accessions. From the 85 taxa, 61 (71.8%) taxa presented a SRS below 3, indicating poor representativeness of the number of genebank accessions in relation to herbarium collections, whilst 16 taxa (18.8%) showed SRS between 3.01 and 5, 4 (4.7%) between 5.01 and 7.5, and 4 (4.7%) greater than 7.5.

The total representativeness (only in terms of the total number of samples, Figure 2 – intermittent line) is above the average representativeness of germplasm collections (continuous line), signifying that on average, species are likely to have fewer genebank accessions than herbarium specimens. *P. vulgaris*, *P. acutifolius* and *P. lunatus* appear well conserved in relation to both the gross number of accessions (compared to other taxa), and in proportion to their respective number of herbarium records.

Potential distribution models for taxa

We used high-resolution global climatic datasets developed by Hijmans *et al.* (2005). WorldClim includes monthly data at 30 arc-seconds resolution (approximately 1 km near the Equator) for total precipitation, and mean, maximum and minimum temperatures. Using such

Table 2. List of derived bioclimatic variables used in the analysis.

ID	Variable name	Units
P1	Annual mean temperature	°C
P2	Mean diurnal temperature range	°C
P3	Isothermality	N/A
P4	Temperature seasonality (standard deviation)	°C
P5	Maximum temperature of warmest month	°C
P6	Minimum temperature of coldest month	°C
P7	Temperature annual range	°C
P8	Mean temperature of wettest quarter	°C
P9	Mean temperature of driest quarter	°C
P10	Mean temperature of warmest quarter	°C
P11	Mean temperature of coldest quarter	°C
P12	Annual precipitation	mm
P13	Precipitation of wettest month	mm
P14	Precipitation of driest month	mm
P15	Precipitation seasonality (coefficient of variation)	%
P16	Precipitation of wettest quarter	mm
P17	Precipitation of driest quarter	mm
P18	Precipitation of warmest quarter	mm
P19	Precipitation of coldest quarter	mm

monthly datasets, 19 bioclimatic variables have been derived (Busby 1991), representing average yearly climates, stressful and extreme conditions, and interannual seasonality (Table 2).

We downloaded WorldClim data at 30 arc-seconds, calculated the bioclimatic indices and aggregated the 30 arc-seconds datasets to 2.5 arc-minutes using a bilinear interpolation in order to reduce the computational time and data storage needs. Although most of the bioclimatic indices used to develop the niche models are highly correlated (particularly in the tropics), we used the complete set of 19 bioclimatic variables in Table 2 because (1) they are useful to provide the best possible description of the climatic requirements of species during a single average year, (2) these correlations might not hold in space and time, (3) the alternative approach of dropping some variables leads to underestimation of distributions and poor performance of Maxent (Hijmans & Graham 2006), (4) the alternative approach of reducing the set of variables to a subset of orthogonal vectors (Dormann 2006) might lead to loss of valuable climatic information and tends to complicate the interpretation of results of the application of the niche model, and (5) the Maxent model prevents over-fitting due to the use of a set of correlated environmental predictors by assigning weights based on the relative importance of the variable to the model (Elith *et al.* 2006, Phillips *et al.* 2006, Phillips & Dudik 2008).

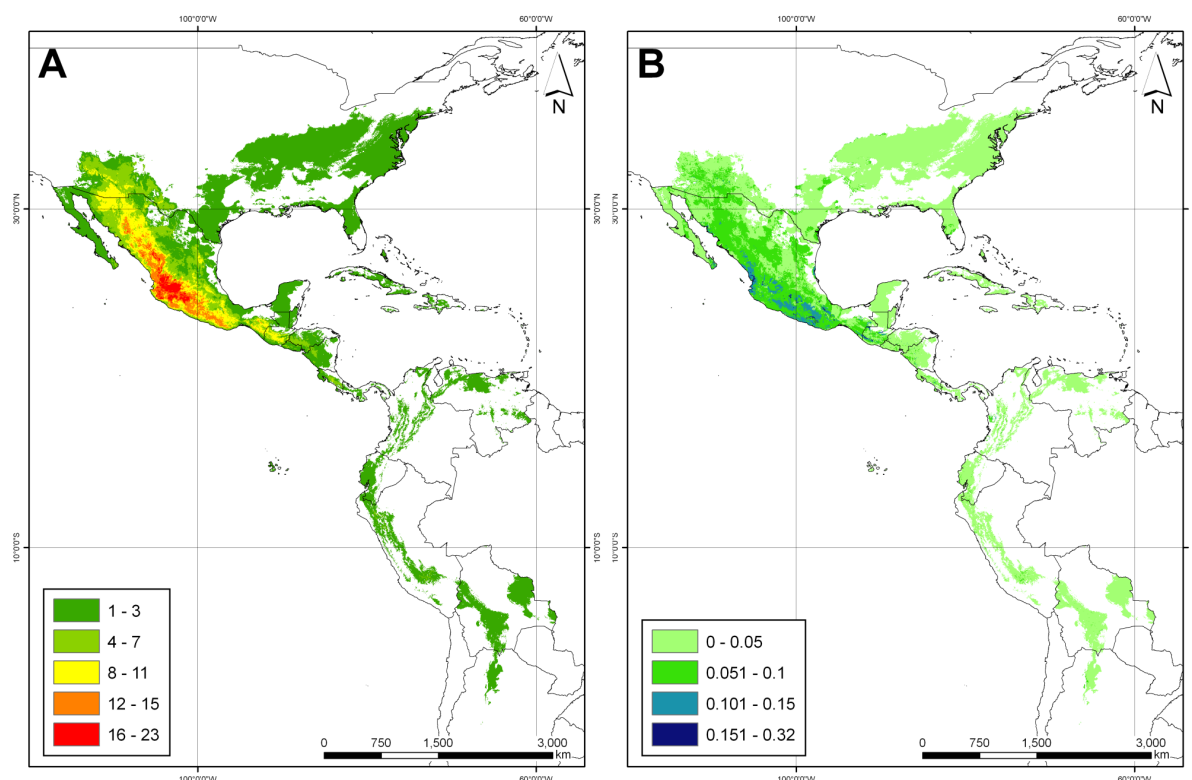


Figure 3. (A) Taxon richness for *Phaseolus* based upon potential area of distribution of all taxa. (B) modeling uncertainties as maximum standard deviations among all modeled taxa.

The geographic distributions of 51 out of the 85 taxa were considered sufficiently accurate and stable to be mapped. Potential distribution coverage was estimated via the convex hull method for 3 additional taxa (*P. marechalii*, *P. salicifolius*, and *P. rotundatus*). Therefore, a total of 54 taxa were assessed further.

The genus was modeled to occur from the northern border of the United States through Central America, and along the Andean chain into northern Argentina (Figure 3A). Potential taxon richness ranged from 1 to 23 taxa per grid cell. Taxon diversity hotspots were mainly found in southern and western Mexico and in the southern United States, as well as some highland areas of Guatemala, Honduras and Costa Rica, where 6 to 11 taxa are potentially distributed in a single 5 km pixel.

Uncertainties in modeling distributional range calculated by the maximum standard deviation among any possible class (i.e., taxon) varied from 0 to 0.32 (Figure 3B), with the vast majority of the potential distribution coverage of the genus presenting a modeling uncertainty below 10%, and only very few areas presenting more than 15% variation in predicted probabilities. High uncertainty areas do not coincide with high diversity areas, confirming the

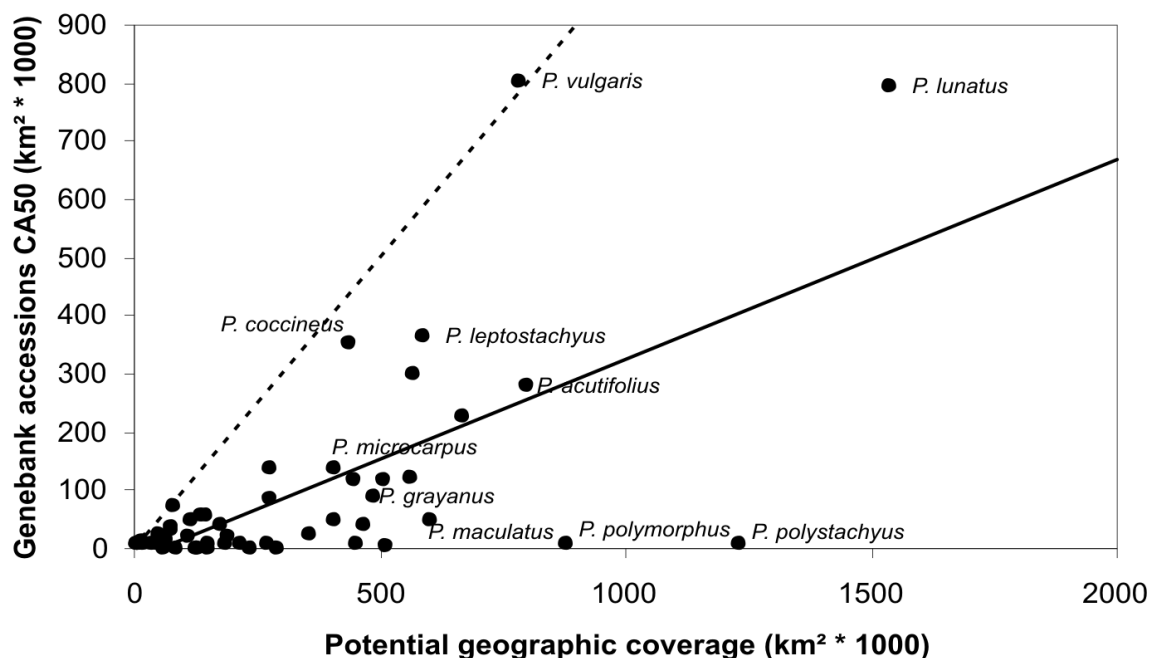


Figure 4. Geographic coverage of genebank accessions against total potential distribution coverage of taxa.

reliability of the Maxent algorithm in predicting the geographic distributions of our set of taxa. These small spots are located in southwestern Mexico along the very western edge of Nayarit (municipalities of El Nayar, Rosamorada, Tepic), along the borders of Guerrero and Oaxaca, in northern Oaxaca, and in northeastern Michoacán. Despite the observed uncertainties, these areas with more than 15% variability among predictions account to less than 10% of the total potential distribution coverage of the genus.

Geographic coverage assessment

The comparison between the CA50 and the size of the potential distribution showed that there are 30 taxa out of the 54 assessed (55.6%) with GRS below 3.01 (less than 30% of representativity in terms of geographic coverage), 12 taxa (22.2%) with GRS between 3.01 and 5, 4 taxa (7.1%) with GRS between 5.01 and 7.5, and 8 taxa (14.8%) with GRS greater than 7.5. The great majority of taxa have germplasm collections covering a geographic range considerably smaller than the potential geographic area in which the taxon is distributed (Figure 4), thus indicating the need for further collecting in order to fill geographic gaps.

The average representativeness line (intermittent line) is above the complete representativeness line (continuous line), indicating that the representativeness of germplasm collections in

comparison to the total potential distribution coverage is low on average, and relatively high only for a few species (namely the wild progenitors *P. vulgaris*, *P. coccineus*, *P. acutifolius* and *P. lunatus*).

Determination of environmental gaps

The principal components analysis showed that the first two components explained up to 81.5% of the total spatial variability among the *Phaseolus* genebank target area (61.2 and 20.3% for PC1 and PC2 respectively). Re-scaling of these two components' weights resulted in a weight of 75.03% for PC1 and 24.97% for PC2. Out of the 54 modeled taxa, 10 (18.5%) presented ERS below 3.01, indicating a significantly low environmental representativeness (i.e., less than 30%) in germplasm collections; 7 (13%) taxa presented an ERS between 3.01 and 5; 7 taxa (13%) between 5.01 and 7.5; and 30 taxa (55.6%) above 7.5. Notably, environmental representativeness of genebank accessions was found to predominantly fit in the two extreme classes (below 30% and above 75%) for most of the taxa.

P. vulgaris and *P. lunatus* showed the highest coverage of potential environmental range, with 8 and 14 respectively out of the 20 classes along PC1, and 8 and 16 classes along PC2 (Figure 5). Germplasm representativeness of these environmental classes is for both species significantly high (90% or more representativeness in both classes). For wild *P. vulgaris*, among other cases (Figure 5), we found the environmental distribution of genebank accessions to be broader than the environmental distribution of the potential distribution coverage, which may be explained as an artifact given the use of the ROC-plot based threshold for binning the species distributions (i.e., the omission rate), the native area (i.e., one or two small localities where the taxon occurs might not be reported in literature), or the use of the CA50 around germplasm locations, which might enlarge the range towards unsuitable habitats, particularly where the landscape changes rapidly (e.g., topographically diverse regions, such as the Andes). A broad range of adaptation to climatic conditions may be covered by current germplasm collections, but it should be noted that small environmental gaps remain even for these well-sampled species.

Rarity of each species

Rarity of species was not included in the analysis since there were significant sampling biases that would lead to inaccurate results. In order to produce accurate results, the weight of the

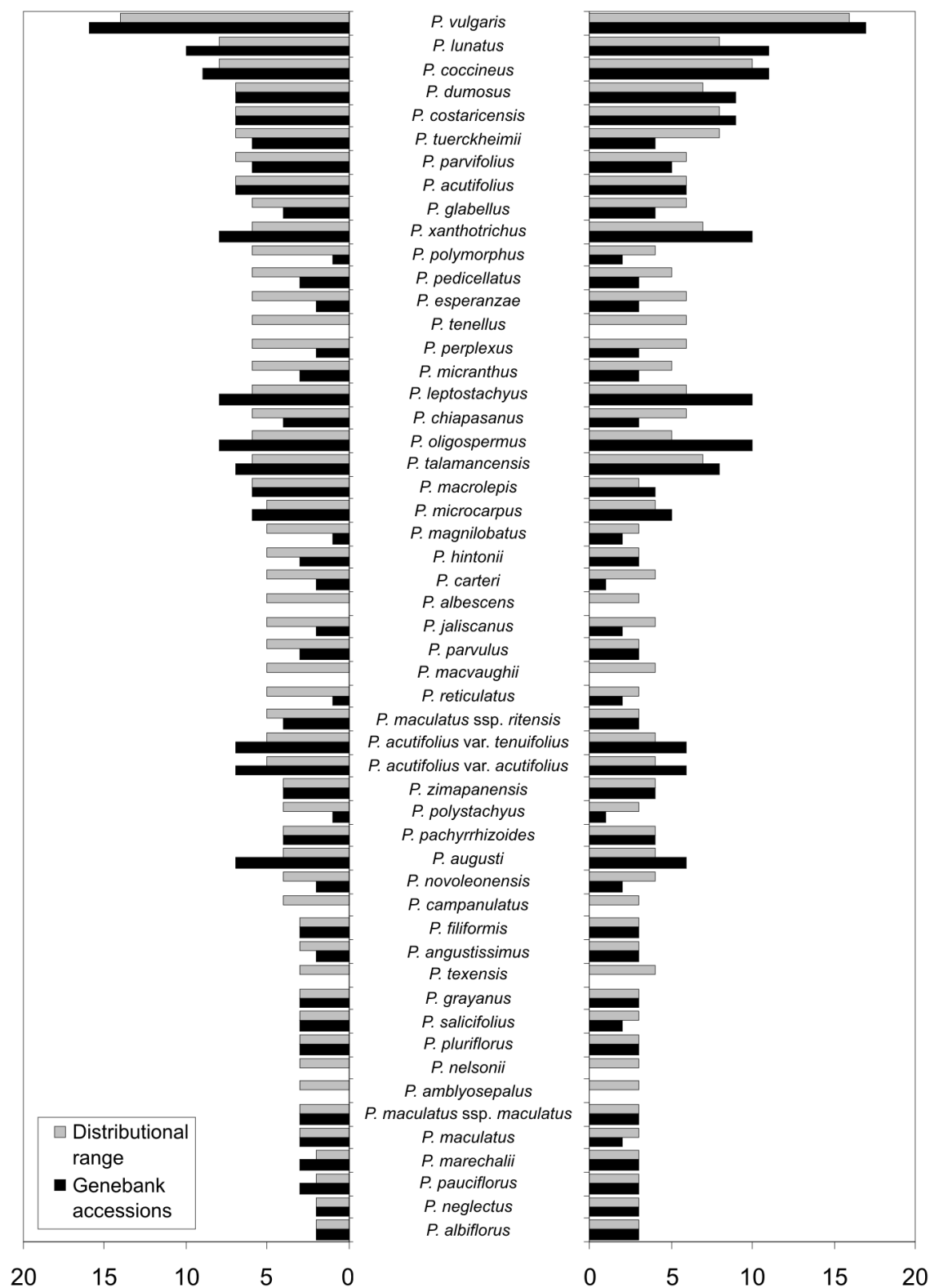


Figure 5. Coverage of genebank accessions versus potential environmental area for modeled species for the first (left) and second (right) principal components.

Table 3. List of taxa and data included in the analysis (see Supplementary Table 1 for full details).

Taxon	HS ¹	HS (RP) ³	GA ²	GA (RP) ³	Total	Total (RP) ³	FPS (GAP)	Class (GAP) ⁴	EPS (DGD)	Class (DGD) ⁴
Sect. Acutifolii										
<i>P. acutifolius</i>	219	119	396	67	615	186	5.7	LPS	NA	NA
<i>P. acutifolius</i> var. <i>acutifolius</i>	87	75	211	81	298	154	6.8	LPS	4	MPS
<i>P. acutifolius</i> var. <i>tenuifolius</i>	177	103	232	93	409	188	6.1	LPS	5	MPS
<i>P. parvifolius</i>	62	56	37	22	99	74	4.5	MPS	4	MPS
Sect. Bracteati										
<i>P. macrolepis</i>	24	6	3	3	27	6	8.3	NFCR	4	MPS
<i>P. talamancensis</i>	13	4	2	1	15	4	7.5	LPS	6	LPS
Sect. Brevilegumeni										
<i>P. campanulatus</i>	4	4	0	0	4	4	0.0	HPS	0	HPS
<i>P. oligospermus</i>	26	22	13	11	39	33	5.8	LPS	3	HPS
<i>P. tuerckheimii</i>	43	24	3	2	46	26	3.5	MPS	3	HPS
Sect. Chiapasana										
<i>P. chiapasanus</i>	53	8	3	3	56	8	4.1	MPS	2	HPS
Sect. Coccinei										
<i>P. coccineus</i>	1041	356	417	206	1458	560	7.3	LPS	4	MPS
Sect. Coriacei										
<i>P. maculatus</i>	106	62	39	17	145	79	4.0	MPS	NA	NA
<i>P. maculatus</i> ssp. <i>maculatus</i>	203	138	30	18	233	151	4.5	MPS	4	MPS
<i>P. maculatus</i> ssp. <i>ritensis</i>	190	120	68	30	258	150	4.6	MPS	2	HPS
<i>P. novoleonensis</i>	4	3	2	1	6	3	3.6	MPS	2	HPS
<i>P. reticulatus</i>	6	4	2	2	8	6	2.3	HPS	3	HPS
<i>P. venosus</i> ++	10	6	0	0	10	6	0.0	HPS	0	HPS
Sect. Digitati										
<i>P. albiflorus</i>	49	4	1	1	50	4	4.3	MPS	6	LPS
<i>P. albiviolaceus</i> +	1	1	0	0	1	1	0.0	HPS	2	HPS
<i>P. altimontanus</i> #	2	2	2	2	4	2	NA	HPS	4	MPS
<i>P. neglectus</i>	9	6	0	0	15	11	0.0	HPS	2	HPS
<i>P. trifidus</i> ++	1	1	0	0	1	1	0.0	HPS	NA	NA
Sect. Falcati										
<i>P. leptostachyus</i>	308	170	115	102	423	270	6.7	LPS	4	MPS
<i>P. macvaughii</i>	11	7	1	1	11	7	1.4	HPS	2	HPS
<i>P. micranthus</i>	21	9	2	1	23	10	2.1	HPS	4	MPS
<i>P. opacus</i> ++	4	1	0	0	4	1	0.0	HPS	NA	NA
<i>P. persistentus</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
Sect. Minkelsia										
<i>P. amabilis</i> ++	8	1	0	0	8	1	0.0	HPS	0	HPS
<i>P. amblyosepalus</i>	10	10	0	0	10	10	0.0	HPS	0	HPS
<i>P. anisophyllus</i> ++	2	2	0	0	2	2	0.0	HPS	0	HPS
<i>P. nelsonii</i>	38	32	0	0	38	32	0.0	HPS	2	HPS
<i>P. parvulus</i>	168	101	29	17	197	118	3.2	MPS	2	HPS
<i>P. pauciflorus</i>	234	161	4	2	238	163	4.4	MPS	2	HPS
<i>P. perplexus</i>	11	7	2	1	13	8	1.7	HPS	3	HPS
<i>P. plagiocylis</i> ++	4	2	0	0	4	2	0.0	HPS	0	HPS
<i>P. pluriflorus</i>	86	56	10	7	96	63	4.0	MPS	3	HPS
<i>P. tenellus</i>	21	9	2	1	22	9	1.0	HPS	0	HPS
Sect. Paniculati										
<i>P. acinaciformis</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
<i>P. albinervus</i> ++	3	1	0	0	3	1	0.0	HPS	0	HPS
<i>P. augusti</i>	27	15	43	39	70	54	7.4	LPS	7	LPS
<i>P. jaliscanus</i>	66	12	2	1	68	12	1.8	HPS	2	HPS
<i>P. juquilensis</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
<i>P. lignosus</i> +	2	2	0	0	2	2	0.0	HPS	0	HPS
<i>P. longiplacentifer</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
<i>P. lunatus</i>	575	275	742	342	1317	616	6.9	LPS	4	MPS
<i>P. maculatifolius</i> ++	2	1	0	0	2	1	0.0	HPS	0	HPS
<i>P. marechalii</i>	10	4	5	2	15	4	8.3	NFCR	3	HPS
<i>P. mollis</i> ++	14	6	0	0	14	6	0.0	HPS	0	HPS
<i>P. nodosus</i> ++	2	2	0	0	2	2	0.0	HPS	2	HPS
<i>P. pachyrrhizoides</i>	5	2	21	20	26	22	7.8	NFCR	8	NFCR
<i>P. polystachyus</i>	580	344	6	2	586	346	0.9	HPS	2	HPS
<i>P. rotundatus</i> ++	3	2	1	1	3	2	6.7	LPS	5	MPS
<i>P. salicifolius</i>	10	3	1	1	11	4	7.2	LPS	0	HPS
<i>P. scrobiculatifolius</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
<i>P. sinuatus</i> #	76	12	1	1	77	12	NA	HPS	2	HPS
<i>P. smilacifolius</i> ++	13	2	0	0	2	1	0.0	HPS	0	HPS
<i>P. sonorensis</i> ++	16	3	0	0	16	3	0.0	HPS	0	HPS
<i>P. xolocotzii</i> ++	1	1	0	0	1	1	0.0	HPS	0	HPS
Sect. Pedicellati										

<i>P. dasycarpus</i> ++	5	5	0	0	5	5	0.0	HPS	0	HPS
<i>P. esperanzae</i>	26	15	7	7	33	15	4.4	MPS	2	HPS
<i>P. grayanus</i>	184	77	49	36	233	113	5.0	MPS	3	HPS
<i>P. laxiflorus</i> ++	5	1	0	0	5	1	0.0	HPS	0	HPS
<i>P. oaxacanus</i> ++	6	2	0	0	6	2	0.0	HPS	3	HPS
<i>P. pedicellatus</i>	129	71	8	8	137	79	2.9	HPS	4	MPS
<i>P. polymorphus</i>	23	5	1	1	24	6	1.4	HPS	3	HPS
<i>P. purpusii</i> ++	5	1	0	0	5	1	0.0	HPS	0	HPS
<i>P. scabrellus</i> +	4	4	0	0	4	4	0.0	HPS	2	HPS
<i>P. teulensis</i> +	1	1	0	0	1	1	0.0	HPS	0	HPS
<i>P. texensis</i>	7	6	0	0	7	6	0.0	HPS	3	HPS
Sect. Phaseoli										
<i>P. albescens</i>	8	8	0	0	8	8	0.0	HPS	2	HPS
<i>P. costaricensis</i>	64	44	4	3	68	46	6.6	LPS	6	LPS
<i>P. dumosus</i>	52	14	9	7	61	14	6.5	LPS	5	MPS
<i>P. vulgaris</i>	284	209	1674	452	1958	661	8.9	NFCR	7	LPS
Sect. Revoluti										
<i>P. leptophyllus</i> +	6	1	0	0	6	1	0.0	HPS	0	HPS
Sect. Rugosi										
<i>P. angustissimus</i>	617	269	17	8	634	275	2.8	HPS	2	HPS
<i>P. carteri</i>	8	3	5	2	13	4	3.9	MPS	2	HPS
<i>P. filiformis</i>	682	397	98	46	780	441	4.6	MPS	2	HPS
Sect. Xanthotricha										
<i>P. esquincensis</i> ++	4	3	0	0	4	3	0.0	HPS	0	HPS
<i>P. gladiolatus</i> ++	1	1	0	0	1	1	0.0	HPS	3	HPS
<i>P. hintonii</i>	12	7	11	7	23	14	4.3	MPS	2	HPS
<i>P. magnilobatus</i>	16	7	2	1	18	8	1.6	HPS	2	HPS
<i>P. xanthotrichus</i>	11	8	38	30	49	38	9.0	NFCR	5	MPS
<i>P. zimapanensis</i>	10	5	16	13	26	17	7.3	LPS	6	LPS
Not classified										
<i>P. glabellus</i>	128	42	15	10	160	42	3.7	MPS	5	MPS
<i>P. microcarpus</i>	223	161	51	35	274	193	5.1	LPS	4	MPS

¹Number of herbarium specimens. ²Number of genebank accessions. ³Refers to the number of populations (unique locations identified) represented by the set of samples. ⁴Prioritization of taxa is done as follows: HPS: High priority species, MPS: Medium priority species, LPS: Low priority species, NFCR: No further urgent conservation required. FPS indicates the result of the method proposed in this paper, and EPS indicates the prioritization given by expert knowledge (based on Daniel G. Debouck's expertise in *Phaseolus*). +Indicates that the taxon had no genebank accessions and no herbarium samples with coordinates or location data; ++indicates a taxon for which a Maxent model was not possible and for which 0-few genebank accessions were available; #indicates a taxon with some genebank accessions but no or limited herbarium samples with coordinates or location data. These taxa are listed as HPS for further collecting in order to inform the gap analysis.

ERTS was finally established at 0.05, which is practically irrelevant and thus the step was dropped. If a subset of species with reliable sampling were to be analyzed separately (e.g., the five wild progenitors of the domesticated species), however, the ERTS could be calculated and weighted equally with the other scores when calculating the FPS.

Numeric assessment to determine the priority of collecting for *ex situ* conservation for each taxon

Out of the 85 taxa under analysis, 48 (56.5%) are either under-represented or not represented in any way in genebanks and therefore flagged as HPS for collecting (Table 3). Of these taxa, 35 had no germplasm accessions, and 11 are listed as HPS due to the average of gross representativeness, geographic, and environmental gaps (FPS below 3.01). A further 2 taxa (*P. sinuatus* and *P. altimontanus*) couldn't be assessed due to uncertainties in the modeling and the

data, and are included as HPS due to the need for collecting in order to provide adequate data for a gap analysis.

Medium priority for further collecting was given to 17 taxa (20%), 15 taxa (17.7%) were given low priority, and only 5 taxa (*P. macrolepis*, *P. marechalii*, *P. pachyrrhizoides*, *P. xanthotrichus* and *P. vulgaris*) were assessed as well represented in *ex situ* collections.

Prioritization of geographic areas for collecting germplasm

Thirty-six priority taxa (i.e., those flagged as high priority and with sufficient location data) were mapped together, along with standard deviations on predicted Maxent probabilities (aggregated for all the taxa using the maximum value) and distances to the nearest population (also aggregated) (Figure 6). Potential collection sites have a richness of up to 7 taxa per grid (Figure 6A). Zones where gaps in *ex situ* collections for many *Phaseolus* taxa overlap are concentrated in central-western Mexico, with an extension along the Sierra Madre Occidental north to Sonora.

Andean environments where *Phaseolus* species are likely distributed appear in general to be adequately represented in genebanks for most of the species. Note that the narrow endemic nature of many of the under- or un-sampled taxa results in a need for very finely targeted collection trips to specific regions outside of the gap richness areas identified, for example to collect from populations of *P. carteri*, *P. novoleonensis*, and *P. plagiocylix* in isolated regions of Mexico, and *P. mollis* in South America.

The maximum modeling uncertainty (given by the maximum standard deviation of the 25 folds per taxon) was slightly greater than 15% in a very small area (dark blue spot in western Nayarit, Figure 6B). Interestingly, modeling uncertainties of high priority taxa had a maximum of 19%, significantly lower than uncertainties of the whole set of taxa under analysis (Figure 3B), strengthening confidence in results regarding high priority taxa. The distance to verified populations (Figure 6C) was greatest (i.e., uncertainty) in northwestern Mexico (southern Sonora, northern Sinaloa, and southwestern Chihuahua). The areas identified in these uncertainty analyses are least likely to contain target species.

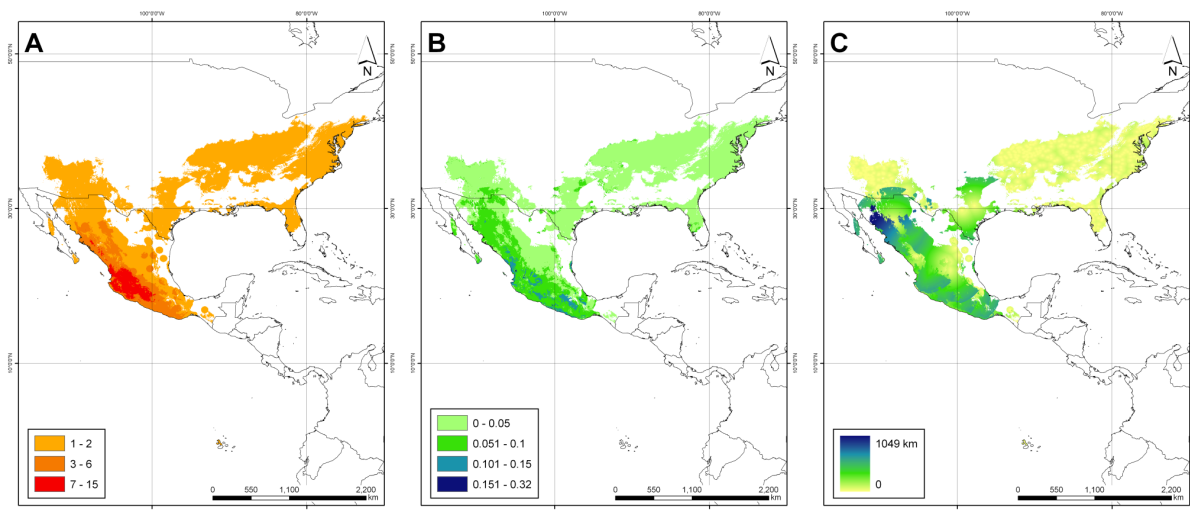


Figure 6. (A) Zones where gaps in *ex situ* collections for multiple taxa overlap (collecting gap richness) for high priority species. (B) modeling uncertainties as standard deviations among high priority modeled taxa. (C) collecting uncertainties as maximum geographic distance to nearest known population.

Comparison with expert opinion

The expert authority for *Phaseolus* was Daniel G. Debouck (DGD), head of the Genetic Resources Unit at the International Center for Tropical Agriculture (CIAT), author and co-author of numerous publications on *Phaseolus*, including a survey of the *Phaseolus* genepool in North and Central America (Freytag & Debouck 2002), who has participated in many collecting missions for the genus throughout the Americas and has extensive expertise in taxonomy (including research at 67 different herbaria in the last 32 years), ecogeographic distributions, and level of *in situ* and *ex situ* conservation of the genepool.

DGD did not assess 4 taxa: *P. maculatus*, and *P. acutifolius* since he considered it enough to assess the subspecies and/or variants, and *P. trifidus* and *P. opacus*, since he considered them as doubtful taxa. All figures below are thus based on the total number of taxa assessed by DGD (81). Further taxonomic analyses of these species are needed in order to inform conservation priorities.

In comparison to expert opinion, the gap analysis approach tended to underestimate priority for collecting in a considerable number of cases (30.9% of the taxa); however, scores for 28 taxa (34.6%) did align with expert opinion (with 0 as score for 24 of these). For 51 taxa (63%), the method and DGD agreed on the priority class, and from the remaining proportion, the difference was of one single class. In addition, the relative difference (RD) varied from -50% to 72.2% and the maximum difference between our approach and the expert's concept was around 7

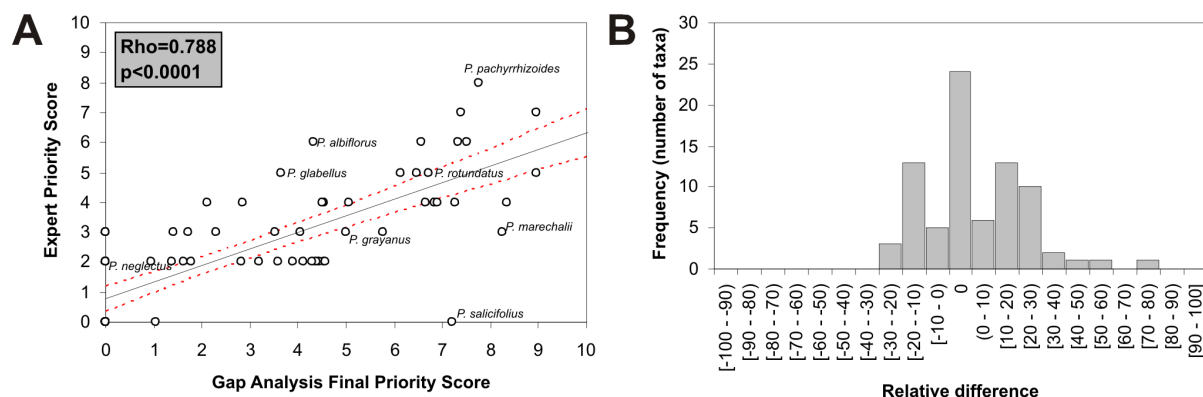


Figure 7. (A) Frequency distribution of the relative difference [RD] and (B) linear trend between Final Priority Score (FPS) and Expert Priority Score (EPS) (the red dotted line indicates a 95% confidence interval).

units in the priority scale of 10 units. Moreover, 87.7% of the validated taxa (81) presented differences lower than 30% or greater than -30%, and only 2 taxa presented more than a 50% or less than -50% difference (*P. salicifolius* with 7.2 in EPS and 0 in FPS, *P. marechalii* with 8.3 in EPS and 3 in FPS). Only *P. salicifolius* was found to have more than 70% difference between EPS and FPS (Figure 7A).

The linear trend between EPS and FPS has a Spearman correlation coefficient of 0.79 ($p < 0.0001$, $n=79$). However, as previously stated, the gap analysis approach tends to underestimate the priorities compared to expert opinion (average underestimation is -10.7%, Figure 7B).

A number of taxa fall far from the linear trend (i.e., *P. neglectus*, *P. albiflorus*, *P. salicifolius* and *P. pachyrrhizoides*). Whilst for *P. pachyrrhizoides* this is due to a very high accuracy (ERS and FPS are equal) in comparison with the propagating error in the regression line (i.e., the underestimation error), differences for other taxa generally result from lack of geographic data for a robust gap analysis, likely taxonomic misidentifications in records, and/or difficulty in eliminating duplicates in records (e.g., *P. neglectus*, *P. albiflorus*).

For species such as *P. xanthotrichus* and *P. oligospermus*, the gap analysis approach indicated little need for further collection, as germplasm has been collected throughout most of the region of recorded herbarium collections and environments occupied by those collections. However, expert knowledge on other areas of distribution of the species, under-recorded in online herbarium data, gave the species higher priority on the EPS.

DISCUSSION

Success of the gap analysis method in identifying priority taxonomic, geographic, and environmental gaps is directly dependent on the quality of input data and robustness of the modeling based upon the data. In this section we discuss uncertainties and limitations concerning the method:

Input data availability, bias and certainty

The quality of the input geographic information (i.e., climatic and occurrence data) directly affects the performance of species distributions models (Dormann 2006, Hijmans & Graham 2006, Graham *et al.* 2008, Loisele *et al.* 2008, Wisz *et al.* 2008, Costa *et al.* 2010). Geographic data for specimens is generally less than optimal and is unevenly distributed across taxa, due to the bias of collecting activities toward particular species or locations, a historically insufficient prioritization of recording and maintaining of geographic data, lack of high quality absence data for species, and limited accessibility of stored data for some collections. Many regions of the world remain un- or under-sampled, particularly highly inaccessible areas, and those chronically affected by war or civil strife.

Recently described and/or under-studied taxa, such as *P. acinaciformis*, *P. juquilensis*, *P. longiplacentifer*, *P. persistentus*, *P. scrobiculatifolius*, *P. teulensis*, *P. albiviolaecus*, *P. leptophyllus*, *P. lignosus*, *P. scabrellus*, and *P. sinuatus*, may require further taxonomic clarification, and are generally in need of further collecting, and characterization of the collected populations, in order to clarify identification and facilitate accurate prioritization.

Intraspecific taxa (variants and subspecies), such as those of *P. maculatus* and *P. acutifolius*, may also be incompletely treated in the analysis due to data constraints. There are several records of these species that remain undetermined at the intraspecific level. Due to overlapping ranges of distribution for various intraspecific taxa, unassigned records cannot be easily differentiated based on collection location. In the gap analysis we have therefore assessed both the species level and the intraspecific taxa.

More germplasm of *Phaseolus* may be conserved worldwide than the accession data used in this analysis indicate, as the data from some genebanks was not accessible. We assume that, with few exceptions, the accessions whose data was not accessible are also generally

inaccessible to crop breeders and researchers worldwide. Areas where these collections were made may not represent a gap for the particular holding collection, yet they are effectively a very real gap for rest of the global community.

Duplication between and within institutes might inflate the numbers of unique records for some of the taxa, leading to bias in the prioritization results. The use of different numbering systems, and lack of tracking of former records, leads to an overestimation of samples held, and difficulty in identifying duplicates, perhaps especially for the most commonly exchanged species (e.g., wild progenitors). For *Phaseolus*, we found that large differences can exist between the number of records and of actual populations both for genebanks (up to 83.1%) and herbaria (up to 87.5%). The data preparation phase of the analysis involved a thorough identification of duplicates in order to avoid inflation of numbers of records and therefore biases in prioritization. Further, the geographic representativeness score (GRS) takes distinctness/uniqueness of populations into account indirectly, and the environmental representativeness score (ERS) addresses the issue by illuminating gaps in the abiotic adaptations of the sampled material (i.e., number of different climatic environments covered by the conserved material).

Location data constraints may also limit the taxa for which the method may be applied, as well as lead to an underestimation of taxon distributions. From the 45 different data sources, 24 (53.3%) had more records without location data than with location data, and only 9 (20%) of the sources presented all of their records with coordinates or with detailed location data (Table 1). For genebanks, 71.1% of the data presented reliable location data and 28.9% had either no location data or location data were unreliable, whilst for herbaria, 70.2% of the data presented coordinates and 29.8% did not present any useful location data.

Additional data, such as absence of the taxon, would certainly improve model-training by increasing the model's ability to discriminate between presence and absence areas. These data are unfortunately not available in conventional genetic resources databases (Phillips & Dudik 2008, Lobo *et al.* 2010). Future collecting should be planned with an eye to the improvement of gap analysis approaches and should thus consider a more systematic recording of absences, geo-referencing all records, and making widely accessible data from all available germplasm and herbarium samples. These actions will improve the performance of species distribution models and any conclusions drawn from them.

Ability of the species distribution model used (i.e., Maxent) to adequately predict the potential and realized niche of taxa

The Maxent modeling technique was chosen for its ability to handle sample bias and spatial autocorrelation of species data (Loiselle *et al.* 2008, Phillips 2008, Costa *et al.* 2010) so as to provide high confidence species distributions models even given limited or biased location data. Maxent is an algorithm known to reliably predict the potential niches of species, and has been tested by several authors under a wide range of conditions and configurations (e.g., Elith *et al.* 2006, Phillips *et al.* 2006, Loiselle *et al.* 2008, Phillips 2008, VanDerWal *et al.* 2009, Costa *et al.* 2010, Smolik *et al.* 2010, among others); although we note that some reports (Jimenez-Valverde *et al.* 2008, Lobo *et al.* 2008, Fitzpatrick & Hargrove 2009) consider niche modeling techniques misleading and of limited use in certain contexts. As the robustness of Maxent is considered in the publications listed above, we do not provide a full analysis here.

We used the average test-data AUC, which showed that 52 species distribution models were reliable (i.e., accurate and stable). Using the current configuration, the AUC statistic is not likely to be biased by the pseudo-absences range (VanDerWal *et al.* 2009). Discrimination between presence and absence sites was therefore considerably positive for most of the taxa (~70%). Particularly good was the performance of taxon distribution models with more than 40 data points.

Moreover, the uncertainties associated with the application of a probabilistic model such as Maxent and depicted by the 25-fold cross-validated models for each of the taxa indicated that standard deviations among predictions ranged from 0 to 0.19. Collecting priorities are more uncertain in limited areas (e.g., along the western coast of Mexico), but are relatively robust across the vast majority of the distributional range of the genepool.

However, there was a set of taxa (those marked with + in Table 3) for which we were not able to develop species distributions models due to either lack of samples or to the distribution of those samples. These species could benefit from other approaches, such as Bayesian techniques (Termansen *et al.* 2006), which are able to develop probability surfaces even from a single point. Here we did not include these additional approaches, given the uncertainties involved with these models. We rather use specimen data (i.e., herbarium sampling points) to depict areas where these species can be potentially collected.

Geographic collecting priorities

To analyze the validity of geographic gap results, we have calculated the stability (standard deviation) of the Maxent models and have also provided the distance to the nearest population within the collection zone (Figure 6).

Additional analyses, including threat level, can be incorporated into the methodology in order to refine conservation priorities. Possible threats that could lead to genetic erosion in wild species populations include fires, grazing pressure, invasive species, deforestation, habitat modification and degradation, urbanization, and climate change, among others (Tanksley & McCouch 1997). Accession-level genetic data may also serve as an input in order to identify gaps in genetic diversity. Additional environmental data, such as soil type, may further define potential distributions of species. These additional inputs are currently only rarely available at high detail over large geographic areas or for all taxa in a genebank, but this may improve with the ongoing development of GIS and decreasing costs of genotyping. Taxon-specific knowledge may also be used to refine or weight priorities, giving some species higher importance in the final result (e.g., focusing on specific traits of interest, adjusting to phytosanitary/noxious weed constraints, recognizing legal constraints to access, prioritizing in order to capitalize on appropriate seasonal collecting windows, etc.)

In our approach, we include all wild relatives of the crop without regard to relatedness to cultivated species, weighting them equally, with the assumption that a wide range of taxa are potentially useful to provide genes for crop improvement (Challinor *et al.* 2007), recognizing the lack of data on relatedness. Information on relatedness and threat level can be added to the prioritization exercise by experts with specific interests or familiar with local conditions.

When this is done for *Phaseolus* the following gaps are highlighted. Collecting a few (1-5) populations is needed for 35 taxa that currently have no genebank samples conserved. Out of the five wild progenitors of the domesticated species, *P. vulgaris* and *P. dumosus* have been relatively well sampled, and only small gaps remain to be filled. Briefly, gaps for wild *P. vulgaris* are present in: Oaxaca, El Salvador, Panama, western Andes of Venezuela, northern central Bolivia, and San Luis in Argentina. For wild *P. dumosus*: eastern Chiapas and Alta Verapaz in Guatemala. For the remaining three progenitors, the gaps are substantially more important. For *P. acutifolius*: Sonora, Chihuahua, many spots in western Mexico and in

Guerrero. For *P. coccineus*: Chihuahua down to Guatemala. For *P. lunatus*, gaps exist throughout the very large range (from the Revillagigedo Islands, Baja California Sur and Sinaloa to Puerto Rico, and down to Salta and Formosa in Argentina).

Regarding the secondary genepool of each of the five cultigens: for common bean, runner bean, and year bean, additional collecting is needed for *P. albescens*, *P. costaricensis*, and *P. persistentus*. For tepary, collecting is needed for *P. parvifolius* (all across its range from Chihuahua down into Guatemala). For Lima bean, concerted effort is required because few (if any) accessions are available for taxa within Section *Paniculati*, as well as *P. maculatus*, *P. novoleonensis*, *P. reticulatus*, *P. ritensis*, and *P. venosus* within Section *Coriacei*.

For the remaining *Phaseolus* species (not highly related to any cultigen given molecular evidence available today), a few accessions exist for taxa such as *P. chiapasanus*, *P. esperanzae*, *P. pluriflorus*, and *P. micranthus*. Remaining species are in need of further collecting in order to secure germplasm *ex situ*.

Comparison with expert knowledge

The method performed well as compared with expert knowledge on the *Phaseolus* genepool, 81.2% of the taxa presenting differences between -30% and 30%, and only one taxon with a difference of more than 70% between EPS and FPS. We note that although the expert will often refine the analysis by adding further insight and by qualifying data, the gap analysis also holds the potential to highlight taxonomic, geographic, and environmental gaps previously unknown to the expert.

In order to provide a more robust test, multiple experts could be consulted. As GIS approaches continue to expand and improve, a more comprehensive validation procedure may be performed with a network of experts, facilitated through an online portal.

Expert intervention within the gap analysis method is especially critical during (1) thorough taxonomical review of the genepool, including variants and/or subspecies changes according to the latest studies, (2) the full evaluation and georeferencing of locality names in the dataset, and (3) the further refining and correction of priorities when a data availability issue is detected.

Expert taxonomic knowledge will of course also be vital in the actual field collecting, especially for understudied species (e.g., *P. albinervus*, *P. leptophyllus*, and *P. purpusii*). This has proven to be important in this genus, as numerous new species have been identified only during germplasm collecting missions (e.g., *P. altimontanus*, *P. costaricensis*, *P. novoleonensis*, *P. persistentus*, *P. rotundatus*, and *P. talamancensis*).

CONCLUSION

This study proposes a method for the rational prioritization of taxa within a gene pool for collection for *ex situ* conservation, using *Phaseolus* as a model. The method builds upon the standard comparison of herbarium samples with genebank accessions via gap analysis (Maxted *et al.* 2008), yet aims to address sampling biases by modeling species distributions with a robust algorithm, and refining these distributions using two different criteria. Furthermore, the method identifies priorities based not only on taxonomic and geographic gaps, but also environmental gaps. Priority locations for sampling of gaps result, as well as gap richness models contributing to the identification of collection locations for maximum efficiency. The results cover the four target outcomes of gap analysis identified by Nabhan (1990). Collecting for *ex situ* conservation should prioritize the resulting taxa, including those not or under-sampled *ex situ*, as well as geographic and environmental gaps in the distribution of taxa with some degree of germplasm currently conserved.

We found 48 high priority taxa (56.5%) (Table 3), 35 (41.1% of total) of these not recorded as represented *ex situ* by even a single accession. Acknowledging that the results for a number of these species may potentially be affected by data availability constraints, in the most optimistic case, around half of the taxa in the gene pool are highly under-represented in *ex situ* conservation. There is therefore a clear need for further collecting in order to cover the full range of taxonomic, geographic and environmental diversity.

The greatest priority regions for further collecting are located in northern Central America (i.e., Mexico and Guatemala), with a maximum potential sampling richness of 7 species per 5 km cell. However, there are a number of species that require individually targeted efforts in other areas (e.g., *P. mollis*, in the Galapagos Islands).

Additional criteria, such as threats to taxa, and degree of relatedness of taxa to cultivated

species, may also be included in the analysis, when data is sufficiently available. In order to include a more complete picture of conservation, the method should ideally be coupled with *in situ* gap analysis results (e.g., Meilleur & Hodgkin 2004), i.e., comparison of distributions with the extent of protected areas. In general, the high priority taxa identified in the analysis are likely to be those also most highly prioritized for *in situ* conservation, although this was not explored in the current analysis.

The method is applicable to any set of related taxa, given adequate geographic data and a thorough taxonomic and geographic referencing process. Genepools whose taxonomy has not received sufficient attention (e.g., *Oryza* in the Americas), or which have not been well sampled for herbarium specimens, will present particular challenges in producing reliable results. As each genepool is different, the analysis must be adapted according to data availability, and tested against expert knowledge, preferably repeatedly. Once the method has been applied to a number of crop genepools, the prioritization of taxa and “gap richness” mapping may be applied for these genepools together, potentially facilitating the identification of priority regions (“plant genetic resource gap megacenters”) for the efficient and effective collecting of CWR diversity on a global scale.

SUPPLEMENTARY INFORMATION

Supplementary Table 1. List of taxa and data included in the analysis.

Taxon	Number of herbarium specimens	Number of herbarium specimens (represented populations)	Number of genebank accessions	Number of genebank accessions (represented populations)	Total (represented populations)	ATAC	STAC	ASD15	SRS	GRS	ERS	FPS (GAP)	Class (GAP)	FPS (DGD)	Class (DGD)
Sect. <i>Acutifolii</i>															
<i>P. acutifolius</i>	219	119	396	67	615	0.962	0.0324	0	3.6	3.5	10	5.7	LPS	NA	NA
<i>P. acutifolius</i> var. <i>acutifolius</i>	87	75	211	81	298	0.965	0.0193	0	5.3	5.3	10	6.8	LPS	4	MPS
<i>P. acutifolius</i> var. <i>tenatifolius</i>	177	103	232	93	409	0.971	0.0308	0	4.9	3.4	10	6.1	LPS	5	MPS
<i>P. parvifolius</i>	62	56	37	22	99	0.972	0.0479	0	3	2.2	8.5	4.5	MPS	4	MPS
Sect. <i>Bracteati</i>															
<i>P. macrolepis</i>	24	6	3	3	27	0.998	0.0055	0.8	5	10	10	8.3	NFCR	4	MPS
<i>P. talamancensis</i>	13	4	2	1	15	0.999	0.001	0.4	2.5	10	10	7.5	LPS	6	LPS
Sect. <i>Brevilegumeni</i>															
<i>P. campanulatus</i>	4	4	0	0	4	0.995	0.002	0.2	0	0	0	0	HPS	0	HPS
<i>P. oligospermus</i>	26	22	13	11	39	0.979	0.0657	0	3.3	3.9	10	5.8	LPS	3	HPS
<i>P. tuerckheimii</i>	43	24	3	2	46	0.996	0.0066	0	0.8	2.1	7.7	3.5	MPS	3	HPS
Sect. <i>Chiapasana</i>															
<i>P. chiapasana</i>	53	8	3	3	56	0.993	0.0055	0	3.8	2.3	6.3	4.1	MPS	2	HPS
Sect. <i>Coccinei</i>															
<i>P. coccineus</i>	1041	356	417	206	1458	0.951	0.0207	0	3.7	8.1	10	7.3	LPS	4	MPS
Sect. <i>Coriacei</i>															
<i>P. maculatus</i>	106	62	39	17	145	0.986	0.0193	0	2.2	0.8	9.2	4	MPS	NA	NA
<i>P. maculatus</i> ssp. <i>maculatus</i>	203	138	30	18	233	0.979	0.0096	0	1.2	2.3	10	4.5	MPS	4	MPS
<i>P. maculatus</i> ssp. <i>ritensis</i>	190	120	68	30	258	0.987	0.0055	0	2	3.2	8.5	4.6	MPS	2	HPS
<i>P. novoleonensis</i>	4	3	2	1	6	0.995	0.002	0	3.3	2.4	5	3.6	MPS	2	HPS
<i>P. reticulatus</i>	6	4	2	2	8	0.959	0.0309	0.2	3.3	0.4	3.2	2.3	HPS	3	HPS
<i>P. venosus</i> ++	10	6	0	0	10	0.957	0.0371	12.9	0	NA	NA	0	HPS	0	HPS
Sect. <i>Digitati</i>															
<i>P. albiflorus</i>	49	4	1	1	50	0.995	0.0131	0	2.5	0.5	10	4.3	MPS	6	LPS
<i>P. albiviolaceus</i> +	1	1	0	0	1	NA	NA	NA	0	NA	NA	0	HPS	2	HPS
<i>P. altimontanus</i> #	2	2	2	2	4	0.5	NA	NA	10	NA	NA	NA	HPS	4	MPS
<i>P. neglectus</i>	9	6	0	0	15	0.997	0.005	0	0	0	0	0	HPS	2	HPS
<i>P. trifidus</i> ++	1	1	0	0	1	NA	NA	NA	0	NA	NA	0	HPS	NA	NA

Taxon	Number of herbarium specimens	Number of herbarium specimens (represented populations)	Number of genebank accessions	Number of genebank accessions (represented populations)	Total (represented populations)	ATAUC	STAUc	ASD15	SRS	GRS	ERS	FPS (GAP)	Class (GAP)	FPS (DGD)	Class (DGD)	
Sect. <i>Falcati</i>																
<i>P. leptostachyus</i>	308	170	115	102	423	270	0.969	0.0082	0	3.8	6.2	10	6.7	LPS	4	MPS
<i>P. macvaughii</i>	11	7	1	1	11	7	0.994	0.0045	0	0.9	0.6	2.1	1.4	HPS	2	HPS
<i>P. micranthus</i>	21	9	2	1	23	10	0.979	0.0264	0	1	0.1	5.2	2.1	HPS	4	MPS
<i>P. opacus</i> ++	4	1	0	0	4	1	NA	NA	NA	0	NA	NA	0	HPS	NA	NA
<i>P. persistentus</i> +	1	1	0	0	1	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
Sect. <i>Minklersia</i>																
<i>P. amabilis</i> ++	8	1	0	0	8	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. amblyosepalus</i>	10	10	0	0	10	10	0.982	0.0475	0	0	0	0	0	HPS	0	HPS
<i>P. anisophyllus</i> ++	2	2	0	0	2	2	0.5	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. nelsonii</i>	38	32	0	0	38	32	0.995	0.0046	0	0	0	0	0	HPS	2	HPS
<i>P. parvulus</i>	168	101	29	17	197	118	0.986	0.0193	0	1.4	1.2	7	3.2	MPS	2	HPS
<i>P. pauciflorus</i>	234	161	4	2	238	163	0.986	0.0069	0	0.1	3.2	10	4.4	MPS	2	HPS
<i>P. perplexus</i>	11	7	2	1	13	8	0.964	0.0419	0	1.3	0.2	3.7	1.7	HPS	3	HPS
<i>P. plagiocylis</i> ++	4	2	0	0	4	2	0.5	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. pluriflorus</i>	86	56	10	7	96	63	0.992	0.008	0	1.1	1	10	4	MPS	3	HPS
<i>P. tenellus</i>	21	9	2	1	22	9	0.994	0.0077	0.1	1.1	0.3	1.7	1	HPS	0	HPS
Sect. <i>Paniculati</i>																
<i>P. acinaciformis</i> +	1	1	0	0	1	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. albinervus</i> ++	3	1	0	0	3	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. augusti</i>	27	15	43	39	70	54	0.987	0.0102	0	7.2	4.9	10	7.4	LPS	7	LPS
<i>P. jaliscanus</i>	66	12	2	1	68	12	0.99	0.0138	0	0.8	0.3	4.2	1.8	HPS	2	HPS
<i>P. juquilensis</i> +	1	1	0	0	1	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. lignosus</i> +	2	2	0	0	2	2	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. longiplacentifer</i> +	1	1	0	0	1	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. lunatus</i>	575	275	742	342	1317	616	0.93	0.0191	0	5.6	5.2	10	6.9	LPS	4	MPS
<i>P. maculatifolius</i> ++	2	1	0	0	2	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. marechalii</i>	10	4	5	2	15	4	0.973	0.0271	40.2	5	9.8	10	8.3	NFCR	3	HPS
<i>P. mollis</i> ++	14	6	0	0	14	6	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. nodosus</i> ++	2	2	0	0	2	2	0.5	NA	NA	0	NA	NA	0	HPS	2	HPS

Taxon	Number of herbarium specimens	Number of herbarium specimens (represented populations)	Number of genebank accessions	Number of genebank accessions (represented populations)	Total (represented populations)	ATAUC	STAC	ASDIS	SRS	GRS	ERS	FPS (GAP)	Class (GAP)	FPS (DGD)	Class (DGD)
<i>P. pachyrrhizoides</i>	5	2	21	20	26	0.996	0.0058	0	9.1	4.2	10	7.8	NFCR	8	NFCR
<i>P. polystachyus</i>	580	344	6	2	586	0.95	0.0117	0	0.1	0.1	2.7	0.9	HPS	2	HPS
<i>P. rotundatus</i> ++	3	2	1	1	3	NA	NA	NA	5	5.1	10	6.7	LPS	5	MPS
<i>P. salicifolius</i>	10	3	1	1	11	0.99	0.0076	12.5	2.5	10	9.2	7.2	LPS	0	HPS
<i>P. scrobiculatifolius</i> +	1	1	0	0	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. sinuatus</i> #	76	12	1	1	77	NA	NA	NA	0.8	NA	NA	NA	HPS	2	HPS
<i>P. smilacifolius</i> ++	13	2	0	0	2	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. sonorensis</i> ++	16	3	0	0	16	0.992	0.0037	37.4	0	NA	NA	0	HPS	0	HPS
<i>P. xolocotzii</i> ++	1	1	0	0	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
Sect. <i>Pedicellati</i>															
<i>P. dasycarpus</i> ++	5	5	0	0	5	0.903	0.1941	83.1	0	NA	NA	0	HPS	0	HPS
<i>P. esperanzae</i>	26	15	7	7	33	0.992	0.0237	2	4.7	4.8	3.7	4.4	MPS	2	HPS
<i>P. grayanus</i>	184	77	49	36	233	0.985	0.0164	0	3.2	1.8	10	5	MPS	3	HPS
<i>P. laxiflorus</i> ++	5	1	0	0	5	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. oaxacanus</i> ++	6	2	0	0	6	0.5	NA	NA	0	NA	NA	0	HPS	3	HPS
<i>P. pedicellatus</i>	129	71	8	8	137	0.982	0.0161	0	1	2.3	5.2	2.9	HPS	4	MPS
<i>P. polymorphus</i>	23	5	1	1	24	0.903	0.0498	0	1.7	0.1	2.5	1.4	HPS	3	HPS
<i>P. purpusii</i> ++	5	1	0	0	5	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. scabrellus</i> +	4	4	0	0	4	NA	NA	NA	0	NA	NA	0	HPS	2	HPS
<i>P. teulensis</i> +	1	1	0	0	1	NA	NA	NA	0	NA	NA	0	HPS	0	HPS
<i>P. texensis</i>	7	6	0	0	7	0.99	0.0069	0	0	0	0	0	HPS	3	HPS
Sect. <i>Phaseoli</i>															
<i>P. albenscens</i>	8	8	0	0	8	0.995	0.0045	0.2	0	0	0	0	HPS	2	HPS
<i>P. costaricensis</i>	64	44	4	3	68	0.995	0.0094	0	0.7	9	10	6.6	LPS	6	LPS
<i>P. dumosus</i>	52	14	9	7	61	0.988	0.0144	2.5	5	4.4	10	6.5	LPS	5	MPS
<i>P. vulgaris</i>	284	209	1674	452	1958	0.949	0.0096	0	6.8	10	10	8.9	NFCR	7	LPS
Sect. <i>Revoluti</i>															
<i>P. leptophyllus</i> +	6	1	0	0	6	NA	NA	NA	0	NA	NA	0	HPS	0	HPS

Taxon	Number of herbarium specimens	Number of herbarium specimens (represented populations)	Number of genebank accessions	Number of genebank accessions (represented populations)	Total	Total (represented populations)	ATAUC	STAUc	ASD ¹ s	SRS	GRS	ERS	FPS (GAP)	Class (GAP)	FPS (DGD)	Class (DGD)
Sect. <i>Rugosi</i>																
<i>P. angustissimus</i>	617	269	17	8	634	275	0.98	0.0049	0	0.3	0.7	7.5	2.8	HPS	2	HPS
<i>P. carteri</i>	8	3	5	2	13	4	0.998	0.0014	6	5	3	3.6	3.9	MPS	2	HPS
<i>P. filiformis</i>	682	397	98	46	780	441	0.971	0.0082	0	1	2.6	10	4.6	MPS	2	HPS
Sect. <i>Xanthotricha</i>																
<i>P. esquincensis</i> ++	4	3	0	0	4	3	0.97	0.0105	61.4	0	NA	NA	0	HPS	0	HPS
<i>P. gladiolatus</i> ++	1	1	0	0	1	1	NA	NA	NA	0	NA	NA	0	HPS	3	HPS
<i>P. hintonii</i>	12	7	11	7	23	14	0.967	0.0545	1.2	5	0.8	7	4.3	MPS	2	HPS
<i>P. magnilobatus</i>	16	7	2	1	18	8	0.985	0.0099	0	1.3	0.4	3.2	1.6	HPS	2	HPS
<i>P. xanthotrichus</i>	11	8	38	30	49	38	0.993	0.014	0	7.9	9	10	9	NFCR	5	MPS
<i>P. zimbabweensis</i>	10	5	16	13	26	17	0.985	0.0375	0	7.6	4.4	10	7.3	LPS	6	LPS
Not classified																
<i>P. glabellus</i>	128	42	15	10	160	42	0.995	0.0042	0	2.4	1.9	6.7	3.7	MPS	5	MPS
<i>P. microcarpus</i>	223	161	51	35	274	193	0.981	0.007	0	1.8	3.4	10	5.1	LPS	4	MPS

¹Number of herbarium specimens. ²Number of genebank accessions. ³Refers to the number of populations (unique locations identified) represented by the set of samples. ⁴Prioritization of taxa is done as follows: HPS: High priority species, MPS: Medium priority species, LPS: Low priority species, NFCR: No further urgent conservation required. FPS indicates the result of the method proposed in this paper, and EPS indicates the prioritization given by expert knowledge (based on Daniel G. Debouck's expertise in *Phaseolus*). +Indicates that the taxon had no genebank accessions and no herbarium samples with coordinates or location data; ++Indicates a taxon for which a Maxent model was not possible and for which 0-few genebank accessions were available; #Indicates a taxon with some genebank accessions but no or limited herbarium samples with coordinates or location data. These taxa are listed as HPS for further collecting in order to inform the gap analysis.

REFERENCES

- Acosta-Gallegos, J.A., Kelly, J.D., Gepts, P. (2007). Prebreeding in common bean and use of genetic diversity from wild germplasm. *Crop Sci.* 47 (3-12), S-44-S-59.
- Brooks, T.M., Mittermeier, R.A., da Fonseca, G.A.B., Gerlach, J., Hoffmann, M., Lamoreux, J.F., *et al.* (2006). Global biodiversity conservation priorities. *Science* 313 (5783), 58-61.
- Burke, M.B., Lobell, D.B., Guarino, L. (2009). Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environ. Chang.* 19 (3) (8), 317-325.
- Busby, J.R. (1991). "BIOCLIM – a bioclimatic analysis and prediction system," in: *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*, eds. C.R. Margules, M.P. Austin (Canberra: Commonwealth Scientific and Industrial Research Organisation [CSIRO]). 64-68.
- Challinor, A.J., Wheeler, T.R., Craufurd, P.Q., Ferro, C.A.T., Stephenson, D.B. (2007). Adaptation of crops to climate change through genotypic responses to mean and extreme temperatures. *Agr. Ecosyst. Environ.* 119, 190-204.
- Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiv. Conserv.* 19, 883-899.
- Debouck, D.G. (1999). "Diversity in *Phaseolus* species in relation to the common bean, Chapter 2," in: *Common Bean Improvement in the Twenty-first Century*, ed. S.P. Singh (Dordrecht: Kluwer Academic Publishers), 25-52.
- Debouck, D.G. (2000). "Biodiversity, ecology, and genetic resources of *Phaseolus* beans- seven answered and unanswered questions," in: *Wild Legumes*, ed. K. Oono (Tsukuba, Ibaraki: Ministry of Agriculture, Forestry and Fisheries, and National Institute of Agrobiological Resources), 95-123.
- Debouck, D.G. (2009). *Cahiers de Phaséologie – Synonymie* (Cali : International Center for Tropical Agriculture [CIAT]). http://isa.ciat.cgiar.org/urg/urgweb_folder/files/unitfiles/CahiersdePhaseologieSynonymie.pdf [Accessed: 10 November 2009]
- Debouck, D.G., Smartt, J. (1995). "Beans," in: *Evolution of Crop Plants*, eds. J. Smartt, N.W. Simmonds (Singapore: Longman Scientific and Technical), 287-294.
- Delgado-Salinas, A. (1985). *Systematics of the Genus Phaseolus (Leguminosae) in North and Central America*. Ph.D. Thesis (Austin: Univ. of Texas).
- Delgado-Salinas, A., Carr, W.R. (2007). *Phaseolus texensis* (Leguminosae: Phaseolinae): a new species from the Edwards Plateau of Central Texas. *Lundellia* 10, 11-17.
- Delgado-Salinas, A., Turley, T., Richman, A., Lavin, M. (1999). Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Syst. Bot.* 24, 438-460.
- Dormann, C.F. (2006). Promising the future? Global change projections of species distributions. *Basic Appl. Ecol.* 8, 387-397.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- FAO (Food and Agriculture Organization of the United Nations). (1997). *State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO). http://apps3.fao.org/wiews/wiewspage.jsp?i_l=EN&show=SOW1.
- FAO (Food and Agriculture Organization of the United Nations). (2009a). FAOSTAT. <http://faostat.fao.org/site/339/default.aspx>.
- FAO (Food and Agriculture Organization of the United Nations) (2009b). *Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (draft) (Rome: FAO). <ftp://ftp.fao.org/docrep/fao/meeting/017/ak528e.pdf>.

- Fitzpatrick, M.C., Hargrove, W.W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodiv. Conserv.* 18, 2255-2261.
- Freytag, G.F., Debouck, D.G. (2002). *Taxonomy, Distribution, and Ecology of the Genus Phaseolus in North America, Mexico and Central America* (Ft. Worth: Botanical Research Institute).
- Garvin, F.D., Weeden, N. (1994). Isozyme evidence supporting a simple geographic origin for domesticated tepary bean. *Crop Sci.* 34, 1390-1395.
- Gepts, P. (2001). "Phaseolus vulgaris (Beans)," in: *Encyclopedia of Genetics* (Academic Press), 1444-1445.
- Graham, C., Hijmans, R.J. (2006). A comparison of methods for mapping species range and species richness. *Global Ecol. Biogeog.* 15, 578-587.
- Graham, C., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., *et al.* (2008). The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239-247.
- Guralnick, R.P., Wieczorek, J., Beaman, R., Hijmans, R.J., and the BioGeomancer Working Group. (2006). Biogeomancer: automated georeferencing to map the world's biodiversity data. *PLoS Biology* 4(11), 1908-1909.
- Hajjar, R., Hodgkin, T. (2007). The use of wild relatives in crop improvement: a survey of developments over the last 20 years. *Euphytica* 156, 1-13.
- Harlan, J.R., de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20 (4), 509-517.
- Hawkins, B., Sharrock, S., Havens, K. (2008). *Plants and Climate Change: Which Future?* (Richmond: Botanic Gardens and Conservation International).
- Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29, 773-785.
- Hijmans, R.J., Graham, C. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob. Change Biol.* 12, 2272-2281.
- Hijmans, R.J., Spooner, D. (2001). Geographic distribution of wild potato species. *Am. J. Bot.* 88, 2101-2112.
- Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965-1978.
- Hijmans, R.J., Guarino, L., Cruz, M., Rojas, E. (2001). Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genet. Res. Newsl.* 127, 15-19.
- Hijmans, R.J., Guarino, L., Jarvis, A., O'Brien, R., Mathur, P., Bussink, C., *et al.* (2005b). *DIVA-GIS version 5.2 manual*. <http://www.diva-gis.org/Materials.htm>.
- Hijmans, R.J., Spooner, D.M., Salas, A.R., Guarino, L., de la Cruz, J. (2002). *Atlas of Wild Potatoes* (Rome: International Plant Genetic Resources Institute).
- Intergovernmental Panel on Climate Change. (2007). *IPCC Fourth Assessment Report: Climate Change 2007* (Geneva: IPCC).
- Jarvis, A., Ferguson, M., Williams, D., Guarino, L., Jones, P., Stalker, H., *et al.* (2003). Biogeography of wild *Arachis*: Assessing conservation status and setting future priorities. *Crop Sci.* 43, 1100-1108.
- Jarvis, A., Guarino, L., Williams, D., Williams, K., Hyman, G. (2002). Spatial analysis of wild peanut distributions and the implications for plant genetic resource conservation. *Plant Genet. Res. Newsl.* 131, 29-35.
- Jarvis, A., Yeaman, S., Guarino, L., Tohme, J. (2005). "The role of geographic analysis in locating, understanding, and using plant genetic diversity," in: *Molecular Evolution: Producing the Biochemical Data, Part B*, ed. E. Zimmer (New York: Elsevier), 279-298.

- Jimenez-Valverde, A., Lobo, J.M., Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Divers. Distrib.* 14, 885-890.
- Jones, P.B., Beebe, S.E., Tohme, J., Galwey, N.W. (1997). The use of geographical information systems in biodiversity exploration and conservation. *Biodiv. Conserv.* 6, 947-958.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57(4), 625-639.
- Kipe-Nolt, J.A., Montealegre, C.M., Tohme, J. (1992). Restriction of nodulation by the broad host range *Rhizobium tropici* strain CIAT899 in wild accessions of *Phaseolus vulgaris*. *New phytol.* 120, 489-94.
- Kornegay, J., Cardona, C. (1991). "Breeding for insect resistance," in *Common Beans: Research for Crop Improvement*, eds. A. van Schoonhoven, O. Voysest (Wallingford: CAB International), 619-41.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28 (3), 385-393.
- Lobo Burle, M., Torres Cordeiro, C.M., Fonseca, J.R., Martins Palhares de Melo, L.A., de Belem das Neves Alves, R., Abadie, T. (2003). Characterization of germplasm according to environmental conditions at the collecting site using GIS- two case studies from Brazil. *Plant Gen. Resour. Newsletter* 135, 1-11.
- Lobo, J.M., Jimenez-Valverde, A., Hortal, J. (2010). The uncertain nature of absences and their importance in species distribution modeling. *Ecography* 33, 103-114.
- Lobo, J.M., Jiménez-Valverde, A., Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17(2), 145-151.
- Loiselle, B.A., Jorgensen, P.M., Consiglio, T., Jimenez, I., Blake, J.G., Lohmann, L.G., *et al.* (2008). Predicting species distributions from herbarium collections, does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35(1), 105-116.
- Maxted, N., Kell, S. (2009). *Establishment of a global network for the in situ conservation of crop wild relatives: status and needs* (Rome: FAO Commission on Genetic Resources for Food and Agriculture).
- Maxted, N., Dulloo, E., Ford-Lloyd, B.V., Iriondo, J.M., Jarvis, A. (2008). Gap analysis: a tool for complementary genetic conservation assessment. *Divers. Distrib.* 14 (6), 1018-1030.
- Maxted, N., Ford-Lloyd, B.V., Jury, S.L., Kell, S.P., Scholten, M.A. (2006). Towards a definition of a crop wild relative. *Biodiv. Conserv.* 15 (8), 2673-2685.
- Maxted, N., Mabuza-Dlamini, P., Moss, H., Padulosi, S., Jarvis, A., Guarino, L. (2004). *An Ecogeographic Study. African Vigna* (Rome: International Plant Genetic Resources Institute).
- McClean, P., Gepts, P., Kami, J. (2004). "Genomics and genetic diversity in common bean," in: *Legume Crop Genomics*, eds. E. Wilson, H.T. Stalker, E.C. Brummer (Champaign: AOCS), 60-82.
- Meilleur, B.A., Hodgkin, T. (2004). *In situ* conservation of crop wild relatives: status and trends. *Biodiv. Conserv.* 13 (4), 663-684.
- Munoz, L.C., Duque, M.C., Debouck, D.G., Blair, M.W. (2006). Taxonomy of tepary bean and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. *Crop Sci.* 46, 1744-1754.
- Nabhan, G.P. (1990). *Wild Phaseolus Ecogeography in the Sierra Madre Occidental, Mexico: Areographic Techniques for Targeting and Conserving Species Diversity. Systematic and Ecogeographic Studies on Crop Gene pools* 5 (Rome: International Board of Plant Genetic Resources).
- Parthasarathy, U., Saji, K.V., Jayarajan, K., Parthasarathy, V.A. (2006). Biodiversity of *Piper* in South

- India – application of GIS and cluster analysis. *Current Science* 91 (5).
- Phillips, S.J. (2008). Transferability, sample selection bias and background data in presence-only modeling: a response to Peterson *et al.* (2007). *Ecography* 31, 272-278.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231-259.
- Phillips, S.J., Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and comprehensive evaluation. *Ecography* 31, 161-175.
- Prescott-Allen, C., Prescott-Allen, R. (1986). *The First Resource: Wild Species in the North American Economy* (New Haven: Yale University).
- Salcedo, C.J., Arroyave, J.A., Toro, O., Debouck, D.G. (2006). *Phaseolus novoleonensis*, a new species (Leguminosae, Phaseolinae) from the Sierra Madre Oriental, Nuevo León, Mexico. *Novon* 16 (1), 106-112.
- Salcedo, C.J., Lépiz, R.I., Castañeda, N.A., Ocampo, C.N., Debouck, D.G. (2009). Additional observations about *Phaseolus rotundatus* (Fabaceae), an endemic bean species from western Mexico. *J. Bot. Res. Inst. Texas* 3(2), 751-762.
- Schinkel, C., Gepts, P. (1988). Phaseolin diversity in the tepary bean *Phaseolus acutifolius* A Gray. *Plant Breeding* 101, 292-301.
- Shelley-Dessert, K.C., Bliss, F.A. (1991). “Genetic improvement in food quality factors,” in: *Common Beans: Research for Crop Improvement*, eds. A. van Schoonhoven, O. Voysest (Wallingford: CAB International), 649-671.
- Singh, R.J., Jauhar, P. (2005). *Genetic Resources, Chromosome Engineering, and Crop Improvement: Grain Legumes, Volume 1* (CRC Press).
- Singh, S.P. (2001). Broadening the genetic base of common bean cultivars: a review. *Crop Sci.* 41, 1659-1675.
- Smarrt, J. (1981). Gene pools in *Phaseolus* and *Vigna* cultigens. *Euphytica* 30, 445-449.
- Smolik, M.G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J., *et al.* (2010). Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *J. Biogeogr.* 37(3), 411-422.
- Tanksley, S.D., McCouch, S.R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277 (5329), 1063-1066.
- Termansen, M., McClean, C.J., Preston, C.D. (2006). The use of genetic algorithms and Bayesian classification to model species distributions. *Ecol. Model.* 192, 410-414.
- The Consultative Group on International Agricultural Research (CGIAR), System-wide Information network for Genetic Resources (SINGER). (2009). <http://singer.grinfo.net/>
- Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, I.C. (2005). Climate change threats to plant diversity in Europe. *Proc. Natl. Acad. Sci.* 102 (23), 8245.
- USDA, ARS, National Genetic Resources Program (NGRP). (2009). Germplasm Resources Information Network - (GRIN) [Online Database]. National Germplasm Resources Laboratory, Beltsville, Maryland. <http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl>.
- VanDerWal, J., Shoo, L.P., Graham, C.H., Williams, S.E. (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecol. Model.* 220, 589-594.
- Veloz, S.D. (2009). Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *J. Biogeogr.* 36, 2290-2299.
- Wisn, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., and NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763-77.

CHAPTER 5

Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I. series Batatas*]

Colin K. Khoury^{1,2}, Bettina Heider³, Nora P. Castañeda-Álvarez^{1,4}, Harold A. Achicanoy¹,
Chrystian C. Sosa¹, Richard E. Miller⁵, Robert W. Scotland⁶, John R.I. Wood⁶,
Genoveva Rossel³, Lauren A. Eserman⁷, Robert L. Jarret⁸, G. Craig Yencho⁹,
Vivian Bernau^{1,10}, Henry Juarez³, Steven Sotelo¹, Stef de Haan^{1,3}, and Paul C. Struik²

¹ International Center for Tropical Agriculture (CIAT), Cali, Colombia.

² Centre for Crop Systems Analysis, Wageningen University, Wageningen, Netherlands.

³ CGIAR Research Program on Roots, Tubers and Bananas (RTB), International Potato Center, Lima, Peru.

⁴ School of Biosciences, University of Birmingham, Birmingham, UK.

⁵ Department of Biological Sciences, Southeastern Louisiana University, Hammond, LA, USA.

⁶ Department of Plant Sciences, University of Oxford, Oxford, UK.

⁷ Plant Biology Department, University of Georgia, Athens, GA, USA.

⁸ Plant Genetic Resources Conservation Unit, United States Department of Agriculture- Agricultural Research Service, Griffin, GA, USA.

⁹ Department of Horticultural Science, North Carolina State University, Raleigh, NC, USA.

¹⁰ Department of Horticulture and Crop Science, The Ohio State University, Columbus, Ohio, USA.

Abstract

Crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I. series Batatas*] have the potential to contribute to breeding objectives for this important root crop. Uncertainty in regard to species boundaries and their phylogenetic relationships, the limited availability of germplasm with which to perform crosses, and the difficulty of introgression of genes from wild species has constrained their utilization. Here we compile geographic occurrence data on relevant sweetpotato wild relatives and produce potential distribution models for the species. We then assess the comprehensiveness of *ex situ* germplasm collections, contextualize these results with research and breeding priorities, and use ecogeographic information to identify species with the potential to contribute desirable agronomic traits. The fourteen species that are considered the closest wild relatives of sweetpotato generally occur from the central United States to Argentina, with richness concentrated in Mesoamerica and in the extreme southeastern United States. Currently designated species differ among themselves and in comparison to the crop in their adaptations to temperature, precipitation, and edaphic characteristics and most species also show considerable intraspecific variation. With 79% of species identified as high priority for further collecting, we find that these crop genetic resources are highly under-represented in *ex situ* conservation systems and thus their availability to breeders and researchers is inadequate. We prioritize taxa and specific geographic locations for further collecting in order to improve the completeness of germplasm collections. In concert with enhanced conservation of sweetpotato wild relatives, further taxonomic research, characterization and evaluation of germplasm, and improving the techniques to overcome barriers to introgression with wild species are needed in order to mobilize these genetic resources for crop breeding.

Keywords: Crop diversity, Crop improvement, Crop wild relatives, Food security, Gap analysis, Plant genetic resources

INTRODUCTION

Sweetpotato [*Ipomoea batatas* (L.) Lam.] counts among the world's most important root crops, grown on at least eight million hectares in 114 countries worldwide, with particular significance to food supplies in the tropics and subtropics of East and Southeast Asia, and Sub-Saharan Africa. The success of the crop in these regions is due to its adaptability to a wide range of agro-ecological conditions, ease of propagation from cuttings, low input cultivation requirements, and high productivity as well as nutritive value (Woolfe 1992). The crop can be cultivated from humid to semi-arid conditions, from sea level to 3000 m.a.s.l. (Huaman 1987), and can translocate photosynthetic products to the storage roots throughout the growing season, thereby mitigating negative effects of temporary adverse conditions (Kays 1985). Sweetpotato produces among the highest amounts of edible energy per hectare of all major food crops (de Vries *et al.* 1967), and is an important source of vitamin A and C, calcium, iron, and a number of essential amino acids (Kays & Kays 1998, Tumwegamire *et al.* 2011). In addition to human consumption of the storage roots and young leaves, the crop is utilized for animal feed, fuel, and starch production.

In Sub-Saharan Africa, sweetpotato is predominantly cultivated in small plots characterized by low fertility and drought-prone soils, producing relatively good yields with low inputs and minimal labor costs. The crop has recently become the focus of targeted bio-fortification for enhanced vitamin A. Orange-fleshed varieties have been bred with 50-fold more β -carotene than standard varieties and these newly-released varieties rank first among roots and tubers in Sub-Saharan Africa for their nutritional quality (Low *et al.* 2007, Hotz *et al.* 2012). Given its adaptability, low-external input requirements, nutritional quality, and improvement potential, it is not surprising that sweetpotato has become a priority in crop based strategies for enhancing food security in the tropics (Pfeiffer & McClafferty 2007, Bill and Melinda Gates Foundation 2011, Bouis & Islam 2012).

The full potential of sweetpotato is far from realized, with particularly large yield gaps (ca. 20 t ha⁻¹) remaining across rain-fed Sub-Saharan Africa due to a range of biotic and abiotic constraints, especially sweetpotato virus disease (SPVD) and sweetpotato weevils (SPW), as well as susceptibility to drought (Sutherland 1986, Valverde *et al.* 2007, Ngailo *et al.* 2013, Kivuva *et al.* 2015a, b). SPVD is a severe constraint on the continent, caused by the synergistic interaction of two viruses transmitted by whiteflies and aphids, causing yield losses of up to

98% under severe infections (Ngailo *et al.* 2013). SPW (*Cylas* spp.) are the most devastating insect pests of the crop. *Cylas formicarius elegantulus* Summers, *C. puncticollis* Boheman and *C. brunneus* Fabricius can cause yield losses of between 67% and 100% in Sub-Saharan Africa (Smit 1997). The concealed feeding behavior, oviposition, and larval development of the weevils in the storage roots make their control very difficult, necessitating the development of improved management options, in particular via enhanced genetic resistance. Drought, and the compounding effect of increasing heat on drought, is a rising concern particularly in regions undergoing significant climatic change, both due to its direct effect on productivity (Low *et al.* 2009, Schafleitner *et al.* 2010) as well as to its association with increased severity of damage from SPW and SPVD (Munyiza *et al.* 2007, Mwololo *et al.* 2007). Lack of drought tolerance in high β -carotene sweetpotato varieties has led to constraints in their adoption (Mwanga & Ssemakula, 2011).

Crop wild relatives (CWR) are increasingly being recognized for their potential to contribute valuable traits to breeding programs (Feuillet *et al.* 2008, Guarino & Lobell 2011, Dempewolf *et al.* 2014). CWR have provided breeders with genes for pest and disease resistance, abiotic stress tolerance, and quality traits in an ever increasing number of food crops, such as banana, barley, bean, cassava, chickpea, maize, lettuce, oat, potato, rice, sugarcane, sunflower, tomato, and wheat, among others (Xiao *et al.* 1996, Hajjar & Hodgkin 2007, McCouch *et al.* 2007, Khoury *et al.* 2010). Yet despite the successful history of contribution to the improvement of major crops, systematic approaches to the use of CWR in the breeding programs of a number of important staples, including sweetpotato, remain underdeveloped.

The morning glory tribe Ipomoeae contains approximately 650-900 species and includes the genus *Ipomoea* and nine other related genera (Wilkin 1999, Mabberley 2008). Although many genera, subgenera, and sections of the Ipomoeae are not monophyletic in phylogenetic analyses, *Ipomoea* series *Batatas* (Choisy) D.F. Austin, which contains sweetpotato and 14 closely related CWR (Austin 1978, McDonald & Austin 1990), does form a monophyletic lineage (Miller *et al.* 1999, McDonald *et al.* 2011, Eserman *et al.* 2014). These species include wild *I. batatas* (L.) Lam. [including *Ipomoea batatas* var. *apiculata* (Martens and Galeotti) McDonald and Austin], *Ipomoea cordatotriloba* Dennstedt, *Ipomoea cynanchifolia* Meisn., *Ipomoea grandiflora* (Dammer) O'Donnell, *Ipomoea lacunosa* L., *Ipomoea leucantha* Jacquin, *Ipomoea littoralis* Blume, *Ipomoea ramosissima* (Poir.) Choisy, *Ipomoea splendor-sylvae*

House, *Ipomoea tabascanana* McDonald and Austin, *Ipomoea tenuissima* Choisy, *Ipomoea tiliacea* (Willd.) Choisy in D.C., *Ipomoea trifida* (H.B.K.) G. Don, and *Ipomoea triloba* L.

Many sweetpotato CWR can be hybridized with the crop through controlled pollinations, somatic cell, and/or ovule culture techniques (Diaz *et al.* 1996). Crosses involving *I. tabascanana*, *I. trifida*, *I. triloba*, *I. littoralis*, *I. grandifolia*, *I. lacunosa*, *I. leucantha*, and wild *I. batatas* in particular have resulted in relatively viable progeny (Nimmakayala *et al.* 2011). The wild conspecific as well as *I. trifida* have been documented for their contribution to increases in protein and starch content, and nematode and SPW resistance (Iwanaga 1988, Shiotani *et al.* 1994), although there is uncertainty for some material as to whether they may actually have been feral forms of the cultivar (Nimmakayala *et al.* 2011). Species that have been explored for potential traits of use in crop improvement include *I. trifida* and *I. littoralis* for yield and SPW, scab [*Elsinoë batatas* (Saw.) Viegas et Jenkins], and black rot disease (*Ceratocystis fimbriata* Ell. et Halst.) resistance; *I. grandifolia* for sweetpotato stem nematode and SPVD resistance; and *I. triloba* for drought tolerance, root rot resistance, and foliar fungal disease resistances (Iwanaga 1988, Jarret *et al.* 1992, Komaki 2004, Zhang & Liu 2005, Nimmakayala *et al.* 2011). Challenges in the creation of viable progeny between the CWR and the cultivated species are not insignificant, though, due to differences in ploidy and interspecific incompatibility (Martin 1970, 1982, Teramura 1979, Shiotani *et al.* 1994, Lu & Li 1992, Diaz *et al.* 1996, Komaki 2004, Nimmakayala *et al.* 2011).

A lack of basic knowledge about boundaries between species within *I. series Batatas* and a dearth of diagnostic characters enabling differentiation of taxa - to facilitate reliable and accurate species identification - is a fundamental stumbling block constraining the utilization of sweetpotato CWR (Austin 1978, Austin 1988, Jarret *et al.* 1992, Diaz *et al.* 1996, McDonald *et al.* 2011, Nimmakayala *et al.* 2011, Eserman *et al.* 2014). Needed studies have been delayed in part due to the absence of plant materials for research, in particular the availability of specimens with flowers and ripe fruits. Studies that have been performed have often been based upon limited sampling (e.g., single accessions for *I. littoralis* and *I. tabascanana*).

Many unanswered questions regarding the relationships of CWR to sweetpotato potentially impact the efficiency of breeding strategies for the crop. For example, do species such as *I. tabascanana* and *I. tenuissima* represent distinct taxa (hybrid or otherwise) or rather rare variants

of the crop? What is the range and genetic diversity present in truly wild forms of the crop conspecific, compared to feral escapees? How accurate are the classifications of species with highly disjunct distributions (e.g., *I. cordatotriloba*)? Are there as yet unrecognized cryptic species within *I. series Batatas* [e.g., *I. 'austinii'* (Duncan & Rausher 2013)]? What are the lineages and genetic resources potential of purported hybrid species (i.e., *I. leucantha* and *I. grandifolia*)? What are the geographic locations of new variation being generated through hybridization among sweetpotato CWR?

The investigation, conservation, and availability of genetic resources of sweetpotato provide a foundation for the crop's long term viability and for its potential for improvement. To contribute to these objectives, we analyzed the comprehensiveness of *ex situ* conservation of sweetpotato CWR through a series of questions: (a) what constitutes a potentially useful wild relative of sweetpotato?, (b) where are these species encountered?, (c) what is the state of conservation and availability of these species to researchers, and what are the highest taxonomic and ecogeographic priorities for further collecting? And finally, (d) what traits do sweetpotato CWR possess that may be valuable to crop improvement?

MATERIALS AND METHODS

Identification of target CWR species and occurrence data compilation

The CWR of sweetpotato analyzed in this study were selected based upon recent and historical taxonomic and molecular phylogenetic research (Austin 1978, Austin & Huaman 1996, Austin 1997, Miller *et al.* 1999, McDonald *et al.* 2011, GRIN 2013, Eserman *et al.* 2014), identifying those wild species with a relatively close phylogenetic relationship to the crop (i.e., members of *I. series Batatas*). We included all 14 wild species comprising the series in the analysis.

Domesticated sweetpotato *I. batatas* (6x) has been proposed as originating from interspecific hybridization involving *I. trifida* (2x, 4x, 6x), *I. littoralis* (2x), and/or *I. leucantha* (2x) (Nishiyama 1971, Nishiyama 1982, Austin 1988). The species most closely related to the crop have been posited to be *I. trifida* followed by *I. tabascana* (4x) (Srisuwan *et al.* 2006, Nimmakayala *et al.* 2011). Following the genepool concept of Harlan and de Wet (1971) and due to ploidy incompatibility with the cultivated species, the putative closest related taxa to sweetpotato are placed in the secondary genepool: wild forms of *I. batatas* (4x), *I. trifida*, *I. littoralis*, and *I. tabascana* (Jarret *et al.* 1992, Jarret & Austin 1994, Rajapakse *et al.* 2004,

GRIN 2013). Species classified as tertiary wild relatives include: *I. cordatotriloba* (syn. *Ipomoea trichocarpa* Elliott), *I. cynanchifolia*, *I. grandifolia*, *I. lacunosa*, *I. leucantha*, *I. ramosissima*, *I. splendor-sylvae* (syn. *Ipomoea umbraticola* House), *I. tenuissima*, *I. tiliacea*, and *I. triloba* (Jarret & Austin 1994, Huang *et al.* 2002, Rjapakse *et al.* 2004, GRIN 2013).

Occurrence records for these species were acquired from online biodiversity, herbaria, and germplasm databases; through communications with herbaria and genebank managers, and other crop researchers; and via direct recording of provenance data during visits to selected herbaria (Supplementary Table 1). Germplasm data were obtained from repositories that provide straightforward access to genetic resources and associated data to the global research community through online information systems. The occurrence data were then compiled in a standardized format, nomenclature was checked against GRIN Taxonomy for Plants (2013) and The Plant List (2010), and duplicate records were eliminated. Existing coordinates were cross-checked to country and being on land (Hijmans *et al.* 1999), and records with locality information but no coordinates were geo-referenced using the Google Maps Geocoder v.3 (2013) application programming interface. Occurrence data were mapped, iteratively evaluated for correctness, and further processed in order to form a final dataset of improved taxonomic and spatial accuracy.

Challenges in using and in improving the large quantities of occurrence data now available from online resources such as the Global Biodiversity Information Facility (GBIF) have been noted (Gaiji *et al.* 2013), including geographic and nomenclatural data quality and the slow speed with which aggregated datasets are updated (Mesibov 2013, Otegui *et al.* 2013, Hjarding *et al.* 2014). In addition, particular caution must be applied to the occurrence records used in the current paper as ongoing work (unpublished data) indicates that many *Ipomoea* occurrence records in such online databases are identified as synonyms, excluded or invalid names, and that many valid names were applied to specimens well outside of species known ranges. We have identified some of these obvious errors but until all specimen records are correctly identified and checked against an accurate taxonomy these data must be treated with caution.

A total of 5,614 occurrence records for the 14 taxa were included in potential distribution modeling and/or in the conservation analysis, including 749 germplasm records sourced from four genebanks, and 4,865 herbarium and other occurrence reference records sourced from 42

providers. Records per species ranged from eight (*I. tabascana*) to 1,409 (*I. trifida*). Of these, 3,650 records containing unique cross-checked coordinates were used to model species potential distributions and to locate the original collecting site of existing germplasm accessions.

Species potential distribution modeling

A potential distribution model for each species was calculated using the maximum entropy (Maxent) algorithm (Phillips *et al.* 2006), with a set of ecogeographic variables and unique species presence records as inputs. We chose Maxent due to its extensive application in predicting species distributions (Elith *et al.* 2006; Phillips & Dudik 2008, Costa *et al.* 2010), including those for wild relatives (Ramírez-Villegas *et al.* 2010, Conolly *et al.* 2012, Khoury *et al.* 2015). We performed modeling at a resolution of 2.5 arc-minutes (~ 5 km² cell size at the equator), employing 10,000 background points for model training over the combined distributional range of the sweetpotato CWR. Ecogeographic inputs included altitude and 19 bioclimatic variables from the WorldClim database (Hijmans *et al.* 2005), and seven major edaphic drivers of plant species distributions with consistent data coverage throughout the range of the sweetpotato CWR species, obtained from ISRIC-World Soil Information (Hengl *et al.* 2014) (Supplementary Table 2). For the edaphic variables we calculated a weighted mean across 0-5, 5-15, 15-30, 30-60, and 60-100 cm soil depth values in order to derive a single data value for 0-100 cm. We then resampled the 30 arc-seconds resolution data to form 2.5 arc-minutes inputs aligned with the WorldClim datasets.

Potential distribution models were produced by calculating the mean of replicates ($k = 5$), and clipped by measuring the shortest distance between the receiver operating characteristic curve (ROC-curve) and the top-left corner of the plot (Liu *et al.* 2005). Models were constrained per species by a native range defined at the country level as given in GRIN Taxonomy for Plants (GRIN 2013), in order to focus prioritization recommendations on those regions with populations exhibiting long-term adaptations to local ecogeographic conditions. We further cross-validated and refined occurrence data based upon our knowledge of native distributions, constraining localities for wild *I. batatas* to Mexico south to Peru, and not in the Caribbean; *I. leucantha* to the USA and Mexico, and *I. littoralis* to points within 100 km of the ocean.

In order to derive robust distribution models for each species, we analyzed Maxent results

across three groups of ecogeographic inputs: a) the full set of 19 bioclimatic variables (Ramírez-Villegas *et al.* 2010); b) the bioclimatic variables, altitude, and the additional set of seven edaphic variables, totaling 27 input variables; and c) a species-specific derivation of the most important drivers of distribution based upon presence data, further refined by removing highly correlated variables. For the ecogeographic variables in the species-specific subset method, we utilized a non-linear iterative partial least squares (NIPALS) algorithm to perform a principal-component analysis (PCA), as NIPALS has the potential to handle data arrays in which the number of observations is less than the number of input variables. We then identified those variables with the greatest contribution (> 0.7 or < -0.7) to the first two principal components per species. Finally, we used a variance inflation factor (VIF) to identify the variables with a low degree of collinearity (see Supplementary Table 3 for a list of variables identified per species). A calibrated area under the ROC curve (cAUC) was obtained to assess the predictive performance of each model (Hijmans 2012, Khoury *et al.* 2015).

The three modeling methods were compared against null models, and the species-specific subset variables method showed the least overall spatial sorting bias among methods (Spearman's p for the 19 variables was 0.65, for 27 variables it was 0.72, and for the subset method it was 0.25), although the differences in median AUC distributions across species for each method were not significant ($p = 0.095$) through a Kruskal-Wallis non-parametric analysis of variance test. Maxent models were subjected to a four-fold assessment process including: a) the 5-fold average area under the ROC curve of test data (ATAUC), b) the standard deviation of the test AUC of the 5 different folds (STAUC), c) the proportion of the potential distribution coverage with standard deviation above 0.15 (ASD15), and d) the cAUC value. Models with ATAUC above 0.7, STAUC below 0.15, ASD15 below 10%, and cAUC exceeding 0.40 were considered accurate and stable (Ramírez-Villegas *et al.* 2010, Khoury *et al.* 2015).

The potential distribution models of the sweetpotato CWR generally performed well in regard to the modeling assessment process. Four species demonstrated low cAUC values and one of these an elevated ASD15 value, indicating greater uncertainty in the models (Supplementary Table 3). Species-specific subset model outputs for taxa with relatively few distinct occurrence points (< 20) (*I. tabascana*, *I. tenuissima*, and *I. cynanchifolia*) lacked sufficient discriminatory power, leading to highly inflated spatial models in comparison to recorded distributions. Potential distribution models for these species were resolved by deriving an ensemble (i.e.,

overlap) of the three input variation methods, verified by researchers knowledgeable in the distribution of the species as more accurately representing the true known distributions. Potential distribution models based upon the species-specific subset variables method were therefore utilized in subsequent analyses for all species with adequate distinct occurrence points (> 20). The ensemble method was used for the three species with limited occurrence data.

Analysis of current *ex situ* conservation and further collecting needs for CWR

We adapted a gap analysis methodology proposed by Ramírez-Villegas *et al.* (2010), combining three metrics used to assess the urgency of further collecting in order to fill gaps in *ex situ* conservation of CWR. The total sample representation of each species in genebank collections was estimated via a sampling representativeness score (SRS), calculated as the number of germplasm samples (G) divided by the total number of samples (G + herbarium samples (H)) (i.e., all other records aside from available genebank accessions).

The sufficiency of geographic coverage of germplasm collections of each species was estimated through a geographic representativeness score (GRS), calculated as the share of the combined total area of circular buffers of 50 km (CA50) placed around existing germplasm collection points compared with the overall potential geographic distribution of the species.

The comprehensiveness of ecological coverage of germplasm collections of each species was estimated through an ecological representativeness score (ERS), calculated by estimating the distinct ecosystem classifications (Olson *et al.* 2001) represented in the CA50 of existing germplasm collection points compared with the diversity of ecosystems in which the overall potential geographic distribution model of the species occurs.

A final priority score (FPS) for further collecting for *ex situ* conservation was assigned to each species by averaging the three gap analysis metrics (SRS, GRS, and ERS). FPS scores were further classified into four categories of urgency for collecting: high priority species (HPS) for taxa whose $0 < FPS \leq 2.5$ or when no germplasm accessions are currently conserved, medium priority species (MPS) when $2.5 < FPS \leq 5$, low priority species (LPS) when $5 < FPS \leq 7.5$, and ‘no further collecting recommended’ (NFCR) when $7.5 < FPS \leq 10$. We produced collecting priorities maps for all species, displaying the geographic areas that have not yet been collected from within the potential distributions of taxa.

The ecogeographic data preparation, species distribution modeling, and gap analysis were written and performed in R v2.15.1 (R Core Team 2013), utilizing packages *maptools* (Bivand & Lewin-Koh 2014), *rgdal* (Bivand *et al.* 2014), *SDMTools* (van der Wal *et al.* 2014), *raster* (Hijmans 2014), *sp* (Pebesma & Bivand 2005, Bivand *et al.* 2013), *dismo* (Hijmans *et al.* 2013), and *plsdepot* (Sanchez 2012). Resulting spatial files were mapped in ArcMap v.10 (ESRI 2011). Collecting priorities spatial files were analyzed using the Zonal Statistics tool in ArcMap to list the countries prioritized for further collecting for *ex situ* conservation.

Expert evaluation of conservation assessment results

In order to validate and/or expose deficiencies in our findings, we subjected the gap analysis numerical and spatial results to an evaluation performed by seven crop experts with experience in the systematics, distribution, and/or conservation status of CWR of sweetpotato: Richard E. Miller, Southeastern Louisiana University; Robert W. Scotland and John R.I. Wood, University of Oxford; Genoveva Rossel, International Potato Center; Lauren A. Eserman, University of Georgia; Robert L. Jarret, USDA-ARS Griffin; and G. Craig Yencho, North Carolina State University. These experts were first asked to provide an evaluation of the sufficiency of germplasm collections per species based only upon their knowledge of the total number of accessions conserved, and geographic as well as environmental gaps. Such an assessment [comparable expert priority score (EPS)] was considered directly comparable to the FPS of the gap analysis results.

A second evaluation score (contextual EPS) based on the entirety of expert knowledge, including additional factors such as threats to species *in situ* and prioritization by usefulness in crop breeding, was collected in order to provide additional information to collecting prioritization efforts. In both cases, an EPS between 0 and 10, aligned with the gap analysis results prioritization scale, was requested. After these steps, experts were shown the gap analysis data and asked to comment on assessed quantitative results, occurrence data, potential distribution models, and maps of collecting priorities. Following these contributions by experts, input occurrence data were further refined by eliminating clearly incorrect points and adjusting country-level native areas, and the potential distribution modeling and gap analyses were re-run using the refined datasets in order to improve the quantitative and spatial outputs. Expert evaluation metrics displayed in the results pertain to a final evaluation of improved gap analysis outputs, performed by five of the experts.

A multiple factor analysis was used in order to compare the various forms of expert evaluation inputs (i.e., comparable expert priority score, contextual expert priority score, evaluation of gap analysis results score, evaluation of occurrence data, evaluation of potential species distribution models, and evaluation of collecting priorities map) with the gap analysis results. An expert evaluation index was created, which estimated the degree of accord between all experts and the gap analysis results for each species, with a scale from 0 (disagreement) to 100 (agreement). Analyses were performed using R package FactoMineR (Husson *et al.* 2009).

Identification of ecogeographic characteristics of CWR

In order to evaluate the pairwise degree of geographic overlap between sweetpotato CWR distribution models, we calculated an overlap measure equal to the frequency of shared 2.5 arc-minute geographic cells between taxa divided by the sum of the total number of cells in which either of the species are present (Kernohan *et al.* 2001, Fieberg & Kochanny 2005). To assess the pairwise degree of ecogeographic niche overlap between species, we used Schoener's index for niche similarity (D) and the adjusted similarity index (I) as outlined in Warren *et al.* (2008), using species distribution models and the species-specific subset of the 27 ecogeographic layers/ensemble models as inputs. Overlap indices range from 0 (no overlap) to 1 (complete overlap) and were implemented in the R package Phyloclim (Heibl 2011).

We utilized ecogeographic information in combination with species presence data in order to identify populations of species with the potential for outstanding adaptations to climatic and/or edaphic conditions of interest to sweetpotato breeding objectives. We assessed the relative importance of the 27 ecogeographic variables (Supplementary Table 2) in explaining the total variation through a PCA after confirming the validity of its use through a Bartlett's test performed in R package psych (Revelle 2015). We created a hierarchical cluster of principal components (HCPC) in order to identify ecogeographic clusters for the sweetpotato wild relatives, driven by those variables demonstrating $\geq 15\%$ difference (\pm) from the mean and a reduction of $\geq 20\%$ from the mean standard deviation exhibited across all species, using R package FactoMineR. Boxplots for each of the 27 ecogeographic variables were created based upon CWR species occurrence data points, displaying the median and variance parameters per species per variable. Comparable ecogeographic variable data for the sweetpotato crop was extracted from area of cultivation maps (Monfreda *et al.* 2008) at a resolution of 5 arc-minutes, with a random sample of 1,000 points weighted by harvested area, taken from the major

cultivation areas in Asia, Africa, and Latin America.

RESULTS

Distributions of the wild relatives of sweetpotato

Sweetpotato CWR were modeled to occur from the central USA to northern Argentina in the Americas, including the Caribbean (Supplementary Figure 1). Species richness was greatest in central Mexico through Central America to the northern Andean region, and in the southeastern USA, with up to nine species potentially overlapping in Mexico from the states of Veracruz through the Yucatan peninsula (Figure 1). The Mexican and Central American regions of distribution represent one of the posited centers of origin and primary diversity of cultivated sweetpotato (Austin 1988, Austin & Huamán 1996, Zhang *et al.* 2000, Gichuki *et al.* 2003, Roullier *et al.* 2013). Northwestern South America, with archeological remains of cultivated sweetpotato from Peru dating back to 8,000 years BP, which are among the oldest recorded domestication events on the continent (Piperno & Pearsall 1998, Shady Solis *et al.* 2001), displayed a considerably lesser but still notable degree of CWR species richness. One Old World species (*I. littoralis*) (Austin 1991) was also modeled to occur in coastal areas of Madagascar, South and Southeast Asia, Australia, and the Pacific region.

Analysis of current *ex situ* conservation and further collecting needs for CWR

Eleven out of the fourteen CWR species were assigned high priority for further collecting due to insufficient genebank accessions in comparison to the total number of occurrence samples (SRS), and to large geographic (GRS) and ecological (ERS) gaps in *ex situ* germplasm collections in comparison to the full potential distributions of the species (Figure 2, Table 1, Supplementary Figure 2). Six of these taxa are currently represented by ≤ 10 accessions conserved *ex situ*, moreover, with few exceptions these accessions lack associated geographic occurrence information (Supplementary Table 3). Total sampling representativeness and geographic coverage of species in germplasm collections were particularly lacking for taxa assessed high priority, while gaps in ecological representativeness were less extreme for some species (e.g., *I. cordatotriloba*, *I. triloba*, and *I. splendor-sylvae*). *Ipomoea grandifolia* and *I. trifida* were assessed as relatively well covered in regard to ecosystems represented *ex situ*, which raised their final priority score into the medium priority for further collecting category. *Ipomoea tabascana* was assessed as of low priority for further collecting due to existing germplasm collections largely covering its very restricted distribution in southern Mexico,

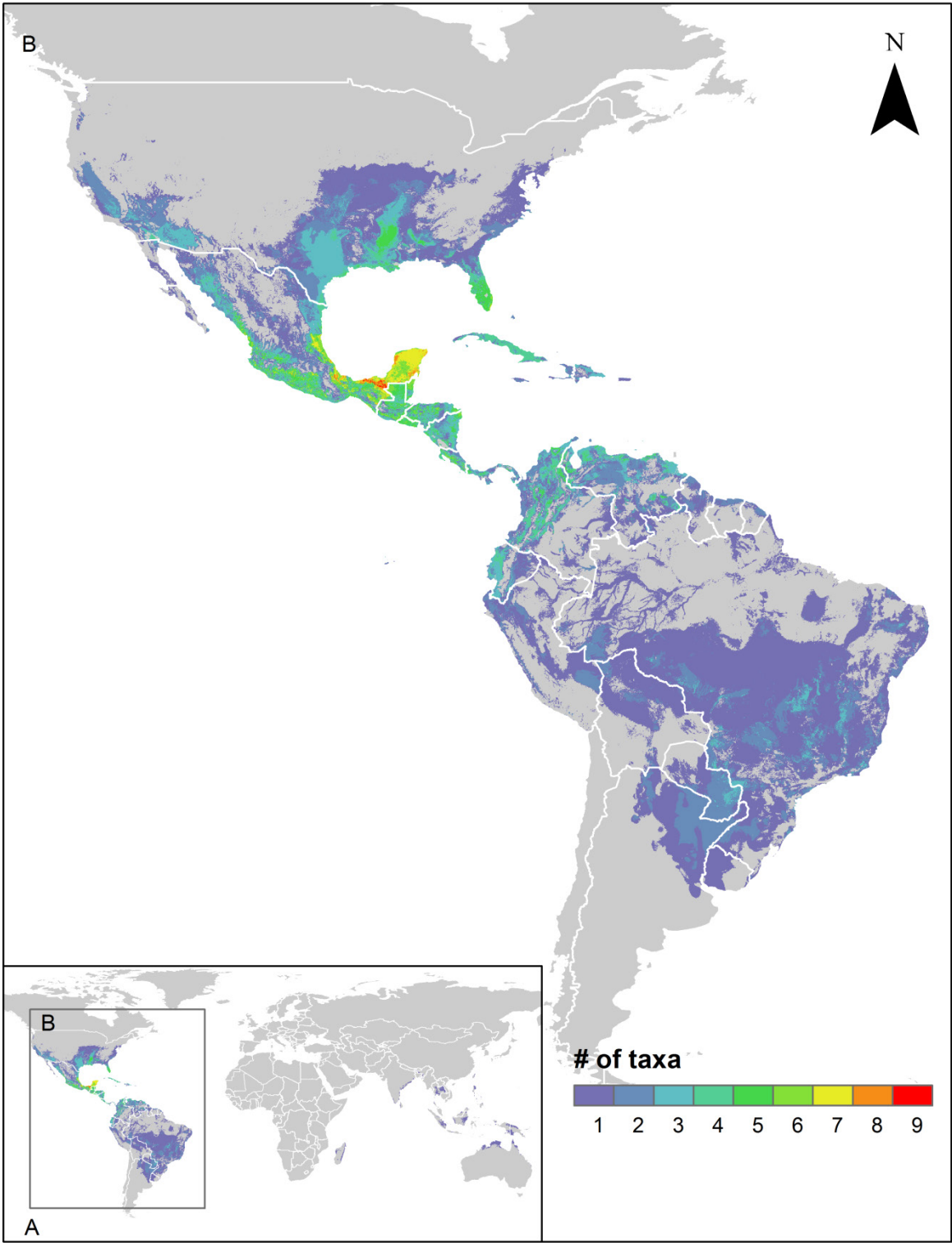


Figure 1. Species richness map for assessed sweetpotato crop wild relative potential distribution models worldwide (A), with concentration of species in the neotropics (B).

resulting in a high GRS score. The mean final priority score across all CWR was 1.75 ± 1.82 .

Paralleling the distribution of species richness of sweetpotato CWR, the regions identified for further collecting of the greatest number of species included central and southern Mexico and the southeastern USA, with up to seven species prioritized for further collecting occurring in the same area (Figure 3). Further collecting priorities were recognized in a total of 50 countries throughout the range of the genepool (Supplementary Figure 3, Supplementary Table 4). Occurrence data, potential distribution models, and collecting priorities maps for all modeled species are available in an interactive map format at <http://www.cwrdiversity.org/distribution-map/>.

Expert evaluation of conservation assessment results

The average of the directly comparable expert priority scores (EPS) across experts was 2.65 (± 1.10) as a mean across species, varying 1.95 points on average from the FPS, with seven taxa designated by the experts as HPS, and seven as MPS (Table 1, Figure 2). For most species, this mean was highly influenced by one or two experts' giving species considerably less priority than the other experts. Species with closest accord between the gap analysis results and the comparable expert analysis included *I. splendor-sylvae*, *I. trifida*, *I. tenuissima*, *I. littoralis*, *I. tileacea*, *I. ramosissima*, and *I. grandifolia*. Taxa with the greatest difference between gap analysis and comparable expert values included *I. tabascanana* and *I. leucantha*.

Two species were assessed by the experts as higher priority for further collecting than the results of the gap analysis. Although *ex situ* collections for the highly restricted distribution of *I. tabascanana* were determined in the gap analysis to be fairly comprehensive (LPS), the experts assigned the species high priority (HPS) for further collecting due to its very limited overall number of germplasm holdings, and to the difficulty in producing viable seed in *ex situ* conservation. *Ipomoea grandifolia* was assessed in the gap analysis as reasonably comprehensively represented in regard to ecosystem diversity, and thus assigned medium priority for further collecting, while the experts gave moderately higher priority to the species.

The contextual expert priority score per species, which also included the expert's knowledge of threats *in situ* as well as usefulness for crop improvement, did not vary widely from the

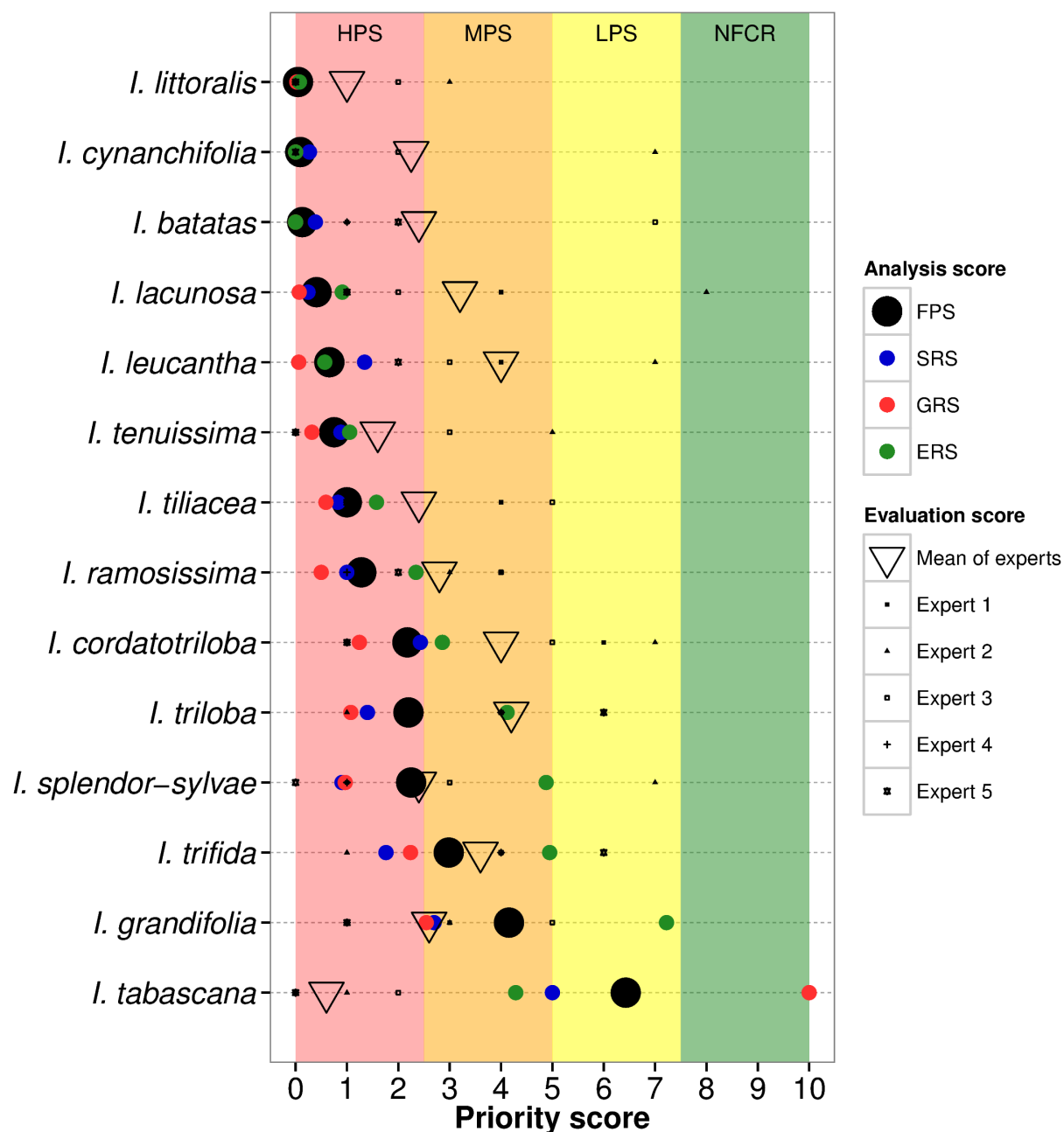


Figure 2. Gap analysis results and comparable expert assessments per sweetpotato crop wild relative. Species are listed by descending priority for further collecting by priority categories [high priority species, HPS (red); medium priority species, MPS (orange); low priority species, LPS (yellow); and no further collecting recommended, NFCR (green)]. The black circle represents the final priority score (FPS) for the species, which is the mean of the sampling representativeness score (SRS), geographic representativeness score (GRS), and ecological representativeness score (ERS).

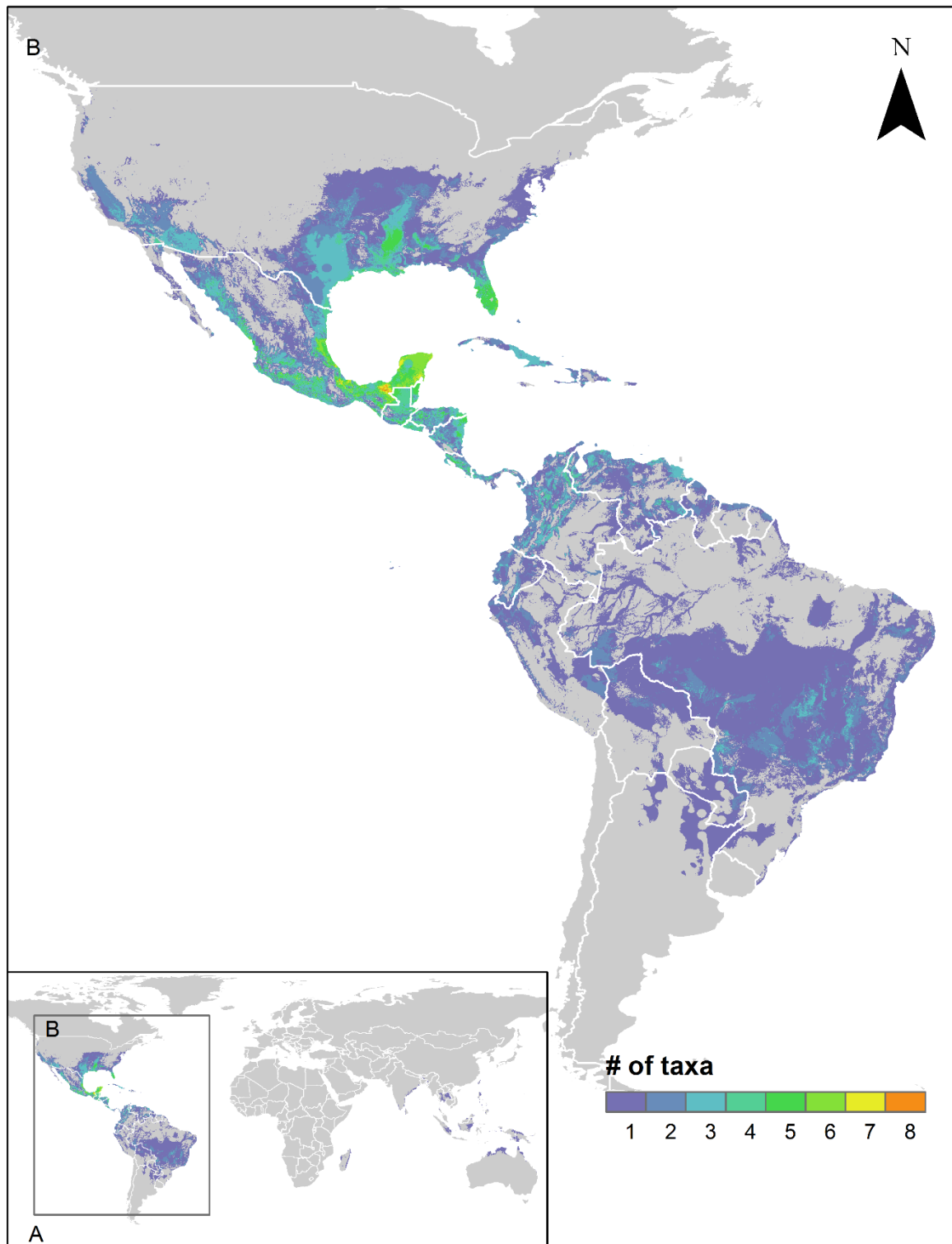


Figure 3. Further collecting priorities hotspots map for high priority (HPS) sweetpotato crop wild relatives. The map displays areas worldwide (A) within the potential distributions of HPS species that have not been previously collected for *ex situ* conservation, with concentration of species in the neotropics (B).

comparable score [mean across all experts and species was 2.76 (\pm 0.85); mean difference between comparable and contextual expert scores across all species and experts was 0.11]. Due largely to differences between the perceptions of relative sufficiency in regard to the overall number of germplasm accessions by the experts versus gap analysis results (e.g., for *I. tabascana*, *I. triloba*, and *I. leucantha*), the comparable and contextual assessments did not correlate well with the gap analysis results for the genepool as a whole (Supplementary Figure 4A-B).

The multiple factor analysis combining the comparable expert priority score, contextual expert priority score, evaluation of gap analysis results score, evaluation of occurrence data, evaluation of potential species distribution models, and evaluation of collecting priorities map, revealed sufficient agreement among experts and variables to justify the use of the expert evaluation index, although variation in expert opinion was notable for many species. Data inputs and resulting distribution models were generally assessed positively as a mean across experts, with eight species receiving very positive index values. Those species with the highest accord between all variables and experts and the gap analysis results included *I. littoralis*, *I. splendorsylvae*, *I. ramosissima*, *I. tenuissima*, and *I. tiliacea*. Those species with the lowest accord included *I. tabascana*, *I. triloba*, and *I. grandifolia* (Supplementary Figure 4). FPS results were particularly influenced by spotty occurrence records for the majority of species, with gaps recognized by the experts.

Identification of ecogeographic characteristics of CWR

The analysis of geographic and ecogeographic accord between pairwise potential species distribution models segregated species well into temperate North American (e.g., *I. lacunosa*, *I. tenuissima*), Mesoamerican (e.g., *I. splendorsylvae*, *I. tabascana*), widely distributed tropical (e.g., *I. triloba*, *I. trifida*), and South American (e.g., *I. grandifolia*, *I. cynanchifolia*) areas (Supplementary Figure 5).

Strong linear relationships were found between bioclimatic variables within the study area, justifying the application of the PCA, with 75.6% of variance explained through four principal components (Supplementary Figure 6). The first principal component (32% of variation) was correlated with temperature extremes and fluctuation. The second component (21.6% of variation) was most occupied by precipitation variables related to drought. The third component

(13.9% of variation), was related with altitude, and the final component (8.1%) with soil texture characteristics.

Species occurrence data segregated into temperate and tropical ecogeographic clusters, with the great majority of species' distributions strongly occurring within a single cluster. Ecogeographic variables most strongly influencing the definition of the temperate cluster (1) included those associated with temperature variation and relatively low precipitation, elevation, and soil organic matter. The most determinant variables in the tropical cluster (2) were related to relatively high and consistent temperatures. Those species displaying a notable proportion of occurrences within both clusters included *I. cynanchifolia* and *I. triloba*, and to a lesser degree *I. leucantha* and *I. ramosissima* (Table 1, Supplementary Figure 6).

Ecogeographic niche assessments of sweetpotato CWR based upon occurrence data points revealed large differences in potential adaptation to temperature, precipitation, and edaphic characteristics (Table 1, Figure 4, Supplementary Figure 7). Such adaptation for many species fell well outside the modeled niche of the cultivated species, particularly for high temperatures both in wet and dry conditions, as well as high precipitation. Species of notable adaptation to high mean annual, monthly, and quarterly temperatures included *I. littoralis*, *I. tabascanana*, *I. trifida*, *I. leucantha*, *I. tiliacea*, *I. tenuissima*, *I. triloba*, *I. splendor-sylvae*, and wild *I. batatas*. USA species *I. lacunosa* stood out for adaptation to low temperatures, with *I. grandifolia*, *I. cordatotriloba*, *I. tenuissima*, and *I. ramosissima* also demonstrating cold tolerance. These same species were among those displaying the widest adaptation to temperature fluctuation.

CWR of sweetpotato occurring in areas of high precipitation included *I. ramosissima*, *I. littoralis*, *I. splendor-sylvae*, *I. tabascanana*, *I. tiliacea*, *I. trifida*, and wild *I. batatas*. *Ipomoea littoralis*, *I. trifida*, *I. splendor-sylvae*, *I. leucantha*, *I. cynanchifolia*, *I. triloba*, *I. lacunosa*, and wild *I. batatas* were distributed in regions with low precipitation. These species were also among those displaying the widest adaptation to precipitation seasonality. Sweetpotato CWR also displayed variation in adaptation to edaphic characteristics. *Ipomoea tabascanana*, *I. grandifolia*, *I. tiliacea*, *I. splendor-sylvae*, *I. ramosissima*, *I. cordatotriloba*, and wild *I. batatas* occurred in clay soils, while *I. tenuissima*, *I. littoralis*, *I. cynanchifolia*, *I. grandifolia*, *I. leucantha*, and *I. cordatotriloba* were distributed in sandy soils.

Table 1. Utilization characteristics, number of germplasm accessions conserved *ex situ*, further collecting priorities, and potential adaptive traits associated with ecogeographic niches of sweetpotato crop wild relatives.

Taxon	Gene pool	Ploidy	Germplasm accessions	Gap analysis priority	Mean expert priority	Eco geographic cluster	Potential adaptation to
<i>I. batatas</i>	2	4x=60	4 (0)	HPS	HPS	2	Heat, high precipitation, drought, precipitation seasonality, clay soils
<i>I. cordatotriloba</i>	3	2x, 4x	103 (67)	HPS	MPS	1	Cold, temperature variation, clay soils, sandy soils
<i>I. cynanchifolia</i>	3	2x=30	1 (0)	HPS	HPS	1,2	Drought, precipitation seasonality, sandy soils
<i>I. grandifolia</i>	3	2x=30	124 (83)	MPS	MPS	1	Cold, temperature variation, clay soils, sandy soils
<i>I. lacunosa</i>	3	2x=30	10 (1)	HPS	MPS	1	Cold, temperature variation, drought
<i>I. leucantha</i>	3	2x=30	18 (15)	HPS	MPS	1,2	Heat, drought, precipitation seasonality, sandy soils
<i>I. littoralis</i>	2	2x=30	2 (2)	HPS	HPS	2	Heat, high precipitation, drought, precipitation seasonality, sandy soils
<i>I. ramosissima</i>	3	2x=30	34 (30)	HPS	MPS	2,1	Cold, high precipitation, clay soils
<i>I. splendor-sylvae</i>	3	2x=30	16 (9)	HPS	HPS	2	Heat, high precipitation, drought, precipitation seasonality, clay soils
<i>I. tabascanana</i>	2	4x=60	4 (2)	LPS	HPS	2	Heat, high precipitation, clay soils
<i>I. tenuissima</i>	3	2x=30	3 (1)	HPS	HPS	1	Heat, cold, temperature variation, sandy soils
<i>I. tiliacea</i>	3	4x=60	61 (44)	HPS	HPS	2	Heat, high precipitation, clay soils
<i>I. trifida</i>	2	2x,3x,4x,6x	248 (159)	MPS	MPS	2	Heat, high precipitation, drought, precipitation seasonality
<i>I. triloba</i>	3	2x=30	121 (74)	HPS	MPS	2,1	Heat, drought

Ploidy data adapted from Nimmakayala *et al.* (2011). Germplasm accessions displays both the total number of accessions recorded in genebanks, as well as the number of accessions with unique geographic coordinates (i.e., unique populations) in parenthesis. HPS = high, MPS = medium, and LPS = low priority species for further collecting.

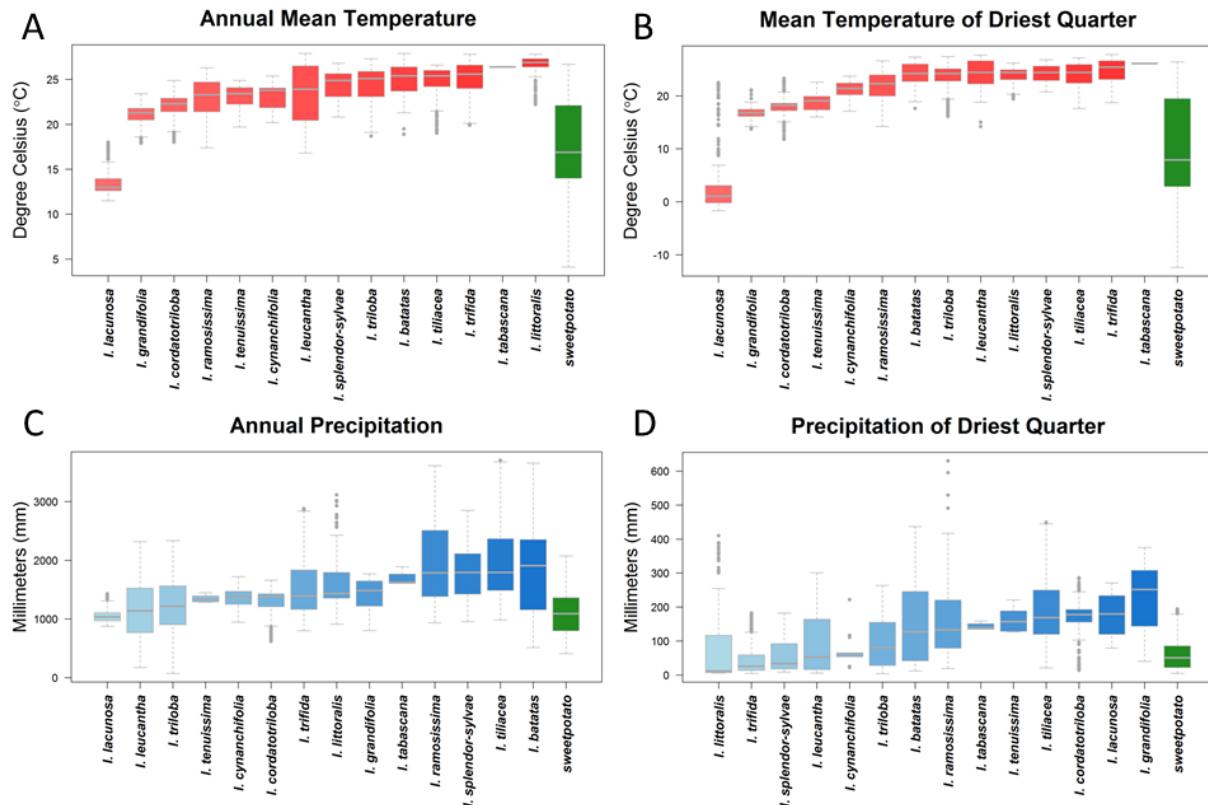


Figure 4. Ecogeographic niches of crop wild relative species based upon their occurrence data presence locations, and the sweetpotato crop, for (A) annual mean temperature, (B) mean temperature of the driest quarter of the year, (C) annual precipitation, and (D) precipitation of the driest quarter of the year. The thick grey line represents median values, boxplots between 25-75% variation, and open circles outliers within 90% of total variation. For a principal component analysis of variables see Supplementary Figure 6 and for ecogeographic niches displaying total variation for all variables per species see Supplementary Figure 7.

DISCUSSION

This article utilizes the most current knowledge on species concepts within *I. series Batatas*. Due to taxonomic uncertainties and to the notable dearth of study material for sweetpotato CWR, the results represent a preliminary understanding of the geography and conservation status of the series, to be further refined following increased levels of collecting sufficient to support the needed taxonomic and crossability research. Further collecting of germplasm thus serves two purposes: a) conserves genetic resources for the long-term and makes these resources available to breeders; and b) provides the basic materials needed by researchers to understand the diversity present in the CWR of sweetpotato.

A total of 78.6% of the wild relatives of sweetpotato considered in this study were assessed as high priority for further collecting for *ex situ* conservation. With general agreement from expert evaluators of medium to high importance for all species, it is clear that much remains to be done

to safeguard the wild genetic resources of this critically important root crop. Included in this list of priorities are species with very few germplasm accessions accessible to the global community in genebank information systems, including *I. cynanchifolia*, *I. littoralis*, *I. tenuissima*, *I. tabascana*, *I. lacunosa*, *I. leucantha*, *I. splendor-sylvae*, and clearly designated wild forms of the crop conspecific *I. batatas*. Such species represent the highest level of priority for further collecting for use in systematic analyses as well as genetic resources conservation. As the species diversity gaps in genebank collections largely align with the geographic distribution of species richness of sweetpotato CWR, hotspots in central Mexico to Central America, and in the extreme southeastern USA, represent particularly high priority regions for efficient collecting of the sweetpotato genepool (Figure 3). Additional unrepresented populations of high priority species such as *I. littoralis* and *I. cynanchifolia* occur outside those regions, thus targeted collecting throughout the geographic distribution of the genepool is necessary in order to form comprehensive germplasm collections.

Due to having relatively large potential distributions occupying a diversity of ecosystems, species such as *I. triloba*, *I. cordatotriloba*, and *I. tiliacea* were categorized as high priority, and *I. trifida* and *I. grandifolia* as medium priority for further collecting despite having sizable currently-existing germplasm collections. As the cost of conserving and investigating germplasm *ex situ* is significant, a further assessment of what constitutes sufficient representation of these species in germplasm collections is warranted. Given adequate resources, further collecting may be worthwhile, as extremely valuable traits sourced from CWR of some crops have been found in only a few populations despite screening of a large number of accessions (Brar & Khush 1997), and accessions of individual CWR species such as *I. triloba* have been shown to possess notable variation in traits such as tolerance to precipitation (Martin & Jones 1973, Nimmakayala *et al.* 2011).

As Maxent models are based upon known presence points for species and are thus subject to sampling bias, they may not fully capture the possibility of occurrence of populations of CWR species in unique climates (Araújo & Guisan 2006, Loiselle *et al.* 2008, Kramer-Schadt *et al.* 2013). Further field exploration of climatic extremes beyond the edges of the distributions created through these methods may therefore lead to the discovery of new populations with particularly valuable adaptations to abiotic stress (Williams *et al.* 2009). Investigation of non-native populations (e.g., *I. trifida* in Asia) may also yield novel adaptations of use in crop

improvement. As techniques for the utilization of distantly related germplasm improve, the exploration of other *Ipomoea* species outside of *I. series Batatas* may also result in the identification of beneficial traits [e.g., *Ipomoea purpurea* (L.) Roth, for stem nematode and SPVD resistance (Cao *et al.* 2009)].

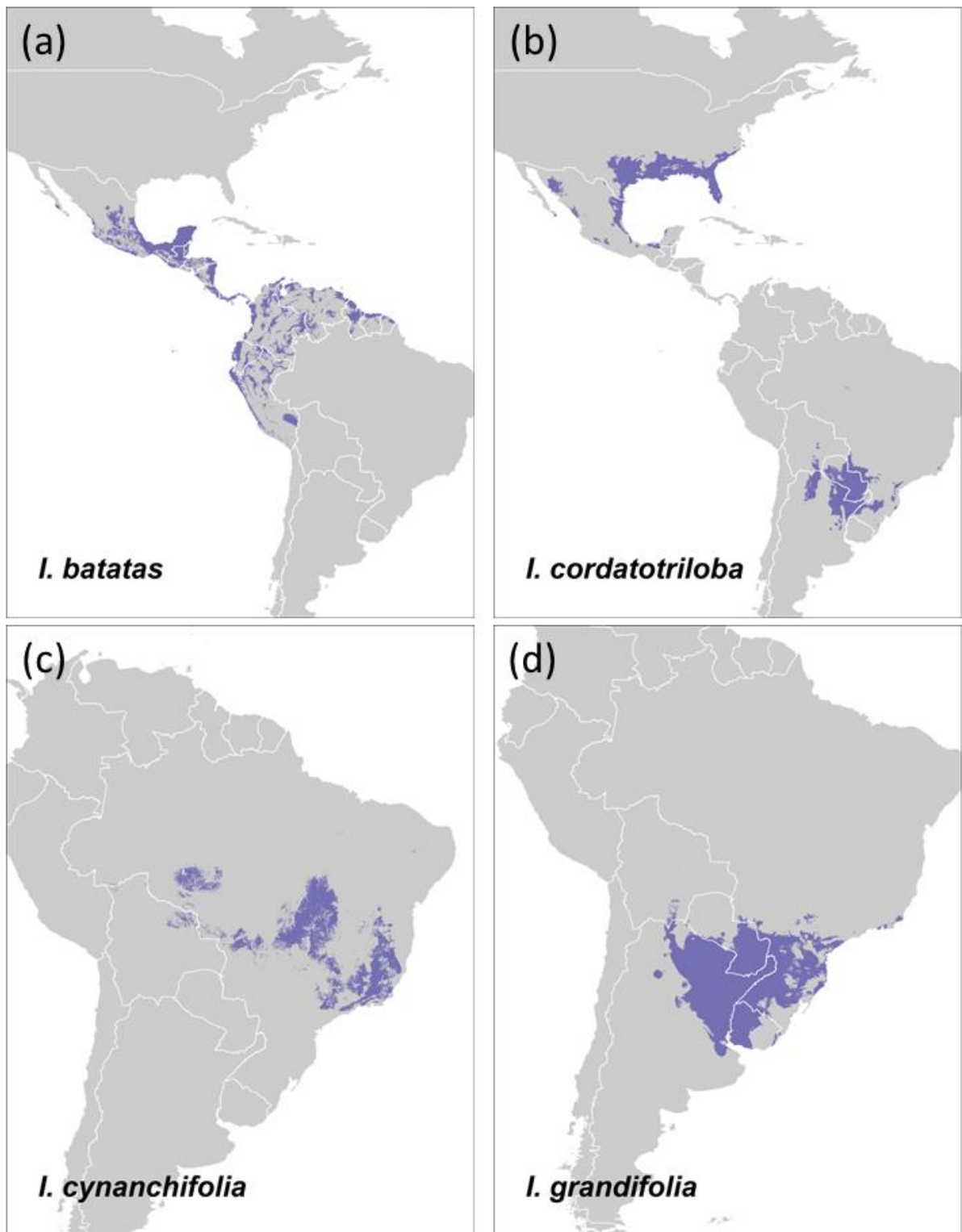
Analysis of geographic overlap and ecogeographic similarities between species, as well as ecogeographic clusters among all species, can supplement morphological and genetic analyses in differentiating useful genetic resources, and can serve as a point of departure for identifying taxonomically problematic populations for further investigation. These analyses may also indicate geographic areas of particular interest in regard to high rates of hybridization, as in the case of *I. cordatotriloba* and *I. lacunosa* (Duncan & Rausher 2013), which indeed were identified as sharing similar geographies and ecogeographic niches. The purported hybrid descendent of these species, *I. leucantha*, was modeled as containing a much more extended latitudinal gradient in the northern hemisphere than its parents, as well as a differing ecogeographic niche, including potential adaptation to high heat and to drought.

Genetic resistance is essential to efforts to overcome major biotic and abiotic constraints in sweetpotato production. As these constraints are often interrelated, e.g., drought stress with SPW and SPVD damage, enhancement of broad resistance for traits such as drought may improve yield across relatively large geographic areas, without the need to breed for resistance to localized viral strains (Ngailo *et al.* 2013). Such broad tolerance may also improve adoption rates for sweetpotato varieties with other desirable characteristics, such as high β -carotene content.

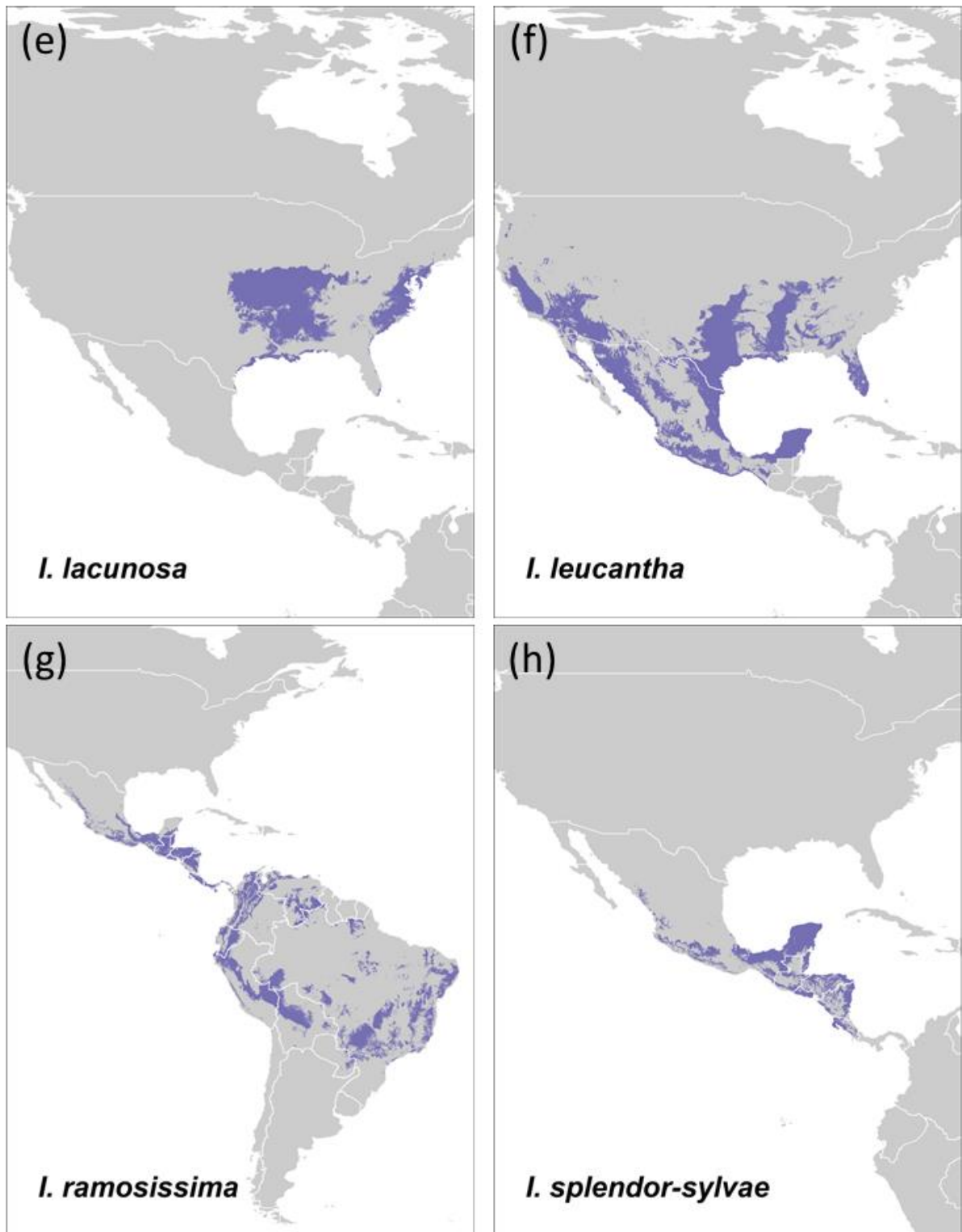
Reliable funding for germplasm collections is paramount in order to safeguard sweetpotato CWR genetic resources in the long-term and to continue to make *ex situ* collections available to the global community. Further investment in genebank information systems, *ex situ* conservation technologies (i.e., storage, testing, and regeneration), safety duplication of unique germplasm, and ensuring facilitated access to genetic diversity is equally essential (FAO 2002, FAO 2010, Khoury *et al.* 2010). In order to maximize the usefulness of conserved germplasm, characterization and evaluation for traits of interest, and increased breeding research, which have been limited for CWR of sweetpotato, are also needed. Further research combining morphological studies, trait evaluations, and genetic diversity analyses is likewise critically

needed for elucidating species boundaries and highlighting accessions of particular value for use in breeding. Recent focused research has produced quick gains, including the identification of new species (Duncan & Rausher 2013, Wood *et al.* 2015). Through these actions the crop research community will contribute to ensuring the long term viability of this important root crop.

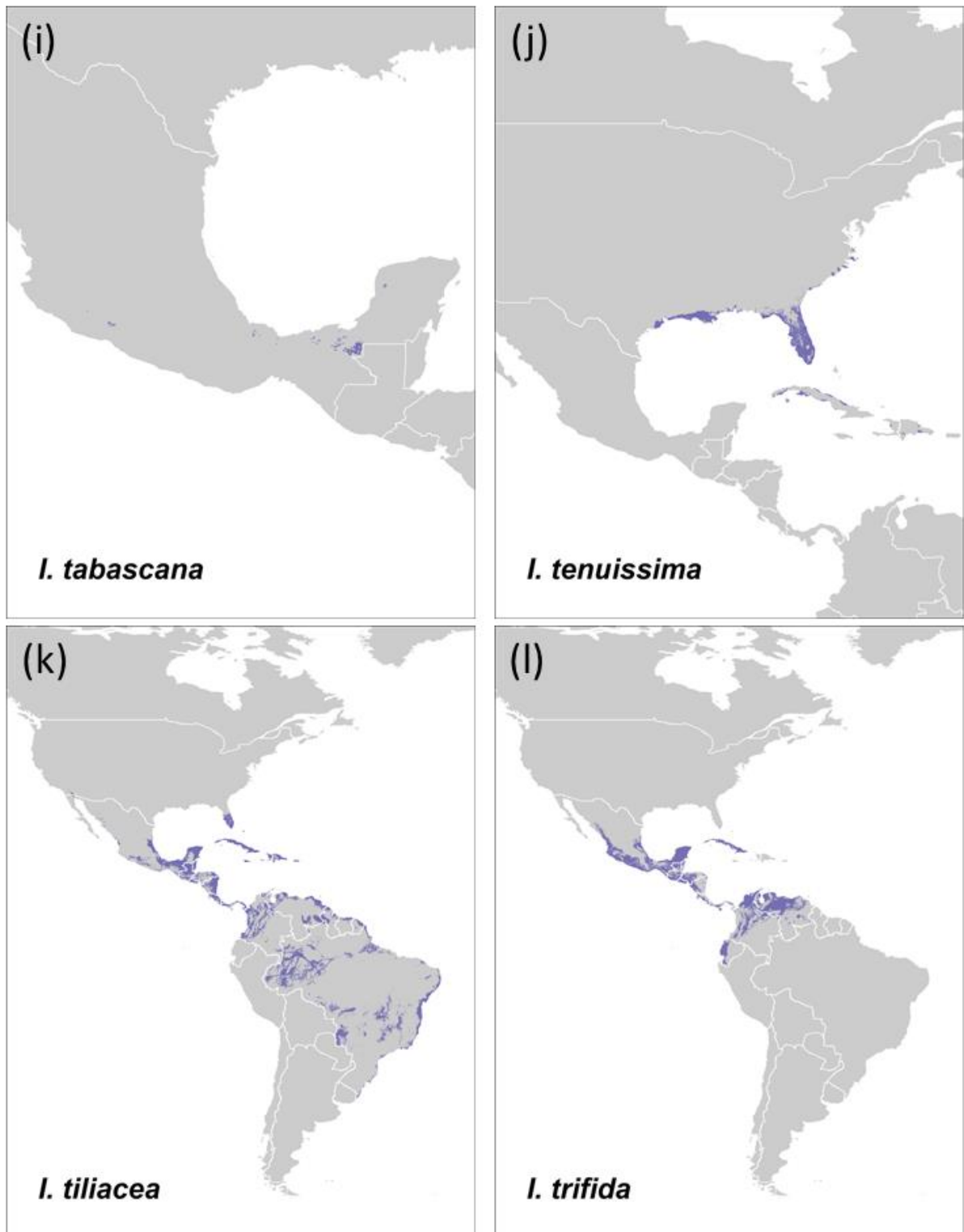
SUPPLEMENTARY INFORMATION



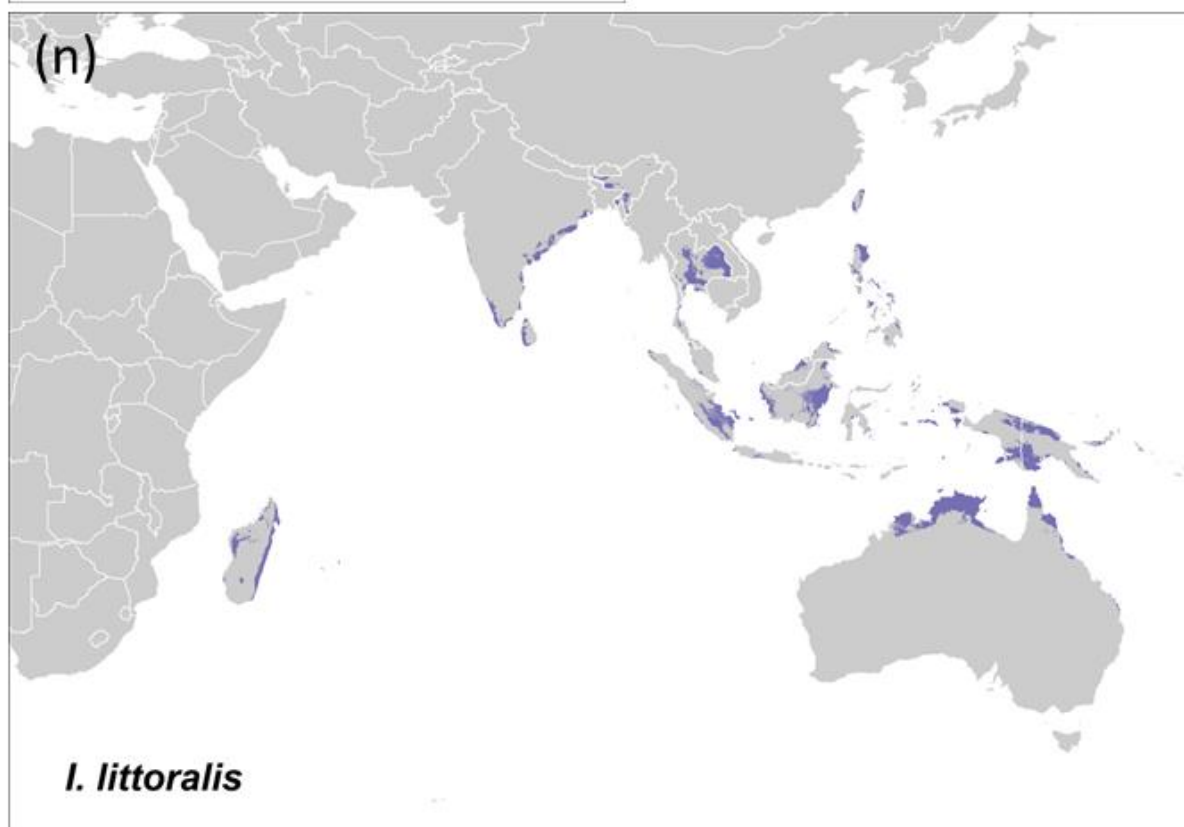
Supplementary Figure 1(A-D). Sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative potential distribution models.



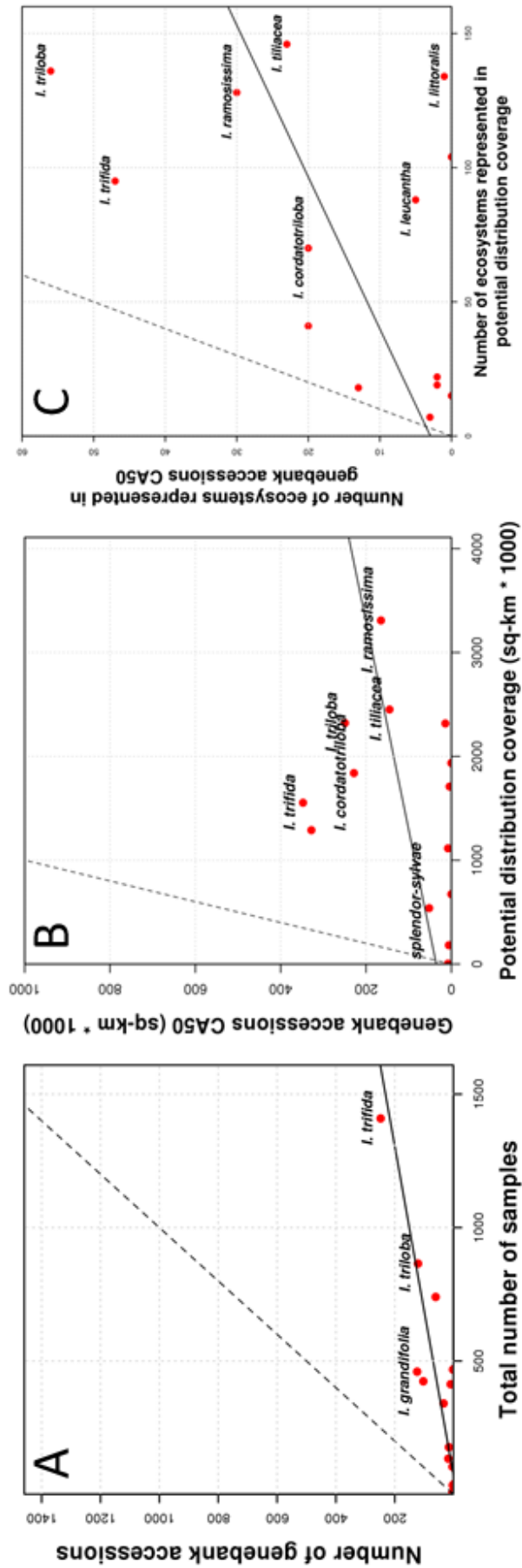
Supplementary Figure 1(E-H). Sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative potential distribution models.



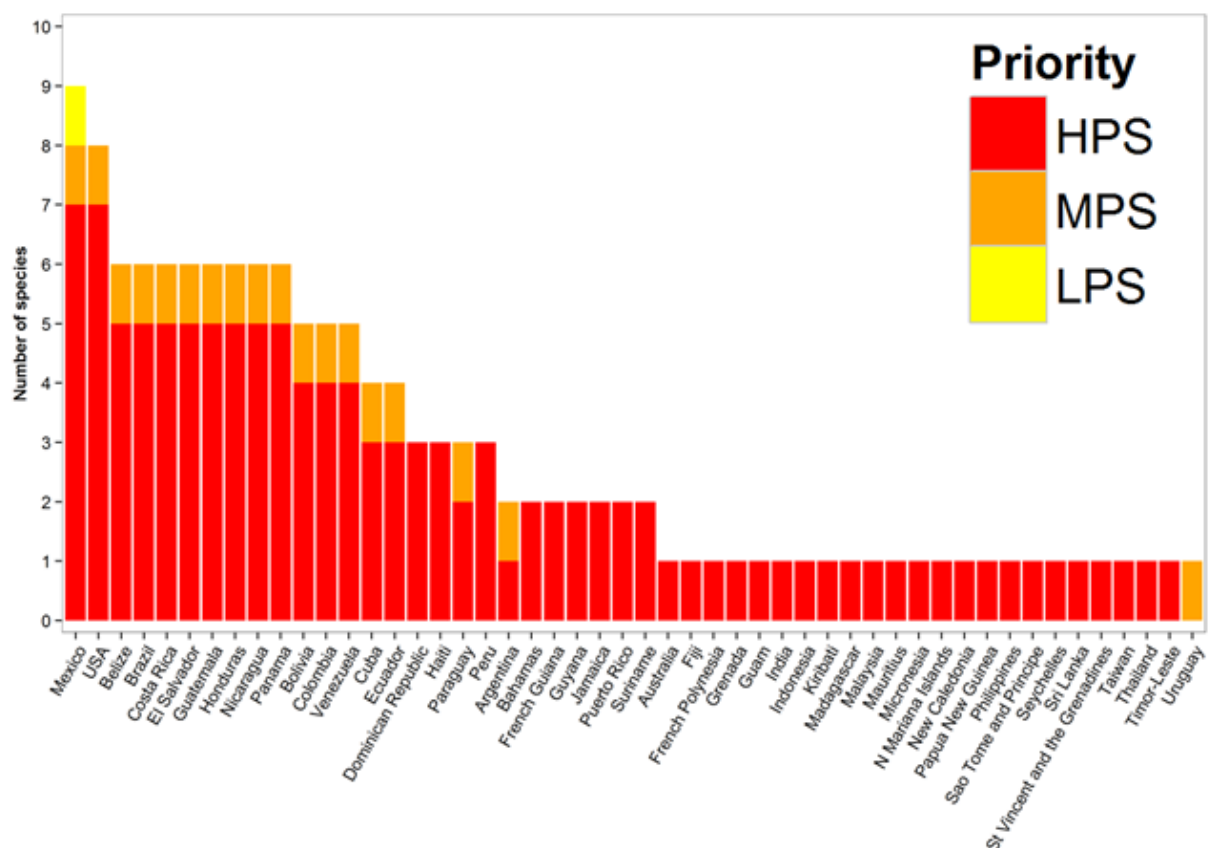
Supplementary Figure 1(I-L). Sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative potential distribution models.



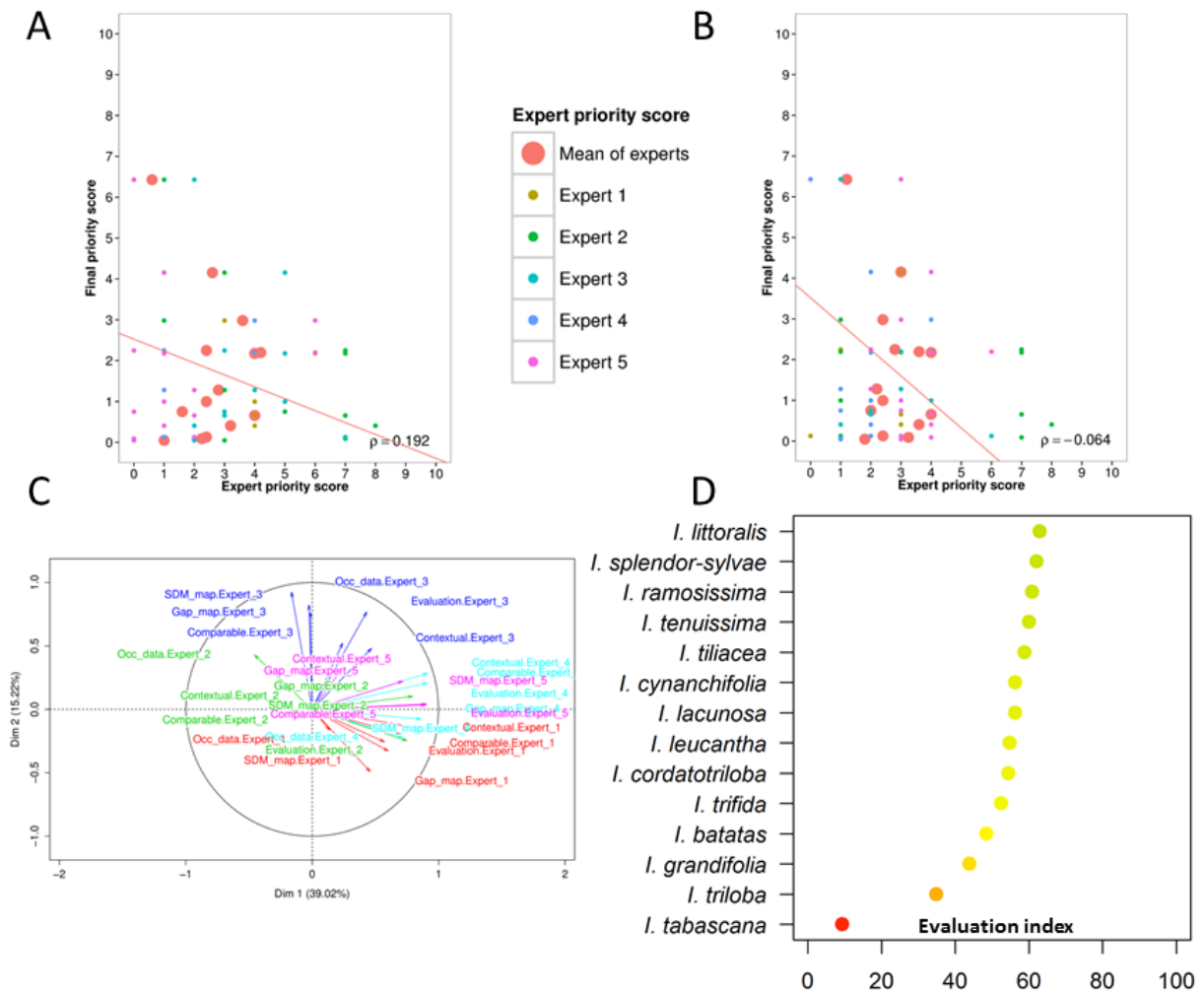
Supplementary Figure 1(M-N). Sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative potential distribution models.



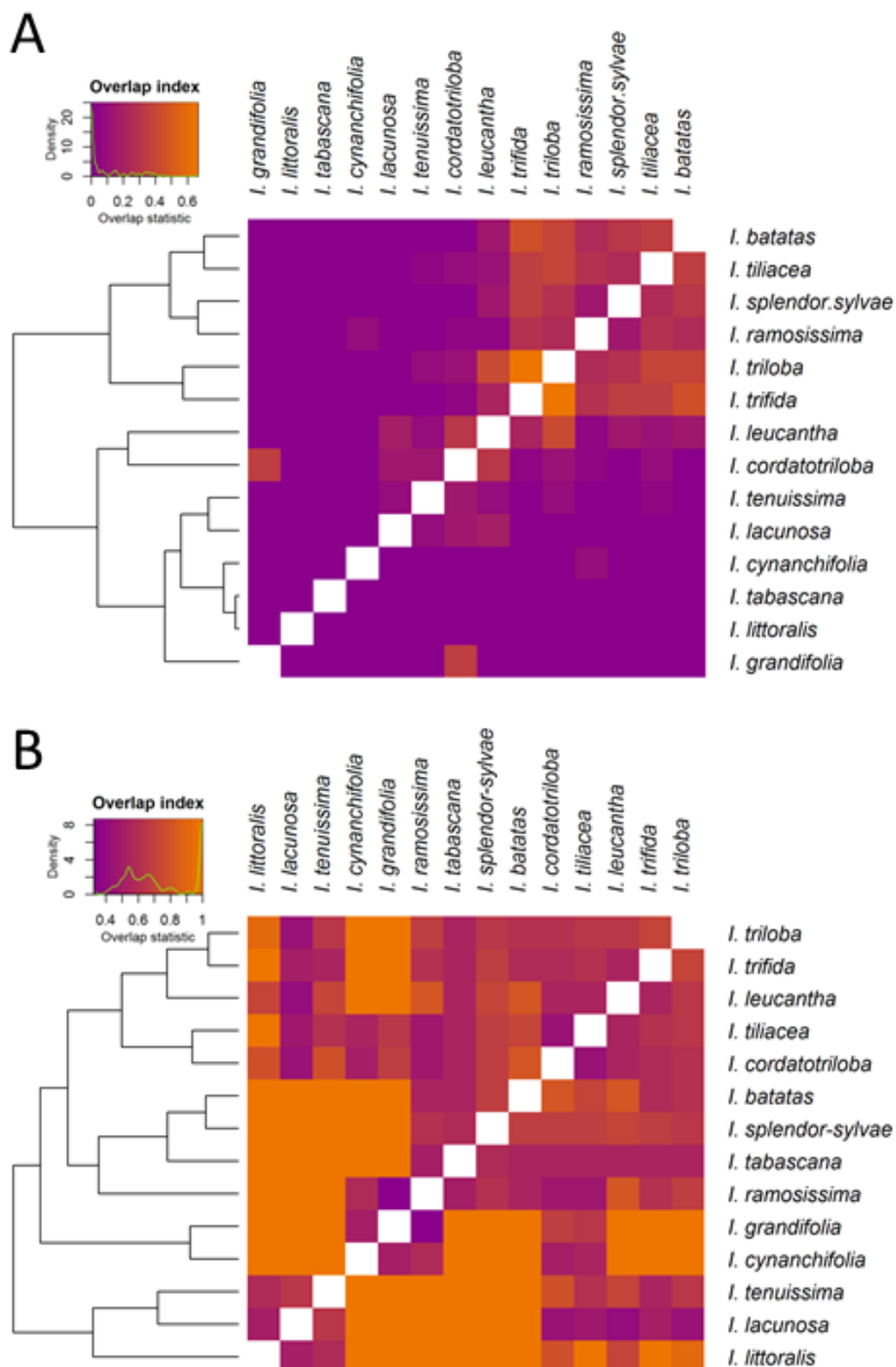
Supplementary Figure 2. Distribution of priority scores per sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative species for (A) sampling representativeness score (SRS), (B) geographic representativeness score (GRS), and (C) ecological representativeness score (ERS).



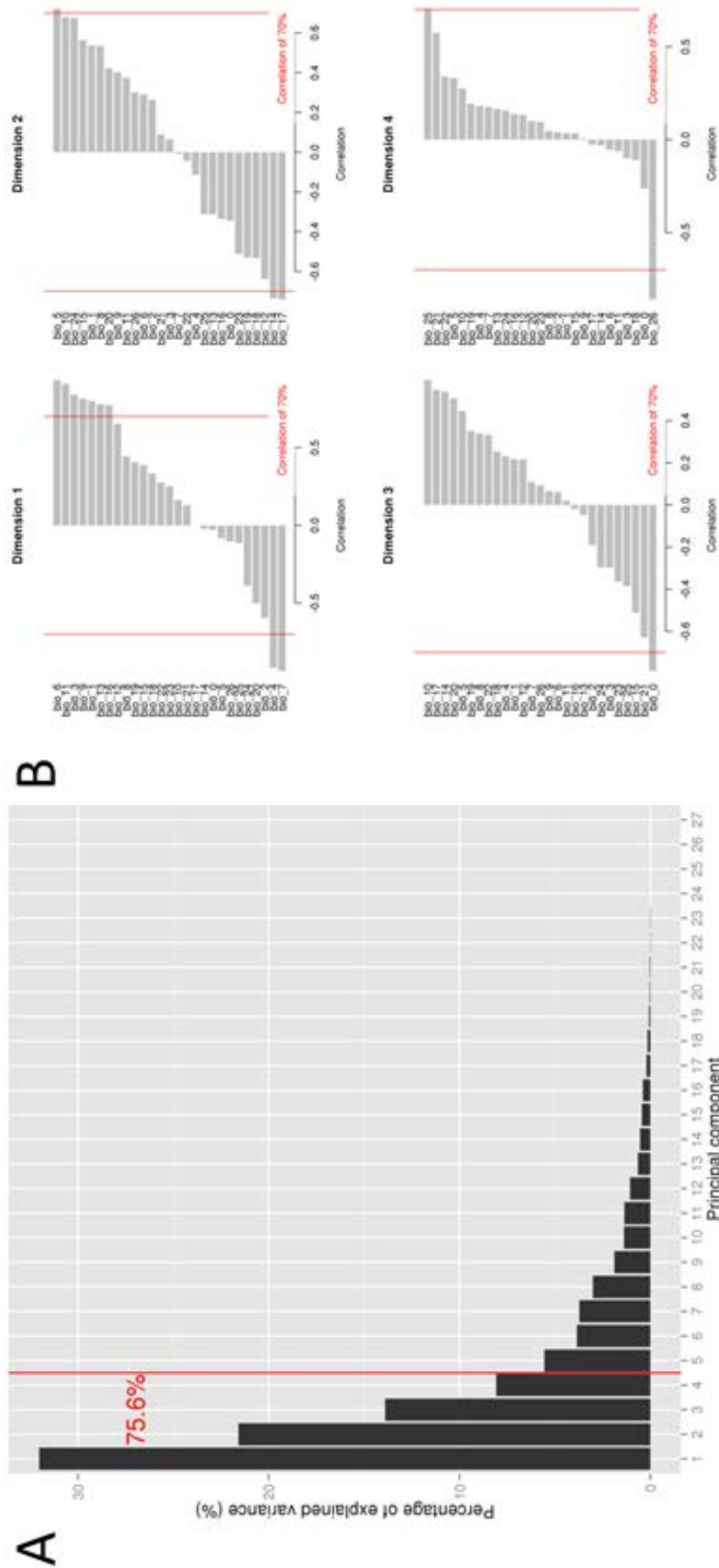
Supplementary Figure 3. Number of sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relatives prioritized for further collecting for *ex situ* conservation per country. HPS = high, MPS = medium, and LPS = low priority species for further collecting.



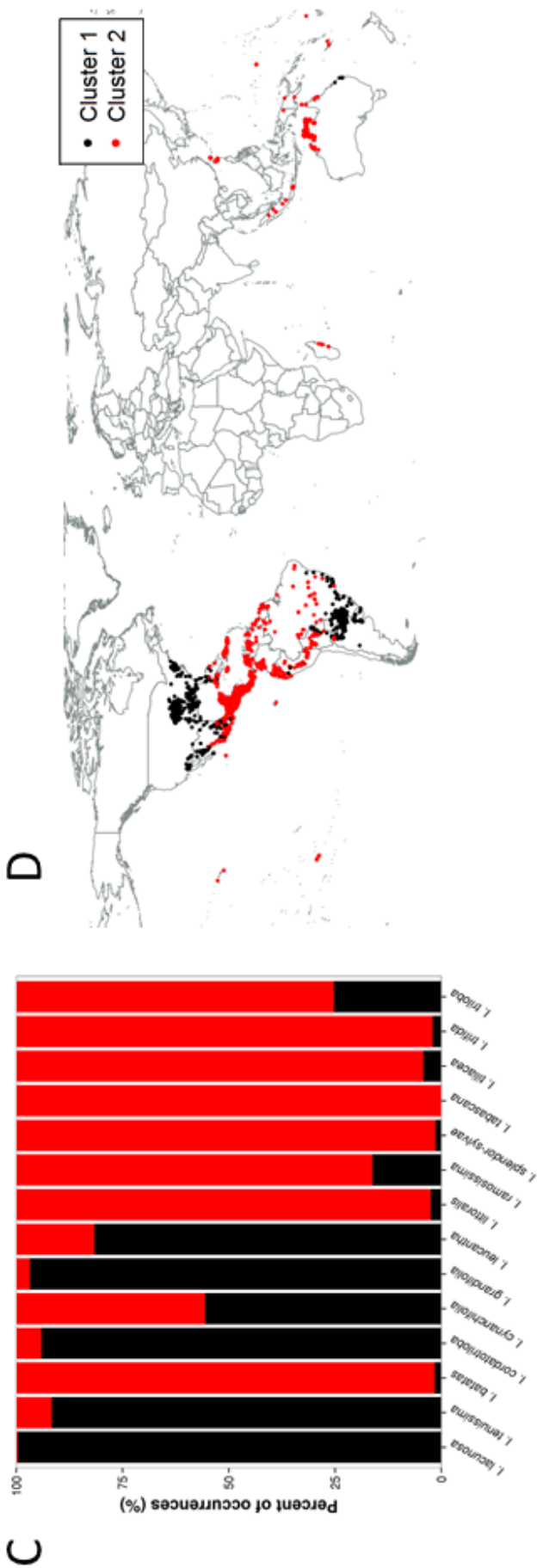
Supplementary Figure 4. Expert evaluation accordance with gap analysis results: (A) correlation between gap analysis results and comparable expert evaluation scores. (B) correlation between gap analysis results and contextual expert evaluation scores. (C) correlation circle of all evaluation variables [comparable expert priority score (Comparable), contextual expert priority score (Contextual), evaluation of gap analysis results score (Evaluation), evaluation of occurrence data (Occ_data), evaluation of potential species distribution models (SDM_map), and evaluation of collecting priorities map (Gap_map)]. (D) combined expert evaluation index score per sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative.



Supplementary Figure 5. Geographic overlap and ecogeographic similarity of potential distribution models between sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative species: (A) geographic overlap of potential distribution models. (B) ecogeographic similarity index (I) of potential distribution models.

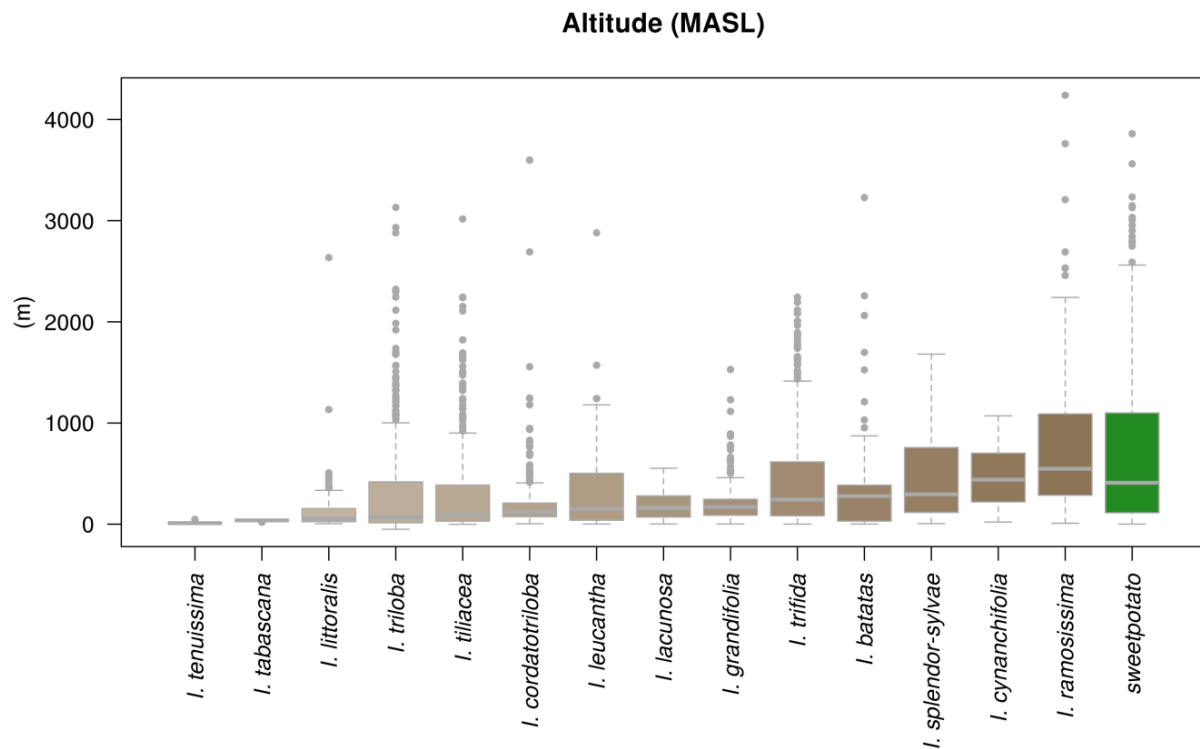


Supplementary Figure 6. Principal component analysis of bioclimatic variables associated with occurrence data for sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relatives: (A) variation represented in the first four principal components. (B) determining bioclimatic variables for the first four principal components.

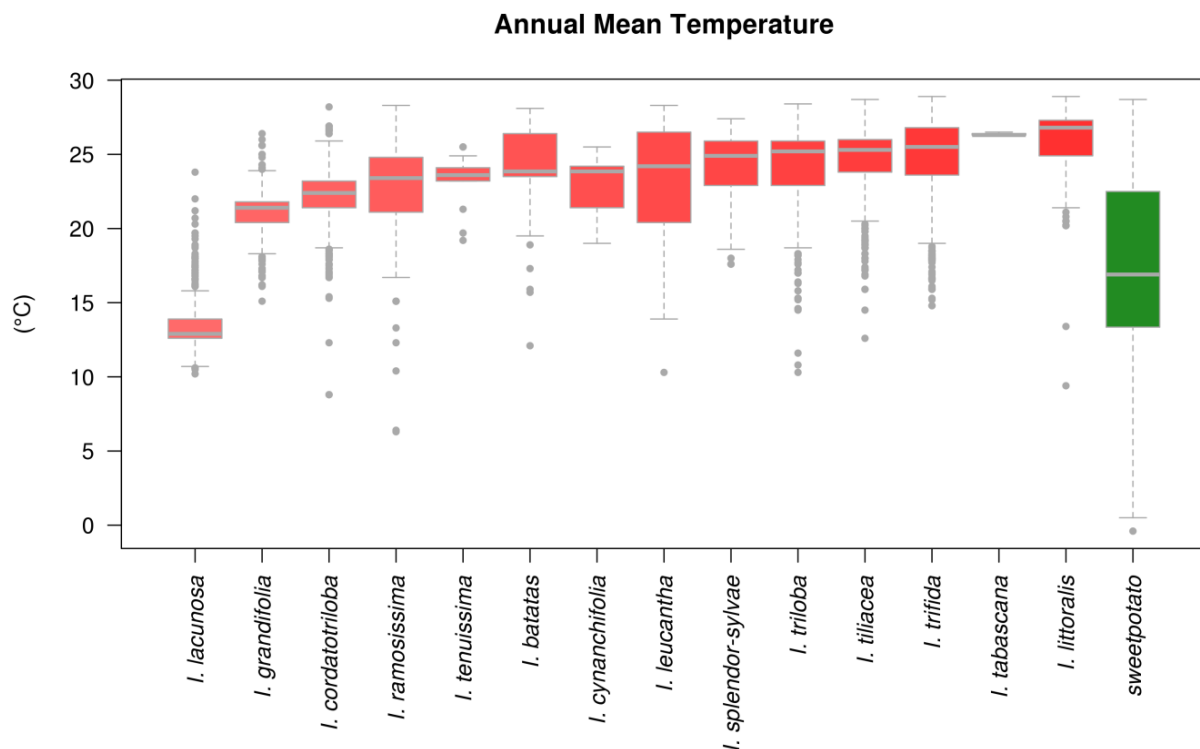


Supplementary Figure 6. Principal component analysis of bioclimatic variables associated with occurrence data for sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relatives: (C) contribution of cluster identified occurrence data points per species. (D) geographic origin of clustered occurrence data points.

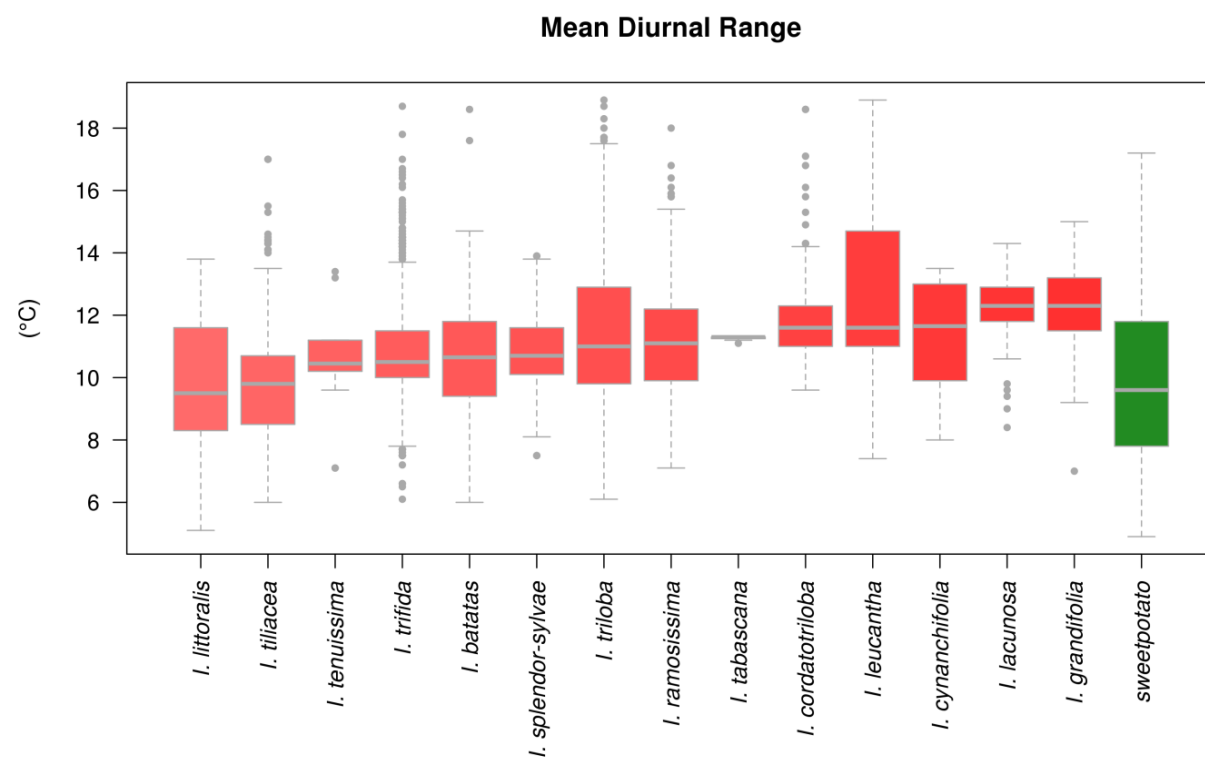
Supplementary Figure 7. Ecogeographic niches of crop wild relative (CWR) species based upon their occurrence data presence locations, and the sweetpotato [*Ipomoea batatas* (L.) Lam.] crop, per bioclimatic and edaphic variable. The bold grey line represents median values, boxplots between 25-75% variation, and circles outliers.



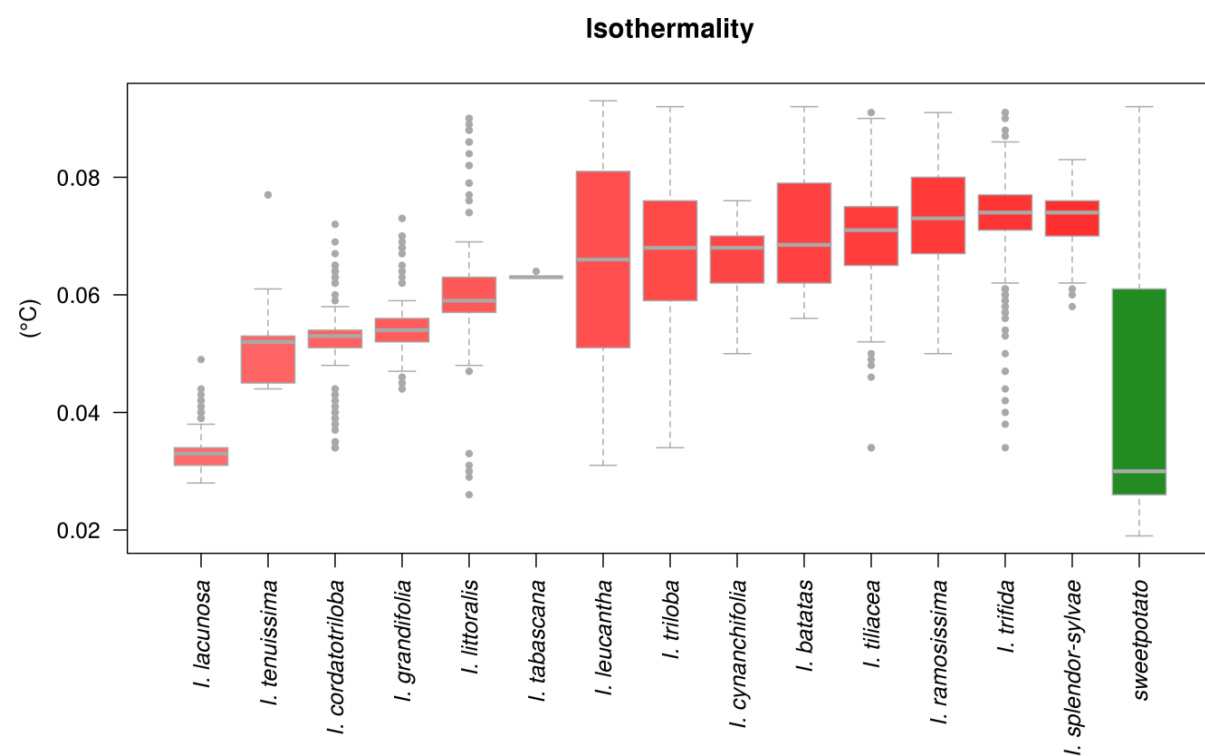
Supplementary Figure 7A. Niches of CWR species and the sweetpotato crop for altitude.



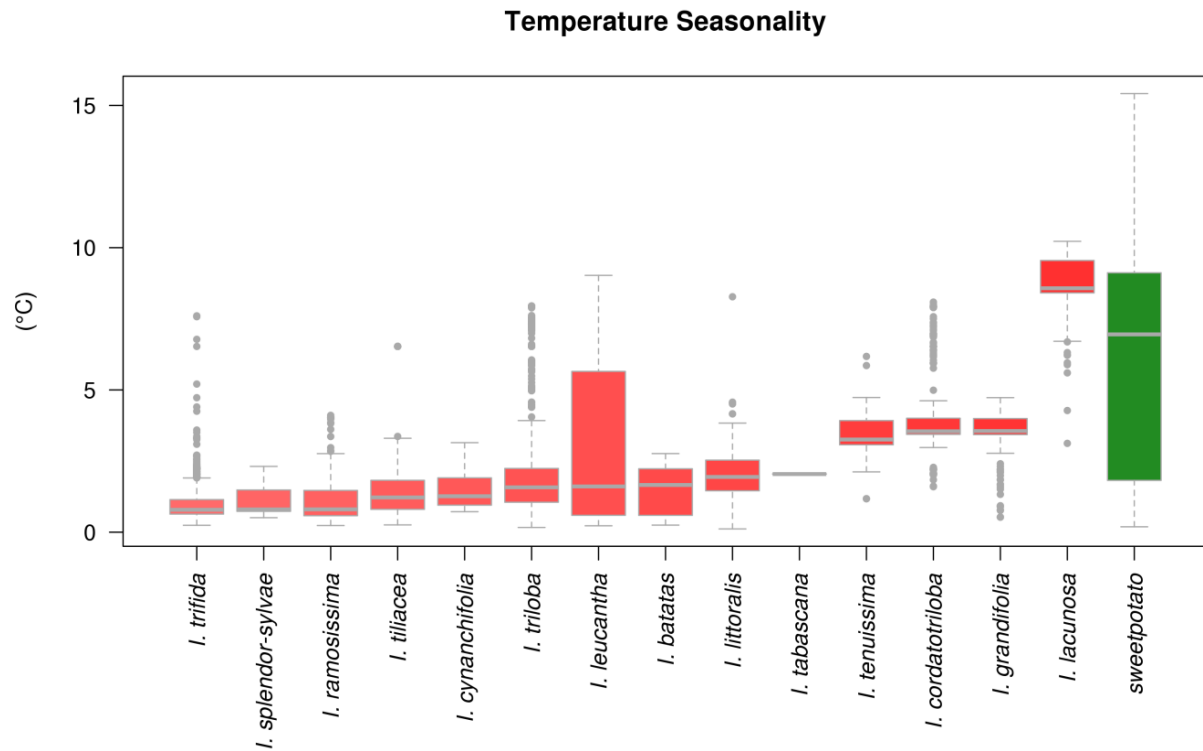
Supplementary Figure 7B. Climatic niches of CWR species and the sweetpotato crop for annual mean temperature.



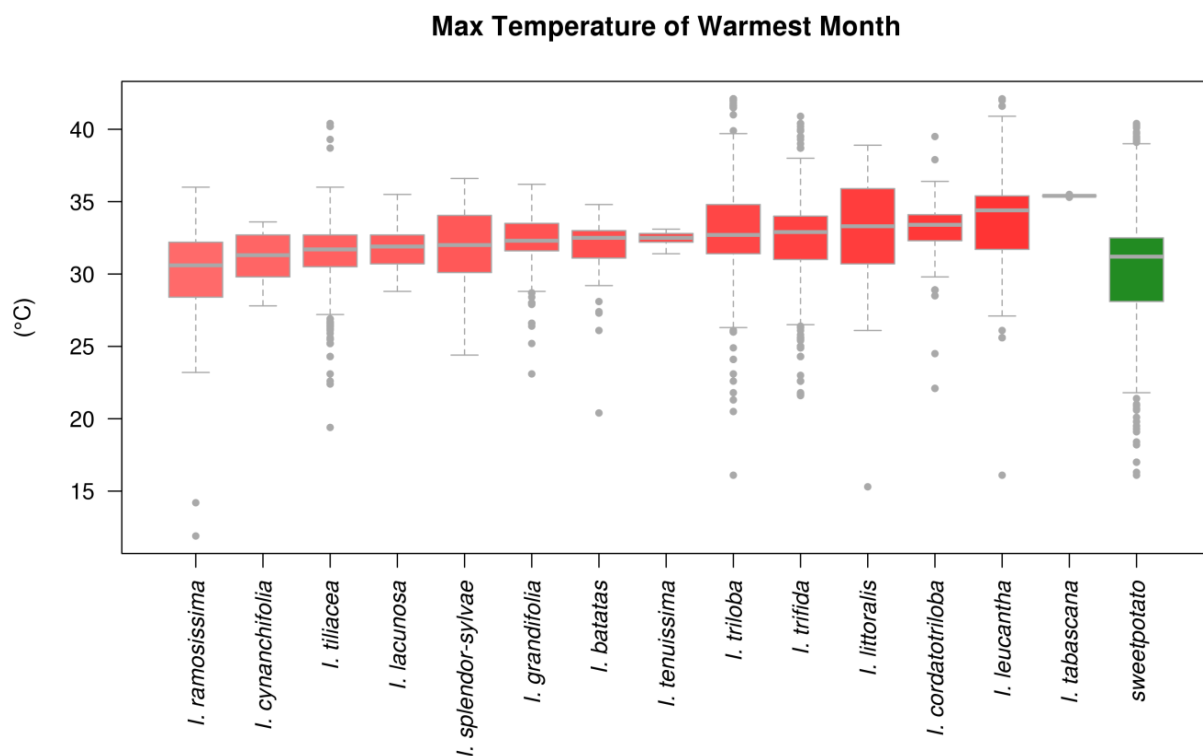
Supplementary Figure 7C. Climatic niches of CWR species and the sweetpotato crop for mean diurnal range.



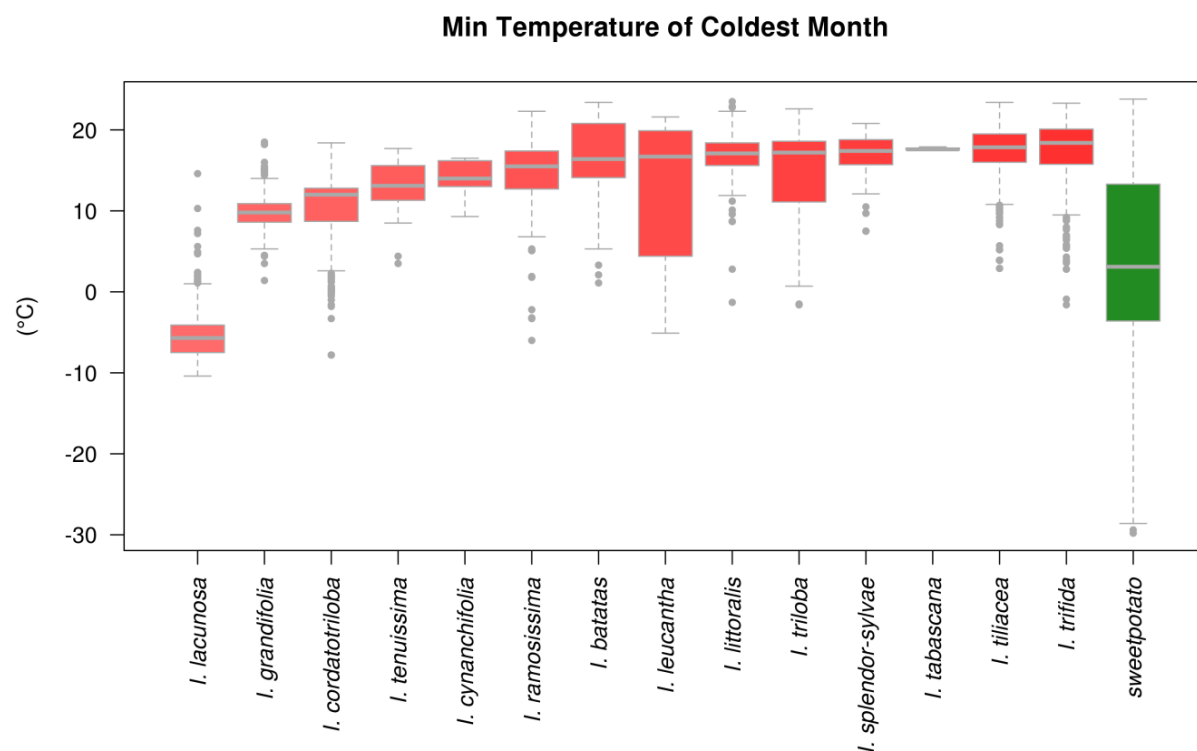
Supplementary Figure 7D. Climatic niches of CWR species and the sweetpotato crop for isothermality.



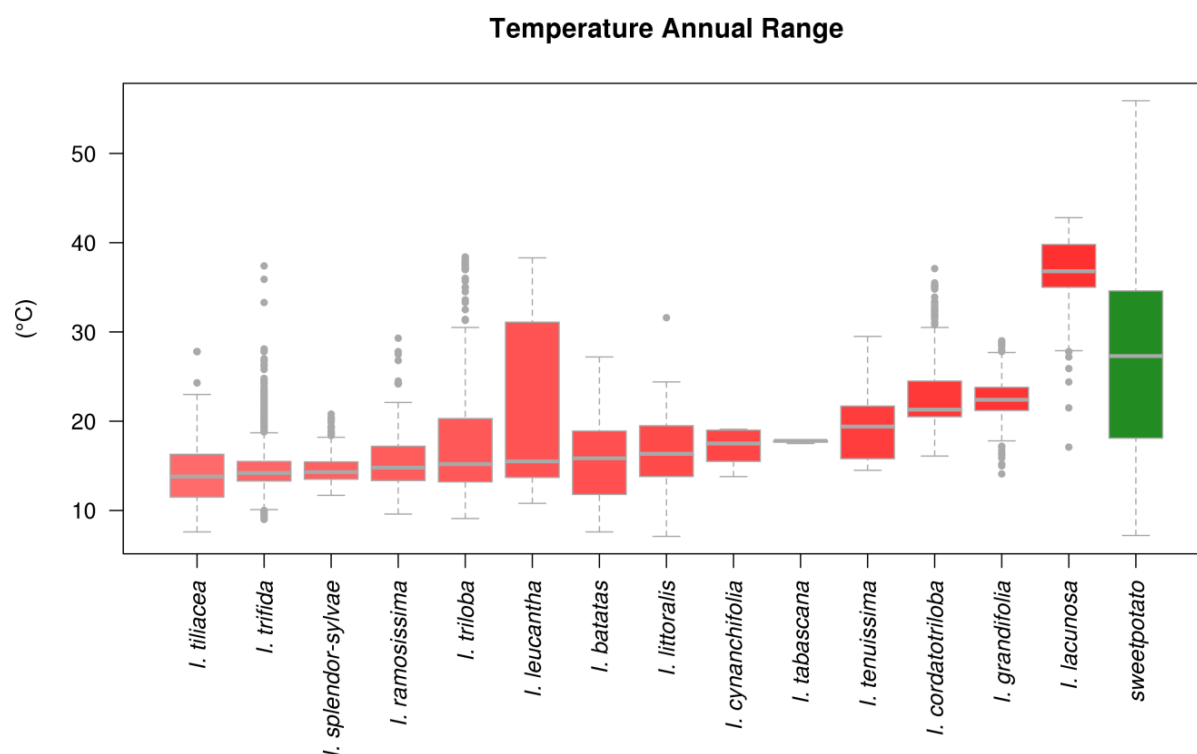
Supplementary Figure 7E. Climatic niches of CWR species and the sweetpotato crop for temperature seasonality.



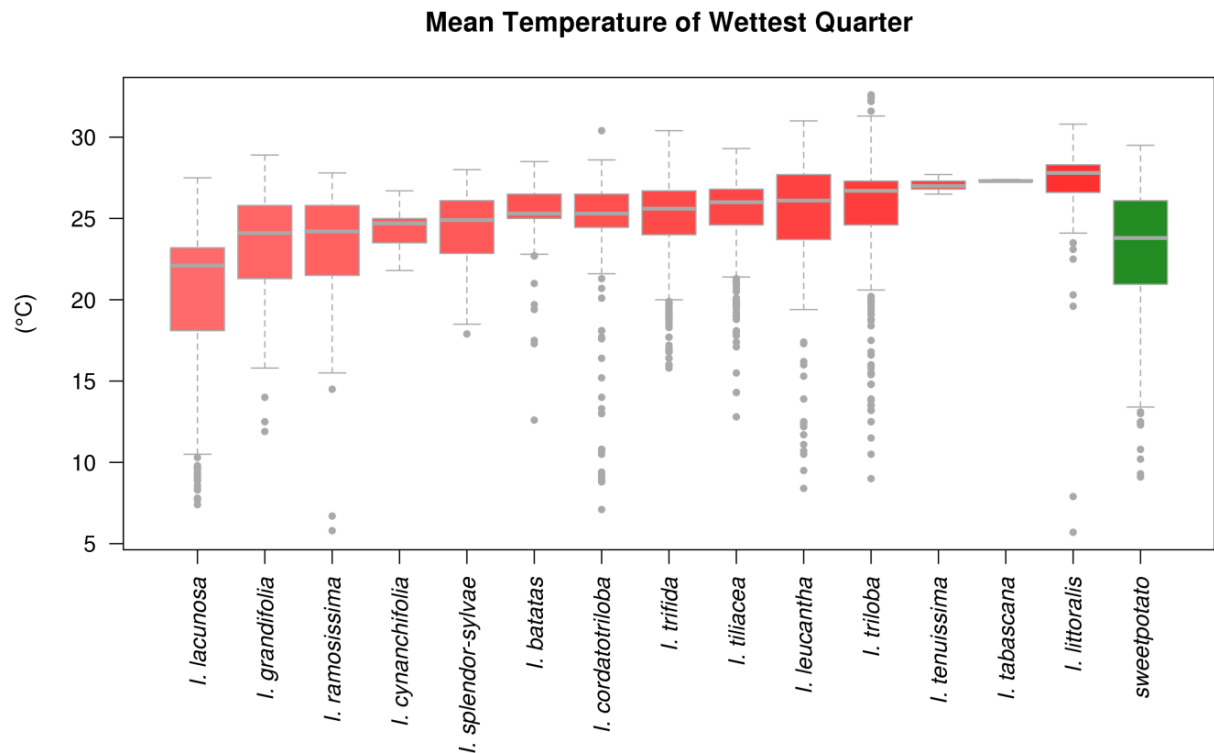
Supplementary Figure 7F. Climatic niches of CWR species and the sweetpotato crop for maximum temperature of the warmest month of the year.



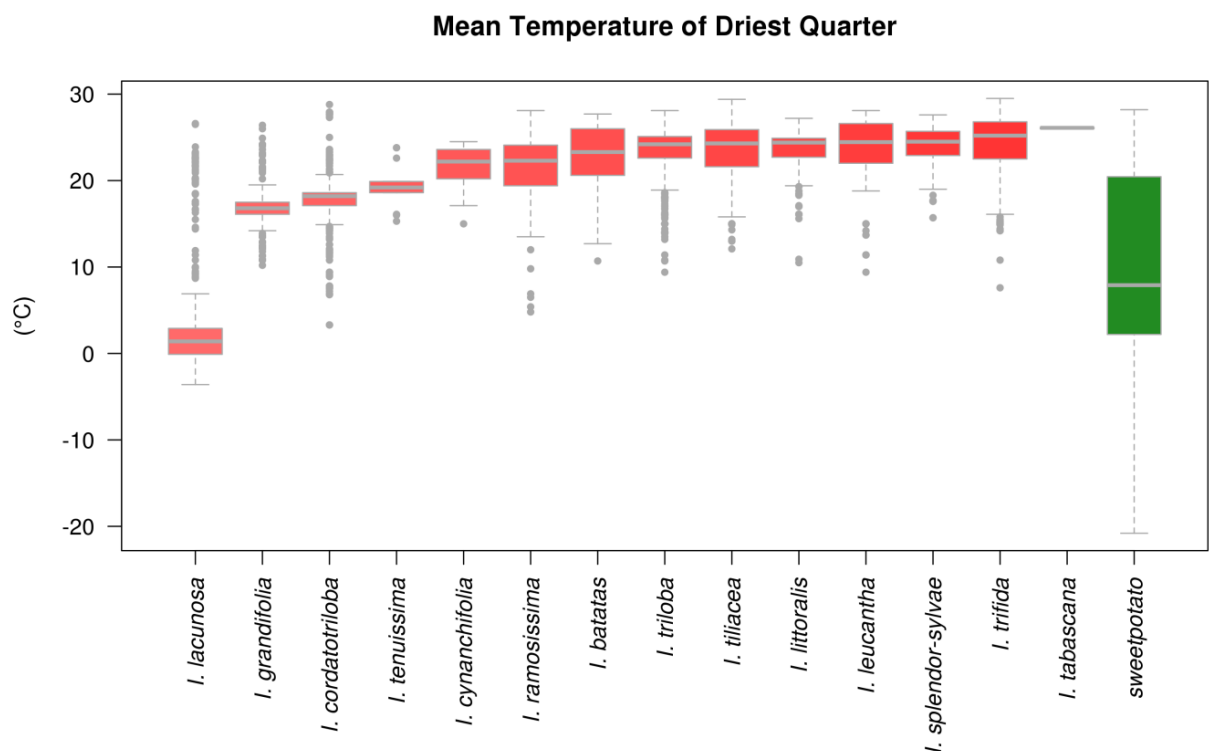
Supplementary Figure 7G. Climatic niches of CWR species and the sweetpotato crop for minimum temperature of the coldest month of the year.



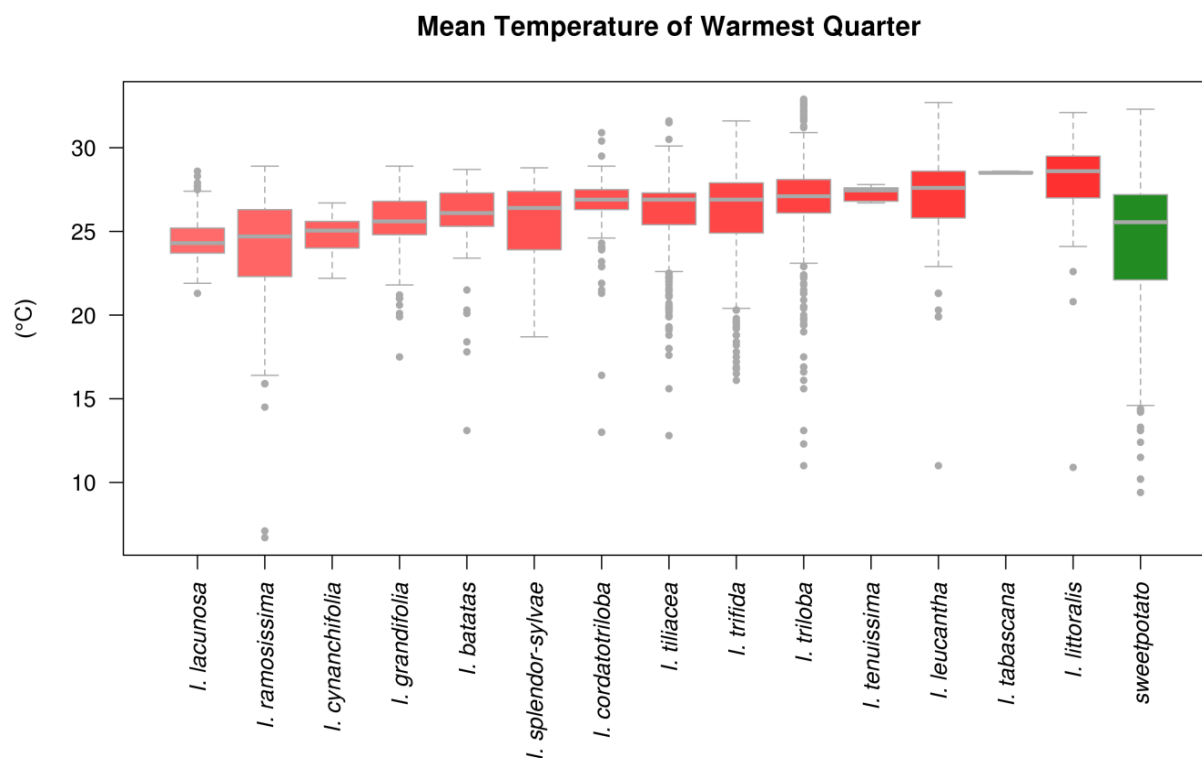
Supplementary Figure 7H. Climatic niches of CWR species and the sweetpotato crop for annual temperature range.



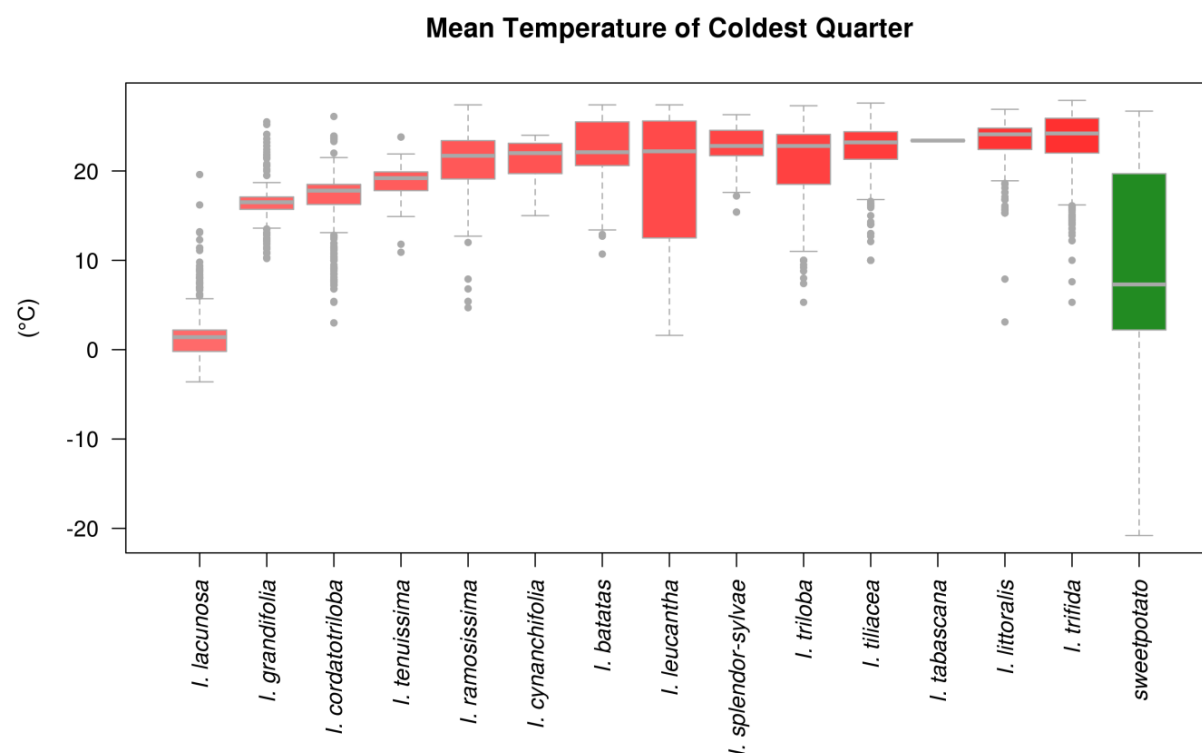
Supplementary Figure 7I. Climatic niches of CWR species and the sweetpotato crop for mean temperature of the wettest quarter of the year.



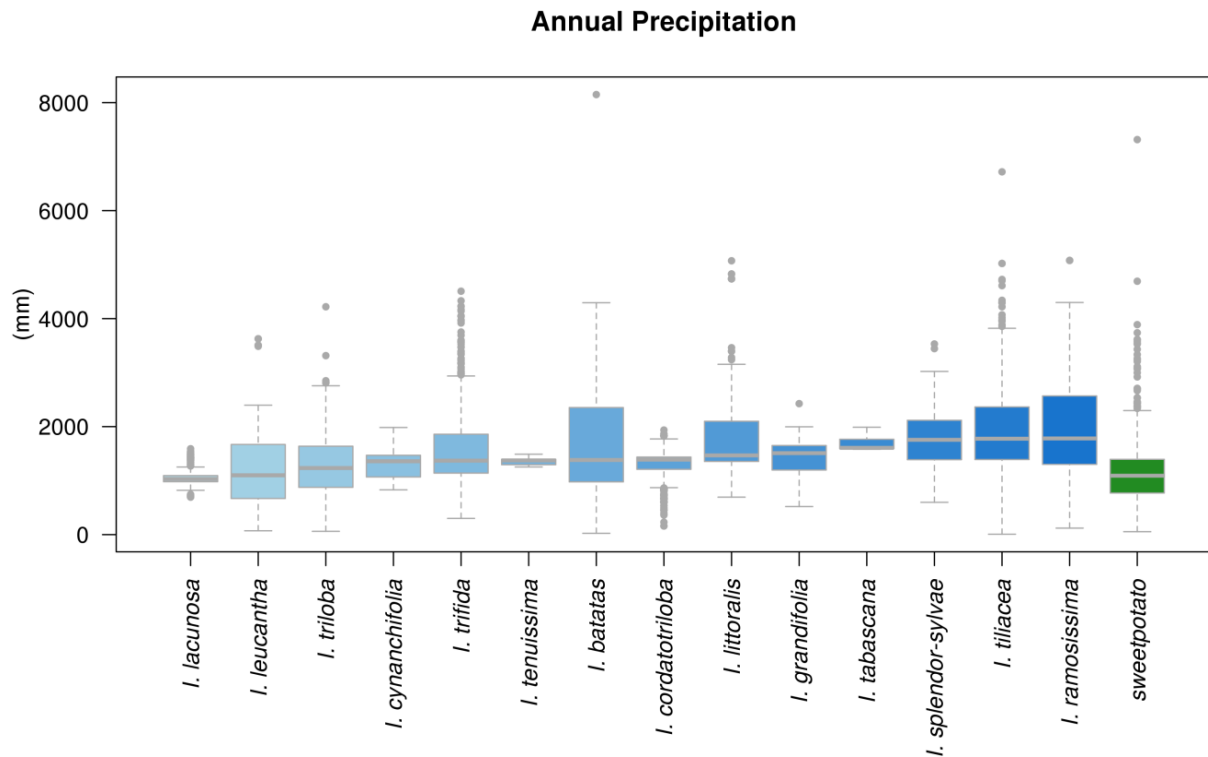
Supplementary Figure 7J. Climatic niches of CWR species and the sweetpotato crop for mean temperature of the driest quarter of the year.



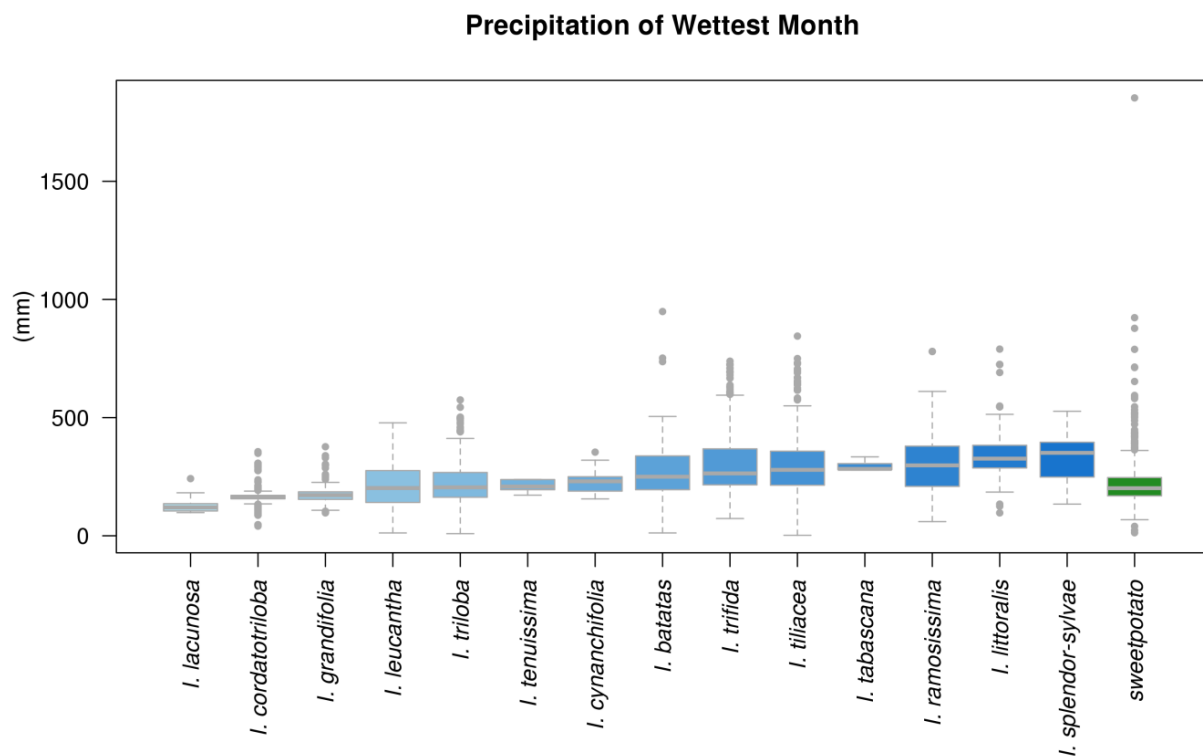
Supplementary Figure 7K. Climatic niches of CWR species and the sweetpotato crop for mean temperature of the warmest quarter of the year.



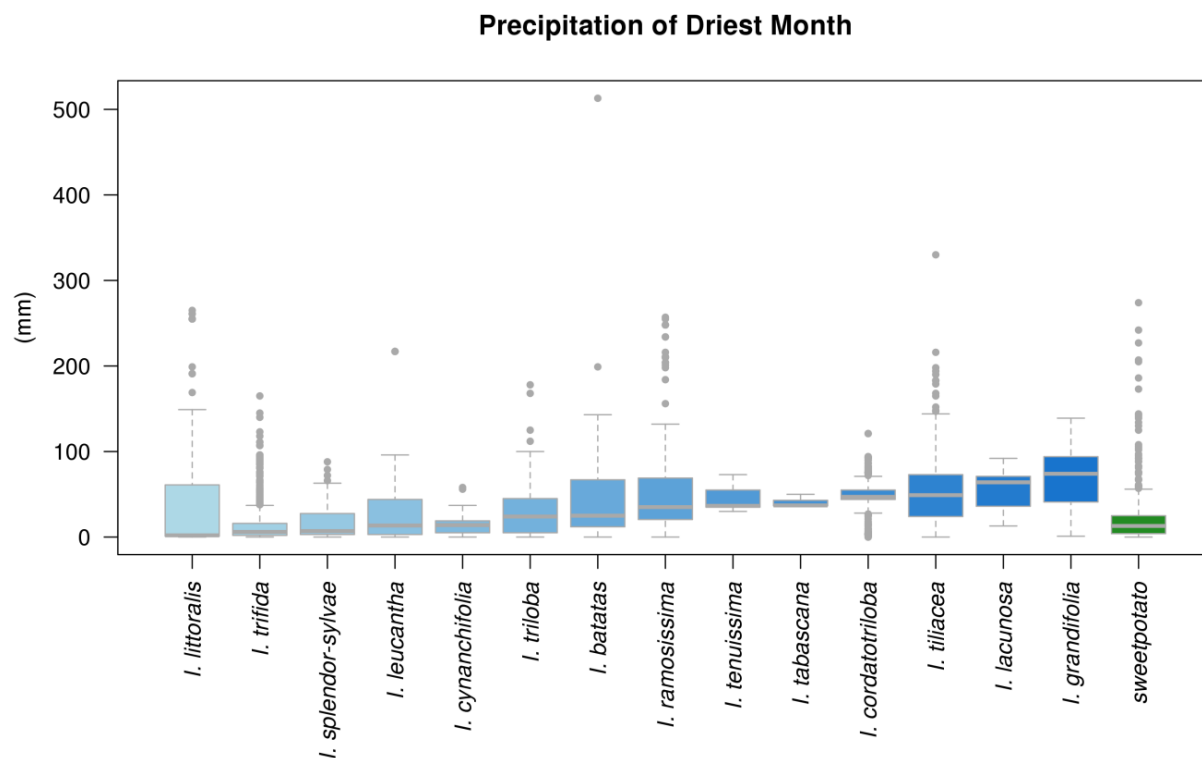
Supplementary Figure 7L. Climatic niches of CWR species and the sweetpotato crop for mean temperature of the coldest quarter of the year.



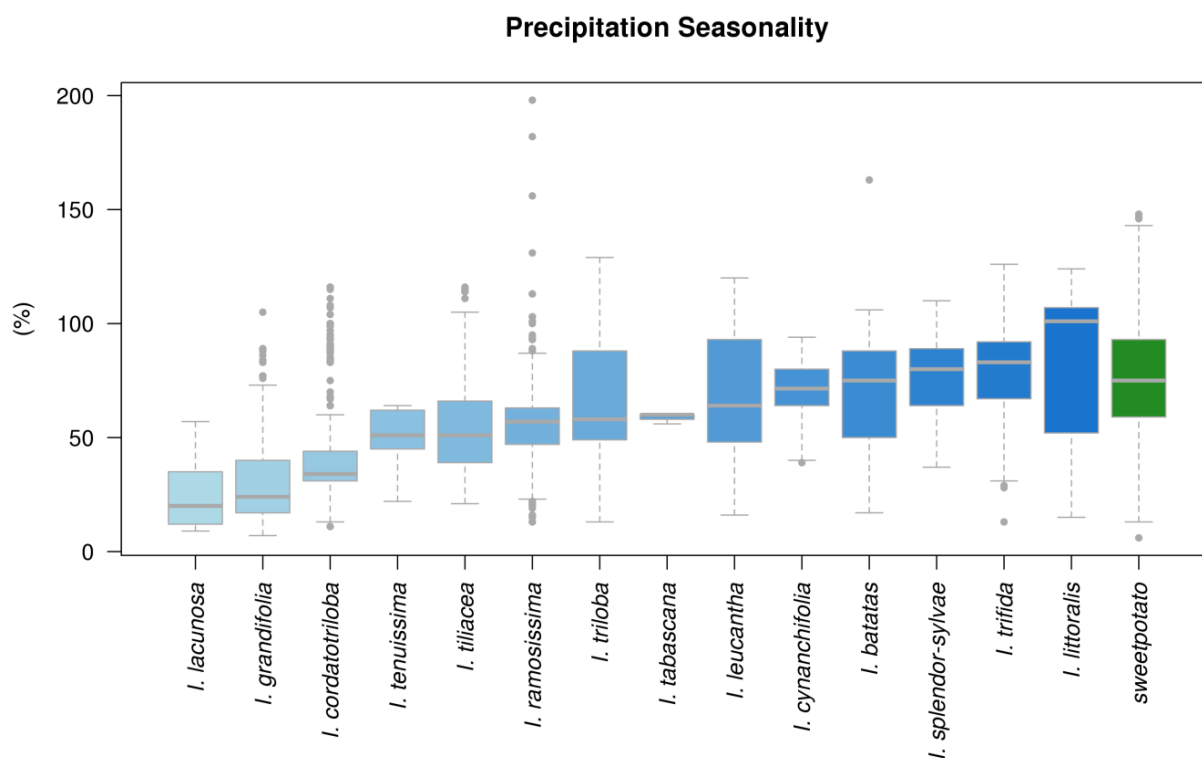
Supplementary Figure 7M. Climatic niches of CWR species and the sweetpotato crop for annual precipitation.



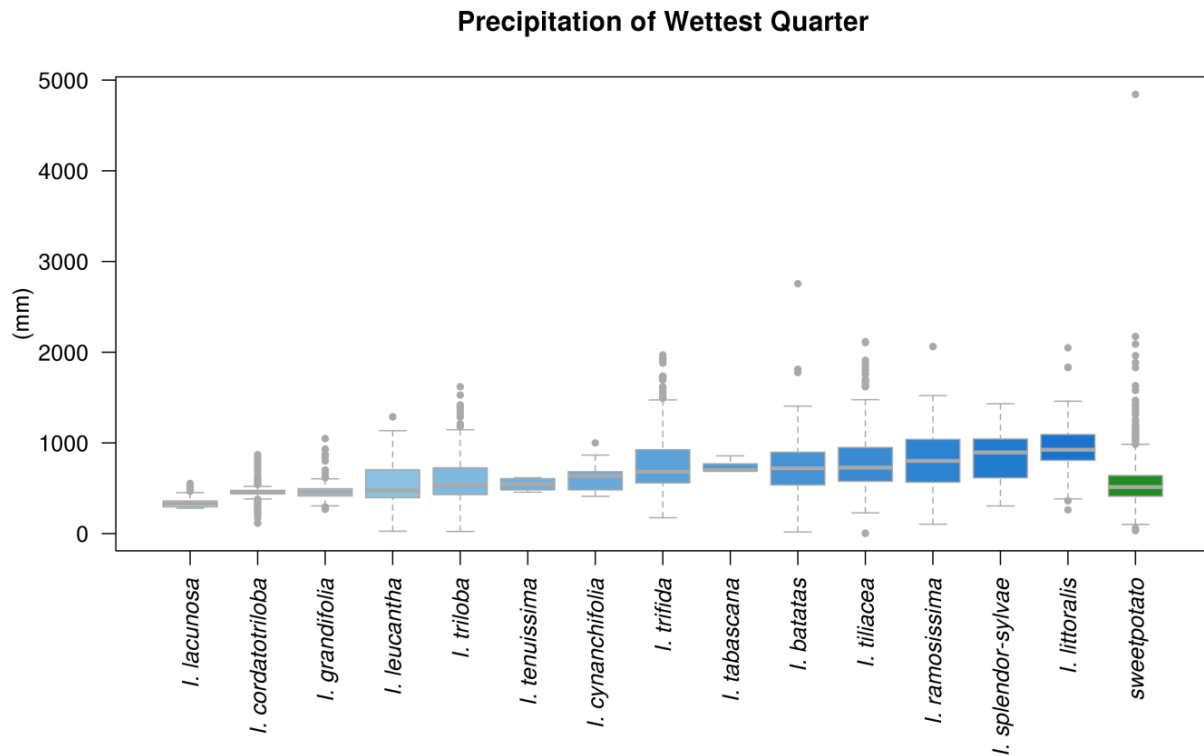
Supplementary Figure 7N. Climatic niches of CWR species and the sweetpotato crop for precipitation of the wettest month of the year.



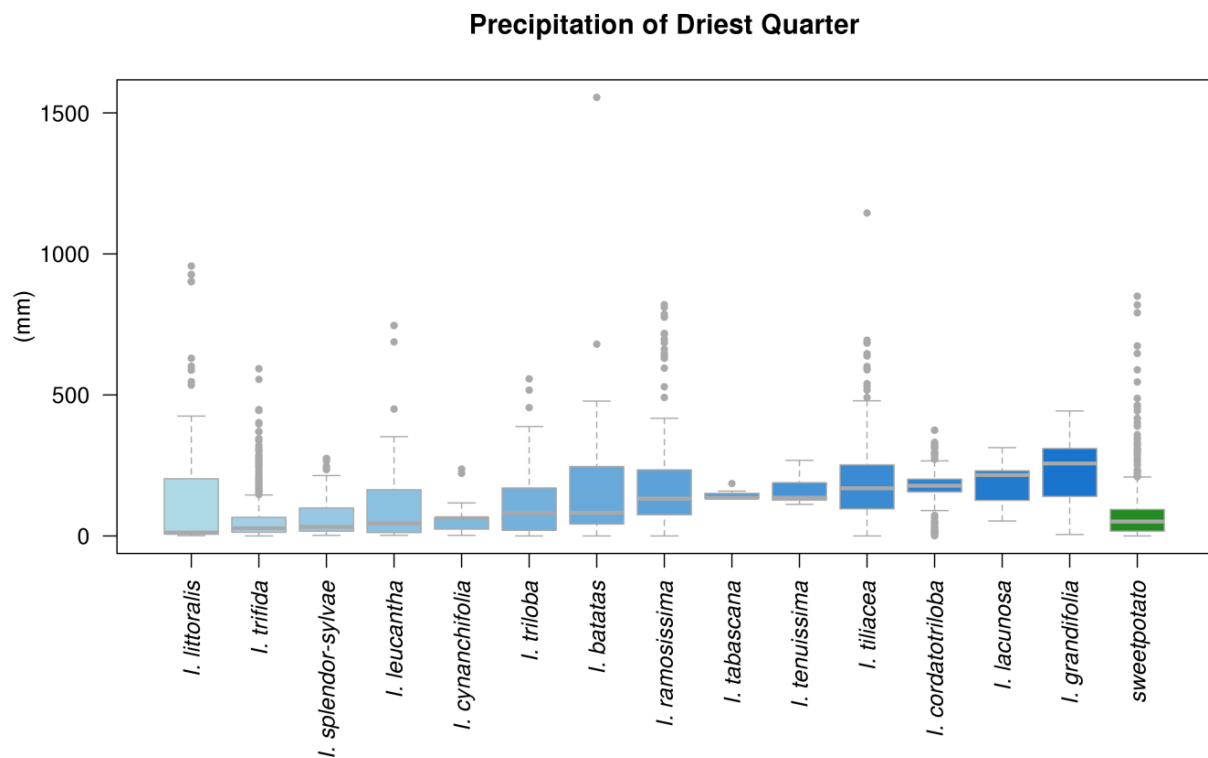
Supplementary Figure 7O. Climatic niches of CWR species and the sweetpotato crop for precipitation of the driest month of the year.



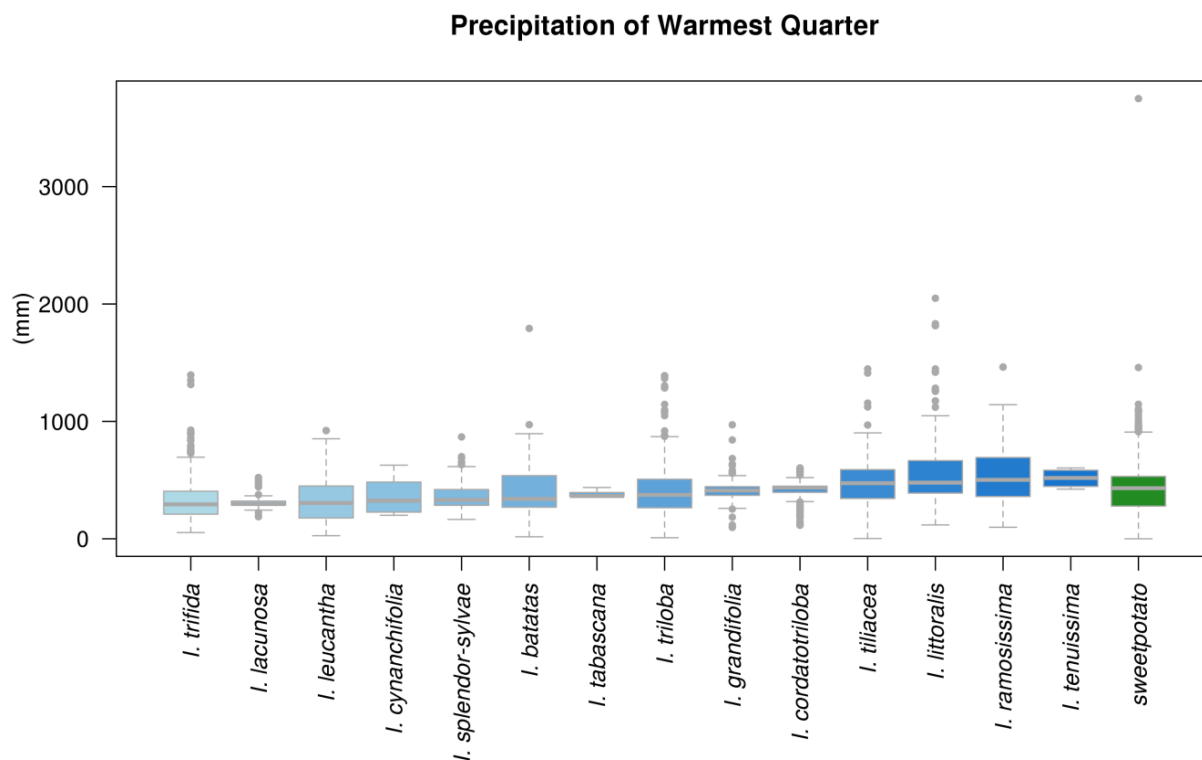
Supplementary Figure 7P. Climatic niches of CWR species and the sweetpotato crop for precipitation seasonality.



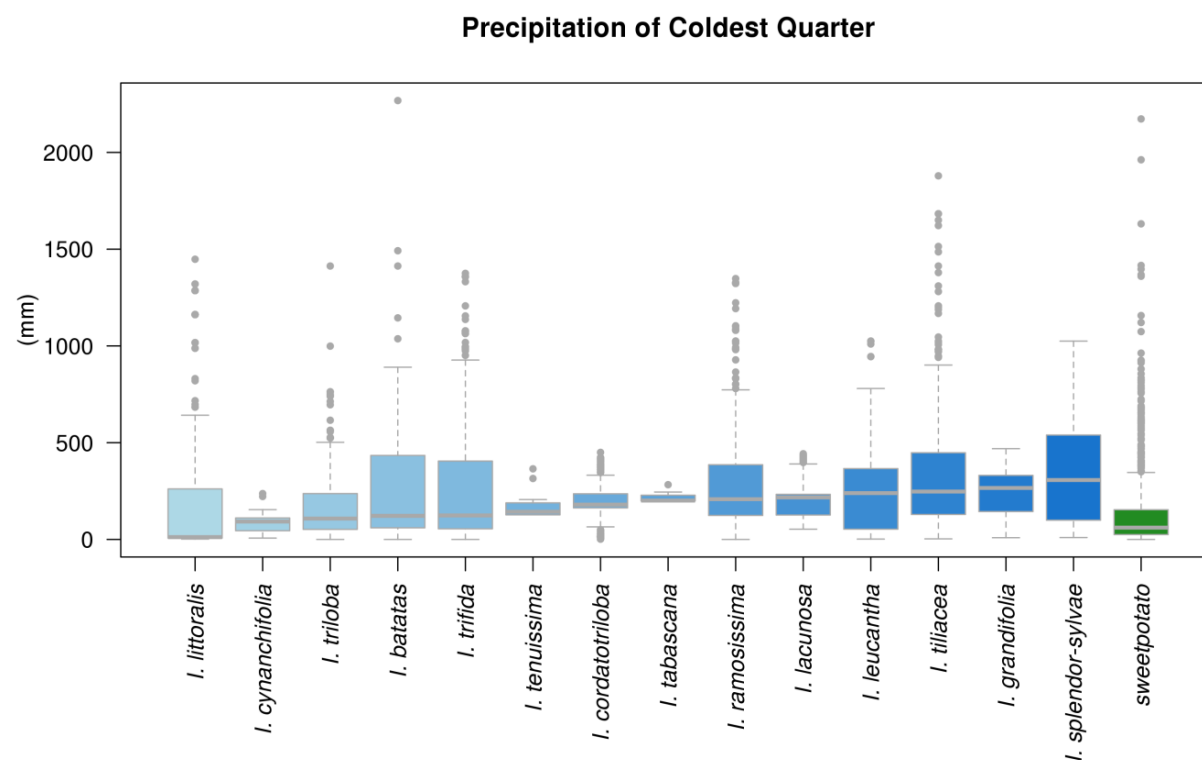
Supplementary Figure 7Q. Climatic niches of CWR species and the sweetpotato crop for precipitation of the wettest quarter of the year.



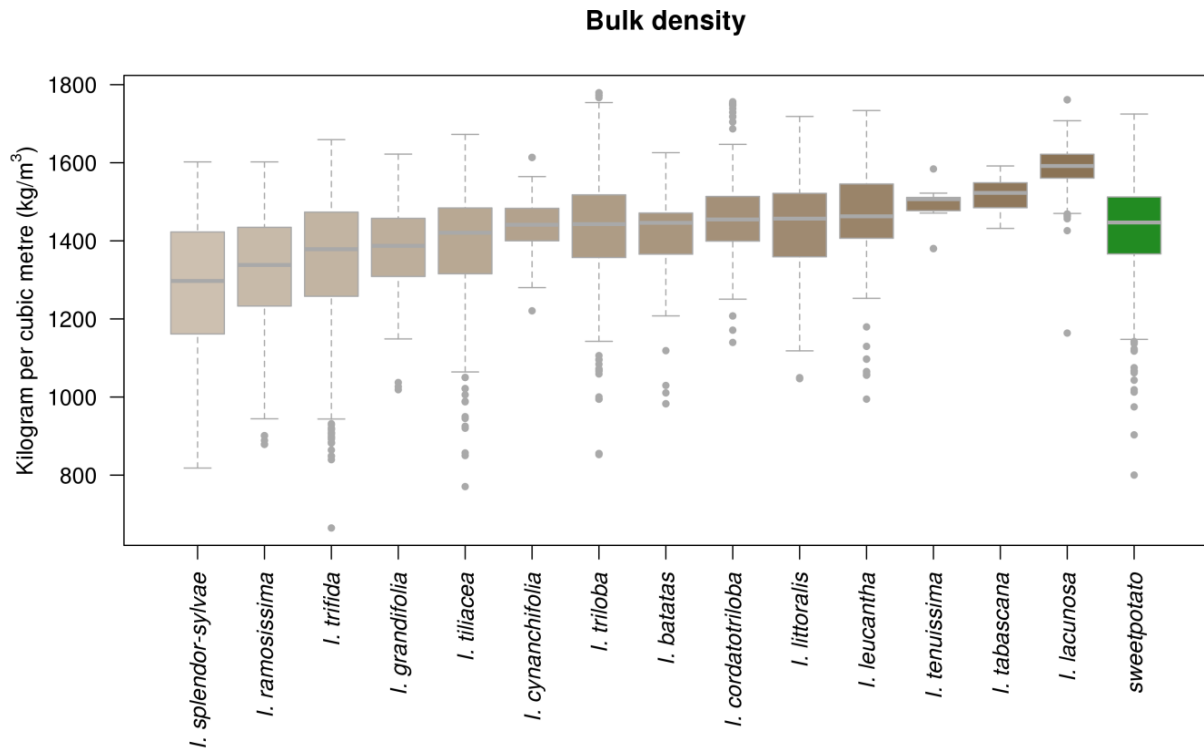
Supplementary Figure 7R. Climatic niches of CWR species and the sweetpotato crop for precipitation of the driest quarter of the year.



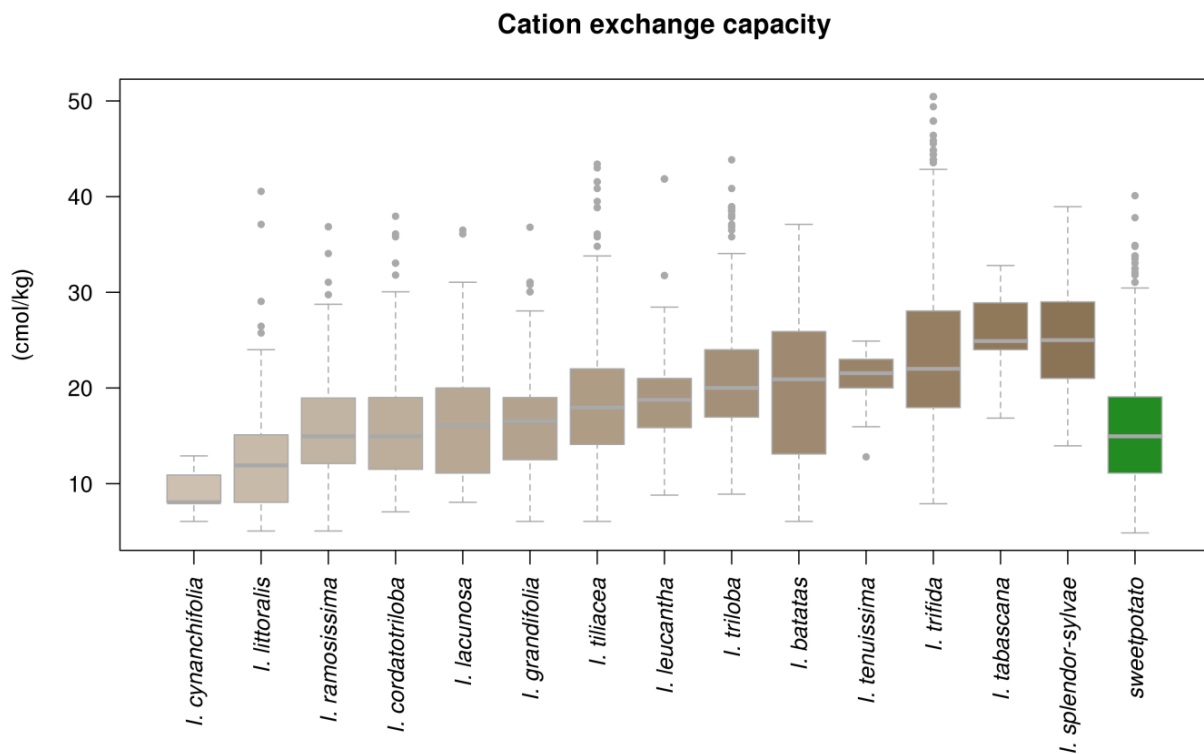
Supplementary Figure 7S. Climatic niches of CWR species and the sweetpotato crop for precipitation of the warmest quarter of the year.



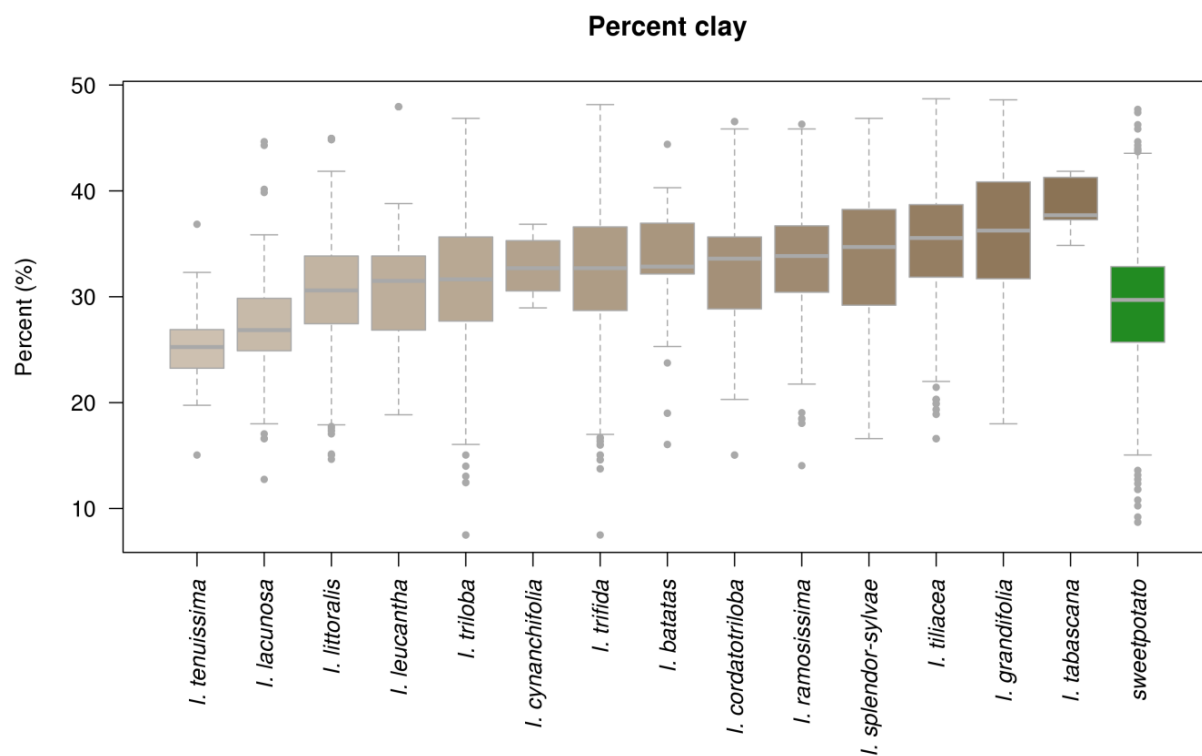
Supplementary Figure 7T. Climatic niches of CWR species and the sweetpotato crop for precipitation of the coldest quarter of the year.



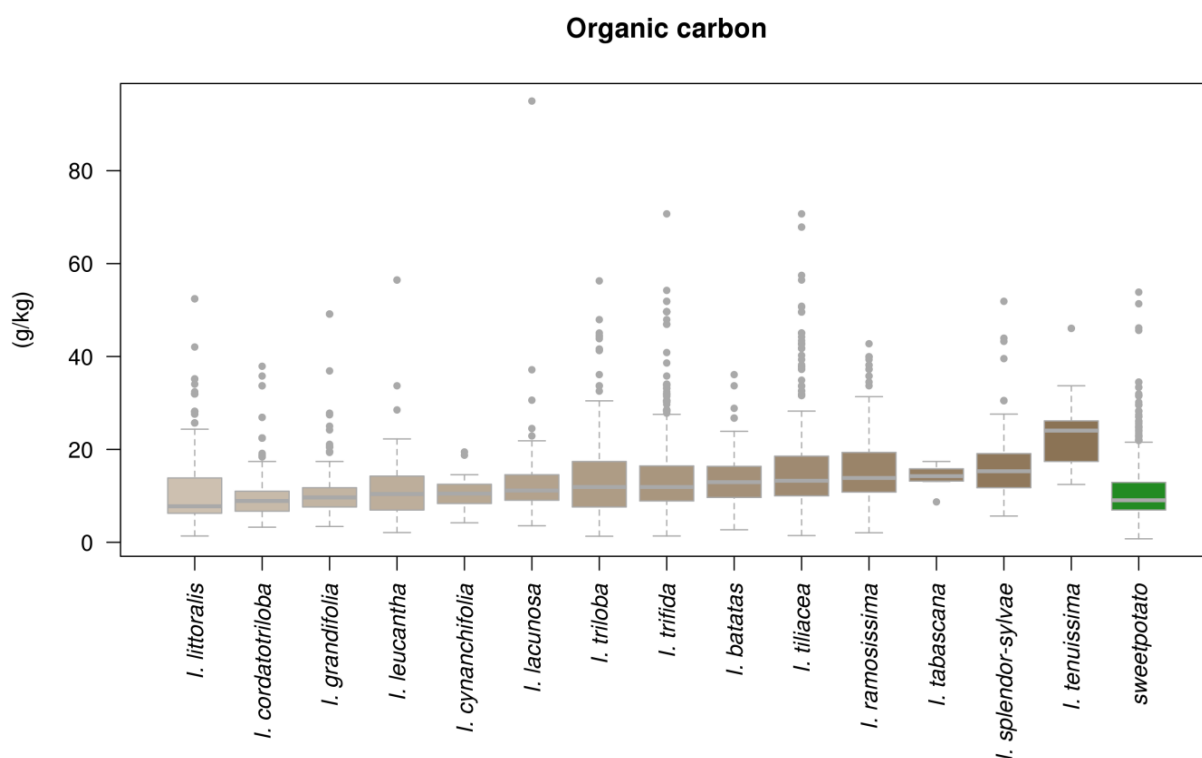
Supplementary Figure 7U. Edaphic niches of CWR species and the sweetpotato crop for bulk density.



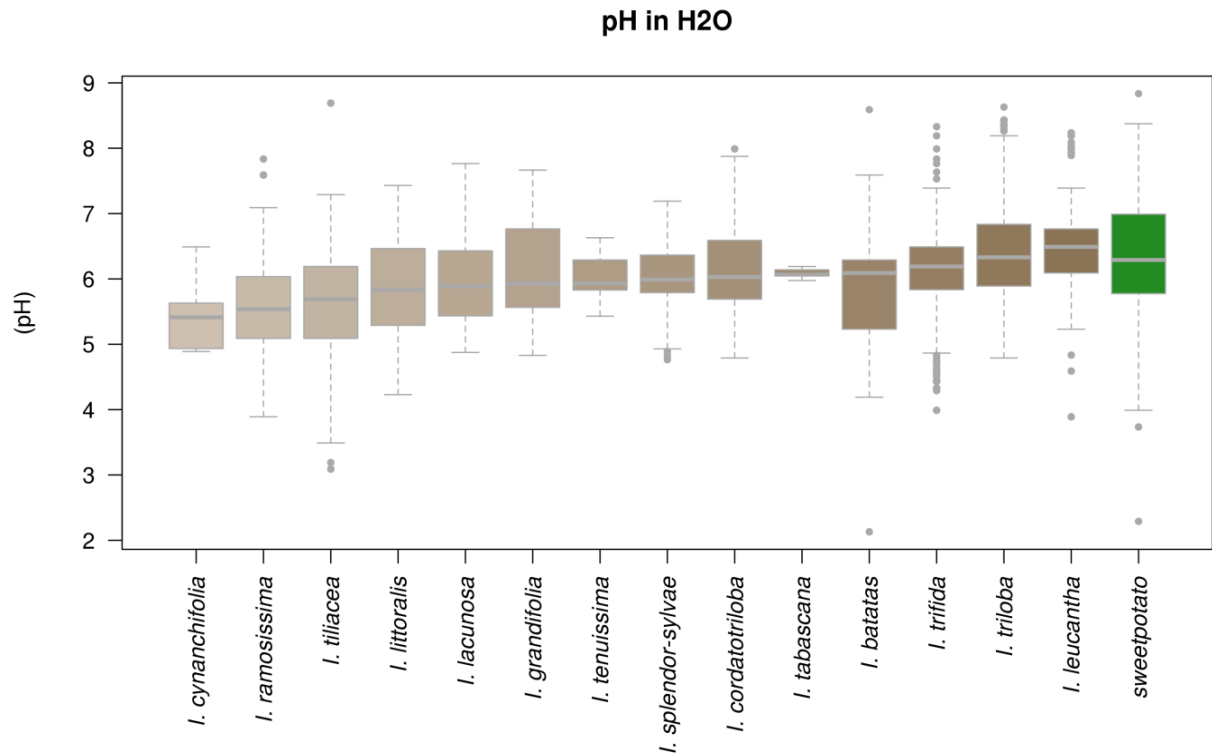
Supplementary Figure 7V. Edaphic niches of CWR species and the sweetpotato crop for cation exchange capacity.



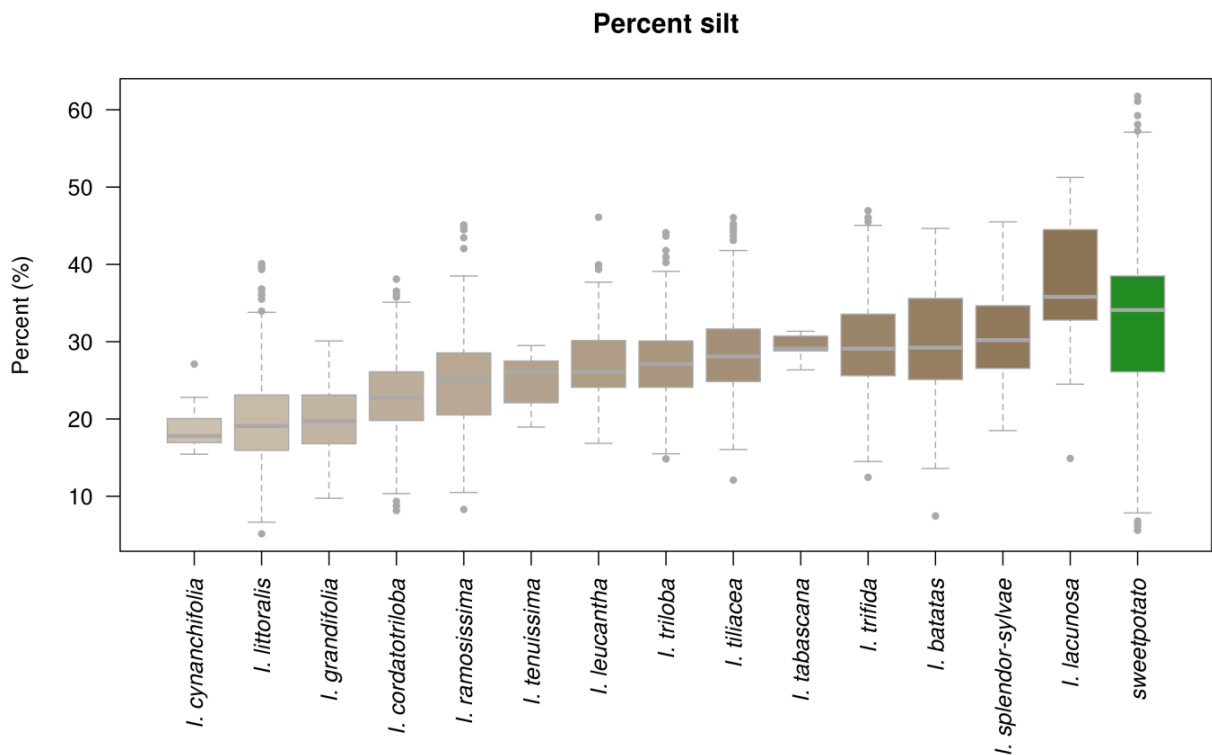
Supplementary Figure 7W. Edaphic niches of CWR species and the sweetpotato crop for percent clay.



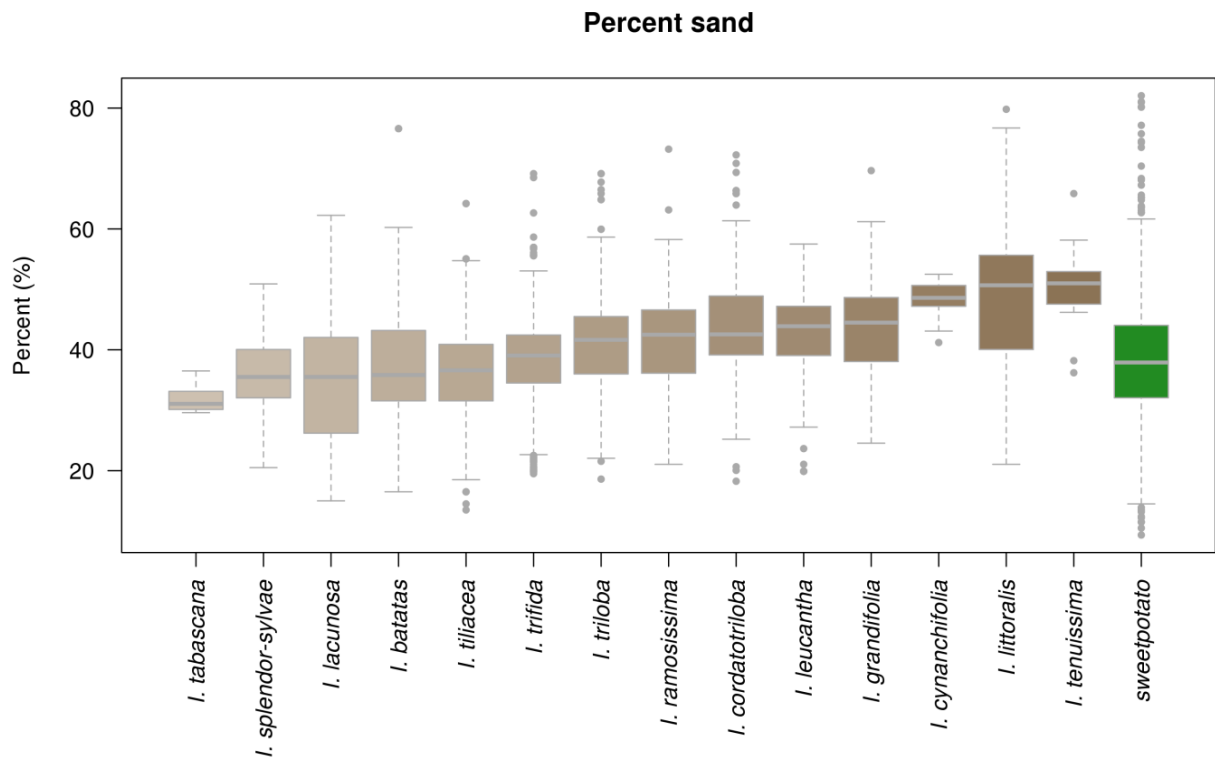
Supplementary Figure 7X. Edaphic niches of CWR species and the sweetpotato crop for organic carbon.



Supplementary Figure 7Y. Edaphic niches of CWR species and the sweetpotato crop for pH.



Supplementary Figure 7Z. Edaphic niches of CWR species and the sweetpotato crop for percent silt.



Supplementary Figure 7AA. Edaphic niches of CWR species and the sweetpotato crop for percent sand.

Supplementary Table 1. Sources of occurrence data for assessed sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative species.

Data provider	Record type	Number of records
International Potato Center (CIP)	Germplasm accession	585
Leibniz Institute of Plant Genetics and Crop Plant Research	Germplasm accession	1
Millennium Seed Bank Partnership, Royal Botanic Gardens, Kew	Germplasm accession	3
USDA, National Plant Germplasm System, Germplasm Resources Information Network (USDA NPGS GRIN)	Germplasm accession	160
Atlas of Living Australia (Flora Atlas NT) (via GBIF)	Herbarium or other record	356
Australia's Virtual Herbarium (AD, BRI, CANB, CBG, DNA, HO, MEL, NSW, PERTH)	Herbarium or other record	39
Bioversity International	Herbarium or other record	208
Brazil Virtual Herbarium (CRIA)	Herbarium or other record	61
California Academy of Sciences Herbarium (CAS)	Herbarium or other record	128
Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) (via GBIF)	Herbarium or other record	97
Consortium of Pacific Northwest Herbaria (CPNWH)	Herbarium or other record	5
DIVEA, DEP, FEEMA Herbario Alberto Castellanos (GUA)	Herbarium or other record	18
Fairchild Tropical Botanical Garden (FTG) (via GBIF)	Herbarium or other record	712
Florida State University Robert K. Godfrey Herbarium (FSU)	Herbarium or other record	6
Harvard University Herbarium (HUH)	Herbarium or other record	8
Herbarium of National Taiwan University (TAI) (via GBIF)	Herbarium or other record	63
Herbarium of Taiwan Forestry Research Institute (TAIF) (via GBIF)	Herbarium or other record	56
Instituto Nacional de Biodiversidad, Costa Rica (INB) (via GBIF)	Herbarium or other record	87
International Center for Tropical Agriculture (CIAT)- Bioversity International, USDA. Atlas of Paraguayan Crop Wild Relatives.	Herbarium or other record	155
International Center for Tropical Agriculture (CIAT)- C. Azurdia, K.A. Williams, D.E. Williams, V. Van Damme, A. Jarvis and S.E. Castaño. 2011. Atlas of Guatemalan Crop Wild Relatives.	Herbarium or other record	139
International Potato Center (CIP)	Herbarium or other record	231
Jardim Botanico de Rio de Janeiro (JABOT/R)	Herbarium or other record	106
Jardim Botânico do Rio de Janeiro (JBRJ) (via GBIF)	Herbarium or other record	119
Louisiana State University Herbarium (LSU) (via GBIF)	Herbarium or other record	116
Manchester University Herbarium (MANCH)	Herbarium or other record	2
McDonald & Austin (1990)	Herbarium or other record	1
Missouri Botanical Garden Herbarium (MO)	Herbarium or other record	44
Museo Nacional de Costa Rica (MNCR) (via GBIF)	Herbarium or other record	80
Museum national d'Histoire naturelle (MNHN) (via GBIF)	Herbarium or other record	509
Nationaal Herbarium Nederland (NHN)	Herbarium or other record	112
Natural History Museum UK Herbarium (BM)	Herbarium or other record	19
New York Botanical Garden Herbarium (NY)	Herbarium or other record	454
Plants of Taiwan	Herbarium or other record	56
Real Jardín Botánico de Madrid (MA)	Herbarium or other record	25
Royal Botanic Gardens, Kew (K)	Herbarium or other record	20
Scotland & Wood, personal communication	Herbarium or other record	8
Smithsonian Institution, National Herbarium (US)	Herbarium or other record	357
The Field Museum (F)	Herbarium or other record	58
Universidad del Valle Colombia Herbarium (CUVC)	Herbarium or other record	4
University of California Jepson Herbarium (UC/Jeps)	Herbarium or other record	63
University of California Riverside Herbarium (UCR)	Herbarium or other record	23
University of Kansas McGregor Herbarium (KU) (via GBIF)	Herbarium or other record	79
University of Oklahoma, Robert Bebb Herbarium (OKL) (via GBIF)	Herbarium or other record	53
US Academy of Natural Sciences Philadelphia Herbarium (PH)	Herbarium or other record	45
USDA, National Plant Germplasm System, Germplasm Resources Information Network (USDA NPGS GRIN)	Herbarium or other record	131
West Virginia University Herbarium (WVU)	Herbarium or other record	12

Supplementary Table 2. Ecogeographic variables utilized in sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative potential distribution modeling and climatic niche analyses.

Variable number	Variable name	Units
0	Altitude	m
1	Annual mean temperature	°C
2	Mean diurnal temperature range	°C
3	Isothermality	N/A
4	Temperature seasonality (standard deviation)	°C
5	Maximum temperature of warmest month	°C
6	Minimum temperature of coldest month	°C
7	Temperature annual range	°C
8	Mean temperature of wettest quarter	°C
9	Mean temperature of driest quarter	°C
10	Mean temperature of warmest quarter	°C
11	Mean temperature of coldest quarter	°C
12	Annual precipitation	mm
13	Precipitation of wettest month	mm
14	Precipitation of driest month	mm
15	Precipitation seasonality (coefficient of variation)	%
16	Precipitation of wettest quarter	mm
17	Precipitation of driest quarter	mm
18	Precipitation of warmest quarter	mm
19	Precipitation of coldest quarter	mm
20	Bulk density	kg/m ³
21	Cation exchange capacity	cmol/kg
22	Percent clay	%
23	Organic carbon	g/kg
24	pH in H ₂ O	pH
25	Percent silt	%
26	Percent sand	%

Supplementary Table 3. Gap analysis and comparable expert evaluation priorities results per sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative.

Taxon	Total Records	Total Records- Distinct	H Records- Total	H Records- Distinct	G Records- Total	G Records- Distinct	SRS	Ecogeographic Variables	ATAUC	STAUc	ASD15	cAUC	GRS	ERS	FPS	FPCAT	EPS 1	EPS 2	EPS 3	EPS 4	EPS 5	EPS mean
<i>I. batatas</i>	104	44	100	44	4	0	0.38	2, 4, 8, 13, 14, 19, 24	0.78	0.06	11.19	0.28	0.00	0.00	0.13	HPS	1	1	7	1	2	2.4
<i>I. cordatoriloba</i>	424	290	321	229	103	67	2.43	1, 2, 3, 9, 13, 15, 18	0.97	0.01	0.93	0.47	1.24	2.86	2.18	HPS	6	7	5	1	1	4
<i>I. cynanchifolia</i>	37	17	36	17	1	0	0.27	0, 15, 16, 18, 19, 22, 23, 26(*)	0.76	NA	NA	0.26	0.00	0.00	0.09	HPS	0	7	2	NA	0	2.3
<i>I. grandifolia</i>	460	334	336	254	124	83	2.70	3, 7, 8, 9, 10, 15, 16	0.95	0.01	0.02	0.45	2.54	7.22	4.15	MPS	3	3	5	1	1	2.6
<i>I. lacunosa</i>	413	252	403	251	10	1	0.24	0, 3, 5, 7, 9, 12, 13, 25	0.97	0.00	0.49	0.47	0.07	0.91	0.41	HPS	4	8	2	1	1	3.2
<i>I. leucantha</i>	134	93	116	79	18	15	1.34	1, 2, 3, 8, 12, 16, 20, 24	0.91	0.04	0.19	0.41	0.06	0.57	0.66	HPS	4	7	3	NA	2	4
<i>I. littoralis</i>	468	340	466	338	2	2	0.04	1, 2, 4, 15, 18, 19, 20, 23	0.98	0.00	0.62	0.48	0.02	0.07	0.05	HPS	0	3	2	0	0	1
<i>I. ramosissima</i>	341	217	307	187	34	30	1.00	0, 7, 11, 13, 19, 20, 23, 24	0.85	0.02	1.07	0.35	0.50	2.34	1.28	HPS	4	3	4	1	2	2.8
<i>I. splendendor-syhae</i>	177	135	161	130	16	9	0.90	0, 5, 11, 19, 20	0.90	0.04	2.76	0.40	0.97	4.88	2.25	HPS	1	7	3	1	0	2.4
<i>I. tabascanae</i>	8	6	4	4	4	2	5.00	22, 23, 24, 26(*)	0.87	NA	NA	0.37	10.00	4.29	6.43	LPS	0	1	2	0	0	0.6
<i>I. tenuissima</i>	34	12	31	12	3	1	0.88	10, 13, 14, 21(*)	0.95	NA	NA	0.45	0.32	1.05	0.75	HPS	0	5	3	0	0	1.6
<i>I. tiliacea</i>	740	441	679	400	61	44	0.82	0, 5, 11, 12, 17, 19, 20, 24	0.94	0.01	0.37	0.44	0.59	1.58	1.00	HPS	4	1	5	1	1	2.4
<i>I. trifida</i>	1409	1023	1161	888	248	159	1.76	0, 5, 11, 12, 13, 24	0.92	0.01	0.01	0.42	2.24	4.95	2.98	MPS	3	1	4	4	6	3.6
<i>I. triloba</i>	865	446	744	376	121	74	1.40	0, 2, 3, 9, 12, 16, 24	0.93	0.01	0.12	0.43	1.07	4.12	2.20	HPS	6	1	4	4	6	4.2

Distinct records denote occurrence records with unique coordinates. H records denote herbarium and other locality data used to inform species distribution models. G records denote germplasm accessions accessible to the global community. Ecogeographic variables lists those species-specific bioclimatic and edaphic variables used in Maxent modeling per taxon (Table S2); (*) denotes species for which an ensemble model was derived. Maxent assessment techniques include the 5-fold average area under the ROC curve of test data (AT AUC), the standard deviation of the test AUC of the 5 different folds (ST AUC), the proportion of the potential distribution coverage with standard deviation above 0.15 (ASD15), and the calibrated AUC (cAUC). The final priority score (FPS) for the species is the mean of the sampling representativeness score (SRS), geographic representativeness score (GRS), and ecological representativeness score (ERS). Final priority categories (FPCAT) for the CWR of sweetpotato included high (HPS), medium (MPS), and low (LPS) priority species for further collecting for *ex situ* conservation. EPS denotes comparable expert priority scores for 5 experts, along with their mean value.

Supplementary Table 4. Countries identified for further collecting per sweetpotato [*Ipomoea batatas* (L.) Lam.] crop wild relative.

Taxon	Priority	Countries identified for further collecting
<i>I. batatas</i>	HPS	Belize, Bolivia, Brazil, Colombia, Costa Rica, Ecuador, El Salvador, French Guiana, Guatemala, Guyana, Honduras, Mexico, Nicaragua, Panama, Peru, Suriname, Venezuela
<i>I. cordatotriloba</i>	HPS	Argentina, Bolivia, Brazil, Mexico, Paraguay, USA
<i>I. cynanchifolia</i>	HPS	Bolivia, Brazil
<i>I. grandifolia</i>	MPS	Argentina, Bolivia, Brazil, Paraguay, Uruguay
<i>I. lacunosa</i>	HPS	USA
<i>I. leucantha</i>	HPS	Mexico, USA
<i>I. littoralis</i>	HPS	Australia, Fiji, French Polynesia, Guam, India, Indonesia, Kiribati, Madagascar, Malaysia, Mauritius, Micronesia, N. Mariana Islands, New Caledonia, Papua New Guinea, Philippines, Seychelles, Sri Lanka, Taiwan, Thailand, Timor-Leste, USA
<i>I. ramosissima</i>	HPS	Belize, Bolivia, Brazil, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Venezuela
<i>I. splendor-syhae</i>	HPS	Belize, Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama
<i>I. tabascana</i>	LPS	Mexico
<i>I. tenuissima</i>	HPS	Cuba, Dominican Republic, Haiti, Puerto Rico, USA
<i>I. tiliacea</i>	HPS	Bahamas, Belize, Brazil, Colombia, Costa Rica, Cuba, Dominican Republic, El Salvador, French Guiana, Grenada, Guatemala, Guyana, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Puerto Rico, Sao Tome and Principe, St. Vincent and the Grenadines, Suriname, USA, Venezuela
<i>I. trifida</i>	MPS	Belize, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, USA, Venezuela
<i>I. triloba</i>	HPS	Bahamas, Belize, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Haiti, Honduras, Jamaica, Mexico, Nicaragua, Panama, Peru, USA, Venezuela

REFERENCES

- Araújo, M.B., Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33 (10), 1677-1688.
- Austin, D.F. (1978). The *Ipomoea batatas* complex - 1. Taxonomy. *Bull. Torrey Bot. Club* 105, 114-129.
- Austin, D.F. (1988). "The taxonomy, evolution and genetic diversity of sweetpotatoes and related wild species," in: *Exploration, Maintenance, and Utilization of Sweetpotato Genetic Resources*, ed. P. Gregory (Lima, Peru: International Potato Center), 27-60.
- Austin, D.F. (1991). *Ipomoea littoralis* (Convolvulaceae) - Taxonomy, distribution, and ethnobotany. *Econ. Bot.* 45, 251-256.
- Austin, D.F., (1997). Dissolution of *Ipomoea* series Anisomerae (Convolvulaceae). *Taxon* 28, 359-361.
- Austin, D.F., Huamán, Z. (1996). A synopsis of *Ipomoea* (Convolvulaceae) in the Americas. *Taxon* 45, 3-38.
- Bill and Melinda Gates Foundation. (2011). *Agricultural Development Strategy Overview*. Available at: <https://docs.gatesfoundation.org/Documents/agricultural-development-strategy-overview.pdf>
- Bivand, R., Lewin-Koh, N. (2014). Maptools: tools for reading and handling spatial objects. R package version 0.8-30. <http://CRAN.R-project.org/package=maptools>.
- Bivand, R., Keitt, T., Rowlingson, B. (2014). Rgdal: bindings for the geospatial data abstraction library. R package version 0.8-16. <http://CRAN.R-project.org/package=rgdal>.
- Bivand, R., Pebesma, E., Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*, Second edition. NY, Springer. <http://www.asdar-book.org/>.
- Bouis, H., Islam, Y. (2012). *Delivering nutrients widely through biofortification: building on orange sweet potato - scaling up in agriculture, rural development and nutrition*. Focus 19, Brief 11. (Washington DC: International Food Policy Research Institute).
- Brar, D.S., Khush, G.S. (1997). Alien introgression in rice. *Plant Mol. Biol.* 35 (1-2), 35-47.
- Cao, Q., Zhang, A., Ma, D., Li, H., Li, Q., Li, P. (2009). Novel interspecific hybridization between sweetpotato (*Ipomoea batatas* (L.) Lam.) and its two diploid wild relatives. *Euphytica* 169, 345-352.
- Conolly, J., Manning, K., Colledge, S., Dobney, K., Shennan, S. (2012). Species distribution modelling of ancient cattle from early neolithic sites in SW Asia and Europe. *Holocene* 22, 997-1010.
- Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodivers. Conserv.* 19, 883-899.
- De Vries, C.A., Ferwerda, J.D., Flach, M. (1967). Choice of food crops in relation to actual and potential production in the tropics. *Neth. J. Agr. Sci.* 15, 241-248.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: a global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sust. Food* 38, 369-377.
- Diaz, J., Schmiediche, P., Austin, D.F. (1996). Polygon of crossability between eleven species of *Ipomoea*: section *Batatas* (Convolvulaceae). *Euphytica* 88, 189-200.
- Duncan, T.M., Rausher, M.D. (2013). Evolution of the selfing syndrome in *Ipomoea*. *Front. Plant Sci.* 4, 301.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- Eserman, L.A., Tiley, G.P., Jarret, R.L., Leebens-Mack, J.H., Miller, R.E. (2014). Phylogenetics and diversification of morning glories (tribe Ipomoeae, Convolvulaceae) based on whole plastome sequences. *Am. J. Bot.* 101, 92-103.

- ESRI. (2011). ArcGIS Desktop: Release 10.1. Available at: <http://www.esri.com/software/arcgis/arcgis-for-desktop>.
- FAO (Food and Agriculture Organization of the United Nations). (2002). *International Treaty on Plant Genetic Resources for Food and Agriculture*. Rome, Italy: FAO.
- FAO (Food and Agriculture Organization of the United Nations). (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture*. Rome, Italy: FAO.
- Feuillet, C., Langridge, P., Waugh, R. (2008). Cereal breeding takes a walk on the wild side. *Trends Genet.* 24, 24-32.
- Fieberg, J., Kochanny, C.O. (2005). Quantifying home-range overlap: the importance of the utilization distribution. *J. Wildlife Manage.* 69, 1346-1359.
- Gaiji, S., Chavan, V., Ariño, A.H., Otegui, J., Hobern, D., Sood, R., *et al.* (2013). Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. *Biodiversity Informatics* 8, 94-172.
- Gichuki, S.T., Berenyi, M., Zhang, D., Hermann, M., Schmidt, J., Glössl, J., *et al.* (2003). Genetic diversity in sweetpotato [*Ipomoea batatas* (L.) Lam.] in relationship to geographic sources as assessed with RAPD markers. *Genet. Resour. Crop Ev.* 50, 429-437.
- Google Maps Geocoder. (2013). Version 3. <https://developers.google.com/maps/documentation/geocoding/?hl=pl>.
- GRIN (Germplasm Resources Information Network), USDA, ARS, National Genetic Resources Program. (2013). USDA, ARS, National Genetic Resources Program. <http://www.ars-grin.gov/cgi-bin/npgs/html/wep.pl>.
- Guarino, L., Lobell, D.B. (2011). A walk on the wild side. *Nature Clim. Change* 1 (8), 374-375.
- Hajjar, R., Hodgkin, T. (2007). The use of wild relatives for crop improvement: a survey of developments over the past 20 years. *Euphytica* 156, 1-13.
- Harlan, J.R., de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20: 509-517.
- Heibl, C. (2011). Phyloclim: integrating phylogenetics and climatic niche modeling. v.0.9-4. <http://cran.r-project.org/web/packages/phyloclim/index.html>.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., *et al.* (2014). SoilGrids1km - global soil information based on automated mapping. *PLoS One* 9 (8), e105992.
- Hijmans, R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93: 679-688.
- Hijmans, R.J. (2014). Raster: geographic data analysis and modeling. R package version 2.2-31. <http://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Cameron, S., Parra, J., Jones, P. G., Jarvis, A. (2005). *WorldClim, Version 1.3* (Berkeley, University of California).
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. (2013). Dismo: Species distribution modeling. R package version 0.9-3. <http://CRAN.R-project.org/package=dismo>.
- Hijmans, R.J., Schreuder, M., De La Cruz, J., Guarino, L. (1999). Using GIS to check co-ordinates of genebank accessions. *Genet. Resour. Crop Ev.* 46, 291-296.
- Hjarding, A., Tolley, K.A., Burgessa, N.D. (2014). Red List assessments of East African chameleons: a case study of why we need experts. *Oryx* First View 1-7. doi: 10.1017/S0030605313001427.
- Hotz, C., Loechl, C., Lubowa, A., Tumwine, J.K., Ndeezi, G., Nandutu Masawi, A., *et al.* (2012). Introduction of β -carotene-rich orange sweet potato in rural Uganda resulted in increased vitamin A intakes among children and women and improved vitamin A status among children. *J. Nutr.* 142, 1871-1880.
- Huaman, Z. (1987). "Current status on maintenance of sweetpotato genetic resources at CIP," in:

- Exploration, maintenance and utilization of sweetpotato genetic resources*. Report of the First Sweetpotato Planning Conference 1987 (Lima: International Potato Center [CIP]), 101-120.
- Huang, J., Corke, H., Sun, M. (2002). Highly polymorphic AFLP markers as a complementary tool to ITS sequences in assessing genetic diversity and phylogenetic relationships of sweetpotato (*Ipomoea batatas* (L.) Lam.) and its wild relatives. *Genet. Resour. Crop Ev.* 49, 541-550.
- Husson, F., Josse, J., Le, S., Mazet, J. (2009). FactoMineR: multivariate exploratory data analysis and data mining with R. v.1.26. <http://factominer.free.fr>.
- Iwanaga, M. (1988). "Use of wild germplasm for sweetpotato breeding," in: *Exploration, Maintenance, and Utilization of Sweetpotato Genetic Resources*, ed. P. Gregory (Lima: International Potato Center [CIP]), 199-210.
- Jarret, R.L., Austin, D.F. (1994). Genetic diversity and systematic relationships in sweetpotato (*Ipomoea batatas* (L.) Lam.) and related species as revealed by RAPD analysis. *Genet. Resour. Crop Ev.* 41, 165-173.
- Jarret, R.L., Gawel, N., Whittemore, A. (1992). Relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *J. Am. Soc. Hortic. Sci.* 117, 633-637.
- Kays, S.J. (1985). "The physiology of yield in sweet potato," in: *Sweet Potato Products: A Natural Resource of the Tropics*, ed. J. Bouwkamp (Boca Raton, Florida: CRC Press), 19-133.
- Kays, S.J., Kays, S.E. (1998). "Sweet potato chemistry in relation to health," in: *Sweet Potato Production System Towards the 21st Century* (Miyakonojo: Kyushu National Agricultural Experimental Station), 231-272.
- Kernohan, B.J., Girzen, R.A., Millspaugh, J.J. (2001). "Analysis of animal space use and movements," *Radio tracking animal populations*. eds. J.J. Millspaugh, and J.M. Marzluff (San Diego: Academic Press), 125-166.
- Khoury, C.K., Castañeda Álvarez, N.P., Achicanoy, H., Sosa, C.C., Bernau, V., Kassa, M.T., *et al.*, (2015). Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: distributions, *ex situ* conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* 184, 259-270.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57, 625-639.
- Kivuva, B.M., Githiri, S.M., Yencho, C., Sibiya, J. (2015a). Combining ability and heterosis for yield and drought tolerance traits under managed drought stress in sweetpotato. *Euphytica* 201, 423-40.
- Kivuva, B.M., Githiri, S.M., Yencho, C., Sibiya, J. (2015b). Screening sweetpotato genotypes for tolerance to drought stress. *Field Crop Res.* 171, 11-22.
- Komaki, K. (2004). "Breeding value of wild species closely related to sweetpotato," in: *Proceedings of international workshop on production, utilization and development of sweetpotato*, Korea, 164-172.
- Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schroder, B., Lindenborn, J., Reinfelder, V., *et al.* (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19, 1366-1379.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28 (3), 385-393.
- Loiselle, B., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., *et al.* (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35, 105-116.
- Low, J.W., Arimond, M., Osman, N., Cunguara, B., Zano, F., Tschirley, D. (2007). A food-based approach introducing orange-fleshed sweet potatoes increased vitamin A intake and serum retinol concentrations in young children in rural Mozambique. *J. Nutr.* 137, 1320-1327.
- Low, J., Lynam, J., Lemaga, B., Crissman, C., Barker, I., Thiele, G., *et al.* (2009). "Sweetpotato in Sub-

- Saharan Africa,” in: *The Sweet Potato*, eds. G. Loebenstein, G. Thottappilly (Springer Netherlands), 359-390.
- Lu, S., Li, T. (1992). Study on the characterization of intra and inter-specific incompatibility of *Batatas* section. *Acta Agron. Sin.* 18, 161-168.
- Mabberley, D.J. (2008). *Mabberley's Plant Book: a portable dictionary of plants, their classifications, and uses, 3rd Edition* (Cambridge: Cambridge University Press).
- Martin, F.W. (1970). Self- and interspecific incompatibility in the Convolvulaceae. *Bot. Gaz.* 131, 139-144.
- Martin, F.W. (1982). “Analysis of the incompatibility and sterility of sweetpotato,” in: *Sweetpotato, Proceedings of 1st international symposium*, No. 82-172, eds. R.L. Villareal, T.D. Griggs (Tainan: AVRDC), 275-283.
- Martin, F.W., Jones, A. (1973). The species of *Ipomoea* closely related to the sweetpotato. *Econ. Bot.* 26, 201-215.
- McCouch, S.R., Sweeney, M., Li, J., Jiang, H., Thomson, M., Septinginsih, E., *et al.* (2007). Through the genetic bottleneck: *O. rufipogon* as a source of trait-enhancing alleles for *O. sativa*. *Euphytica* 154, 317-339.
- McDonald, J.A., Austin, D.F. (1990). Changes and additions in *Ipomoea* section *Batatas* (Convolvulaceae). *Brittonia* 42, 116-120.
- McDonald, J.A., Hansen, D.R., McDill, J.R., Simpson, B.B. (2011). A phylogenetic assessment of breeding systems and floral morphology of North American *Ipomoea* (Convolvulaceae). *J. Bot. Res. Inst. Texas* 5, 159-177.
- Mesibov, R. (2013). A specialist's audit of aggregated occurrence records. *ZooKeys* 293, 1-18.
- Miller, R.E., Rausher, M.D., Manos, P.S. (1999). Phylogenetic systematics of *Ipomoea* (Convolvulaceae) based on ITS and waxy sequences. *Syst. Bot.* 24, 209-227.
- Monfreda, C., Ramankutty, N., Foley, J.A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles* 22: GB1022. Data available at <http://www.geog.mcgill.ca/landuse/pub/Data/175crops2000/>.
- Munyiza, H., Stevenson, P., Mwanga, R., Talwana, H., Murumu, J., Odongo, B. (2007). The relationship between stem base and root damage by *Cylas* spp. on sweet potato. *Afr. Crop Sci. J.* 8, 955-957.
- Mwanga, R., Ssemakula, G. (2011). Orange-fleshed sweet potatoes for food, health and wealth in Uganda. *Int. J. Agric. Sustain.* 9, 42-49.
- Mwololo, J.K., Mburu, M.W.K., Njeru, R.W., Ateka, E.M., Kiarie, N., Munyua, J.K., *et al.* (2007). Resistance of sweet potato genotypes to sweet potato virus disease in Coastal Kenya. *Proc. ACSS Conf.* 8, 2083-2086.
- Ngailo, S., Shimelis, H., Sibiya, J., Mtunda, K. (2013). Sweet potato breeding for resistance to sweet potato virus disease and improved yield: Progress and challenges. *Af. J. Ag. Resear.* 8, 3202-3215.
- Nimmakayala, P., Vajja, G., Reddy, U.K. (2011). “*Ipomoea*,” in: *Wild Crop Relatives: Genomic and Breeding Resources*, ed. C. Kole (Berlin: Springer), 123-132.
- Nishiyama, I. (1971). Evaluation and domestication of sweetpotato. *Botanical Magazine Tokyo* 84, 377-387.
- Nishiyama, I. (1982). Autohexaploid evolution of the sweetpotato. In: *Sweetpotato. Proceedings of 1st international symposium*, No. 82-172, eds. R.L. Villareal, T.D. Griggs (Tainan: AVRDC), 263-274.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E. C., *et al.* (2001). Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51, 933-938.
- Otegui, J., Ariño, A.H., Encinas, M.A., Pando, F. (2013). Assessing the primary data hosted by the

- Spanish node of the Global Biodiversity Information Facility (GBIF). *PLoS One* 8: e55144.
- Pebesma, E.J., Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News* 5 (2). Available at: <http://cran.r-project.org/doc/Rnews/>.
- Pfeiffer, W.H., McClafferty, B. (2007). HarvestPlus: breeding crops for better nutrition. *Crop Sci.* 47, 88-105.
- Phillips, S.J., Dudik, M. (2008). Modeling of species distributions with Maxent: new extensions and comprehensive evaluation. *Ecography* 31, 161-175.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190, 231-59.
- Piperno, D.R., Pearsall, D.M. (1998). *The Origins of Agriculture in the Lowland Neotropics*. San Diego: Academic Press.
- R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. <http://www.r-project.org/>.
- Rajapakse, S., Nilmalgoda, S.D., Molnar, M., Ballard, R.E., Austin, D.F., Bohac, J.R. (2004). Phylogenetic relationships of the sweetpotato in *Ipomoea* series *Batatas* (Convolvulaceae) based on nuclear β -amylase gene sequences. *Mol. Phylogenet. Evol.* 30, 623-632.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. *PLoS One* 5, e13497.
- Revelle, W. (2015). *Psych: Procedures for Psychological, Psychometric, and Personality Research*, v. 1.5.1. <http://cran.r-project.org/web/packages/psych/index.html>
- Roullier, C., Duputié, A., Wennekes, P., Benoit, L., Fernández Bringas, V. M., Rossel, G., *et al.* (2013). Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS One* 8, e62707.
- Sanchez, G. (2012). *Package 'plsdepot'. Partial Least Squares (PLS) Data Analysis Methods*, v. 0.1.17. <http://cran.r-project.org/web/packages/plsdepot/index.html>.
- Schafleitner, R., Tincopa, L.R., Palomino, O., Rossel, G., Robles, R.F., Alagon, R., *et al.* (2010). A sweet potato gene index established by de novo assembly of pyrosequencing and Sanger sequences and mining for gene-based microsatellite markers. *BMC Genomics* 11, 604.
- Shady Solis, R., Haas, J., Creamer, W. (2001). Dating caral, a preceramic site in the Supe valley on the central coast of Peru. *Science* 292, 723-726.
- Shiotani, I., Huang, Z.Z., Sakamoto, S., Miyazaki, T. (1994). The role of the wild *Ipomoea trifida* germplasm in sweet potato breeding. *Acta Hort.* (ISHS) 380, 388-398.
- Smit, N.E.J.M. (1997). *Integrated Pest Management for Sweetpotato in Eastern Africa*. PhD thesis, (Wageningen: Wageningen University).
- Srisuwan, S., Sihachakr, D., Siljak-Yakovlev, S. (2006). The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Sci.* 171, 424-433.
- Sutherland, J.A. (1986). A review of the biology and control of the sweetpotato weevil, *Cylas formicarius elegaluntus* (F). *Trop. Pest Manag.* 32, 304-315.
- Teramura, T. (1979). Phylogenetic study of *Ipomoea* species in the section *Batatas*. *Mem. Coll. Agric., Kyoto Univ.* 114, 29-48.
- The Plant List. (2010). Version 1. <http://www.theplantlist.org/>.
- Tumwegamire, S., Kapinga, R., Rubaihayo, P.R., LaBonte, D.R., Grüneberg, W.J., Burgos, G., *et al.* (2011). Evaluation of dry matter: protein, starch, β -carotene, iron, zinc, calcium, and magnesium in East African sweetpotato [*Ipomoea batatas* (L.) Lam] germplasm. *Hort. Science* 46, 348-357.
- Valverde, R.A., Clark, C.A., Valkonen, J.P.T. (2007). Viruses and virus disease complexes of sweetpotato. *Plant Viruses* 1, 116-126.

- van der Wal, J., Falconi, L., Januchowski, S., Shoo, L., Storlie, C. (2014). *SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises*. R package version 1.1-20. <http://CRAN.R-project.org/package=SDMTools>.
- Warren, D.L., Glor, R.E., Turelli, M. (2008). Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62, 2868-2883.
- Wilkin, P. (1999). A morphological cladistic analysis of the Ipomoeae (Convolvulaceae). *Kew Bull.* 54, 853-876.
- Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brian, J.M., *et al.* (2009). Using species distribution models to predict new occurrences for rare plants. *Divers. Distrib.* 15, 565-576.
- Wood, J.R.I., Carine, M.A., Harris, D., Wilkin, P., Williams, B., Scotland, R.W. (2015). *Ipomoea* (Convolvulaceae) in Bolivia. *Kew Bull.* In press.
- Woolfe, J.A. (1992). *Sweetpotato, an untapped food source* (New York: Cambridge University Press).
- Xiao, J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D. (1996). Genes from wild rice improve yield. *Nature* 384, 223-224.
- Zhang, D.P., Cervantes, J., Huamán, Z., Carey, E., Ghislain, M. (2000). Assessing the genetic diversity of sweetpotato (*Ipomoea batatas* (L.) La.) cultivars from tropical America using AFLP. *Genet. Resour. Crop Ev.* 47, 659-665.
- Zhang, S.S., Liu, L.F. (2005). "Utilization of the wild relatives in sweetpotato breeding," in: *Sweetpotato Breeding and Industrialization in China*, ed. Q.C. Liu. (Beijing: China Agricultural University Press), 29-32.

CHAPTER 6

Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: distributions, *ex situ* conservation status, and potential genetic resources for abiotic stress tolerance

Colin K. Khoury^{1,2}, Nora P. Castañeda-Álvarez^{1,3}, Harold A. Achicanoy¹, Chrystian C. Sosa¹,
Vivian Bernau^{1,4}, Mulualet T. Kassa⁵, Sally L. Norton⁶, L. Jos G. van der Maesen⁷,
Hari D. Upadhyaya^{8,9,10}, Julian Ramírez-Villegas^{1,11,12}, Andy Jarvis^{1,11} and Paul C. Struik²

¹ International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia.

² Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

³ School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.

⁴ Department of Horticulture and Crop Science, The Ohio State University, 202 Kottman Hall, 2021 Coffey Rd, Columbus, Ohio 43210, USA.

⁵ Cereal Research Centre, Agriculture and Agri-Food Canada, 195 Dafoe Road, Winnipeg, Manitoba, T3T 2M9, Canada.

⁶ Australian Grains Genebank, Department of Environment and Primary Industries, Private Bag 260, Horsham, Victoria 3401, Australia.

⁷ Naturalis Biodiversity Center, Darwinweg 2, 2333 CR, Leiden, the Netherlands.

⁸ International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru 502 324, Telangana, India.

⁹ Department of Agronomy, Kansas State University, Manhattan, Kansas 66506, USA.

¹⁰ Institute of Agriculture, University of Western Australia, Crawley WA 6009, Australia.

¹¹ CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, Cali, Colombia.

¹² School of Earth and Environment, University of Leeds, Leeds, UK.

Published in *Biological Conservation* 184, 259-270. doi: 10.1016/j.biocon.2015.01.032

Abstract

Pigeonpea [*Cajanus cajan* (L.) Millsp.] is a versatile, stress-tolerant, and nutritious grain legume, possessing traits of value for enhancing the sustainability of dry sub-tropical and tropical agricultural systems. The use of crop wild relatives (CWR) in pigeonpea breeding has been successful in providing important resistance, quality, and breeding efficiency traits to the crop. Current breeding objectives for pigeonpea include increasing its tolerance to abiotic stresses, including heat, cold, drought, and waterlogging. Here we assess the potential for pigeonpea CWR to be further employed in crop improvement by compiling wild species occurrence and *ex situ* conservation information, producing geographic distribution models for the species, identifying gaps in the comprehensiveness of current germplasm collections, and using ecogeographic information to identify CWR populations with the potential to contribute agronomic traits of priority to breeders. The fifteen prioritized relatives of pigeonpea generally occur in South and Southeast Asia to Australia, with the highest concentrations of species in southern India and northern Australia. These taxa differ considerably among themselves and in comparison to the crop in their adaptations to temperature, precipitation and edaphic conditions. We find that these wild genetic resources are broadly under-represented in *ex situ* conservation systems, with 80% of species assessed as high priority for further collecting, thus their availability to plant breeders is insufficient. We identify species and highlight geographic locations for further collecting in order to improve the completeness of pigeonpea CWR germplasm collections, with particular emphasis on potential traits for abiotic stress tolerance.

Keywords: Crop diversity, Crop improvement, Crop wild relatives, Food security, Germplasm conservation, Plant genetic resources

INTRODUCTION

Challenges to global food production are compounding. Our growing population and dietary expectations are projected to increase demand on food systems for at least the next four decades, outpacing current yield trends (Ray *et al.* 2013). Limitations in land, water, and natural resource inputs, competition for arable soils with non-food crops and other land uses, the need to minimize harmful impacts on biodiversity and other ecosystem services, and greater climatic variability further constrain production potential (Lobell *et al.* 2008, Cordell *et al.* 2009, Rosenzweig, *et al.* 2013). Although gains in food availability may partially be obtained through dietary change and waste reduction (Tilman & Clark 2014, West *et al.* 2014), a transition toward more sustainable, yet highly productive, agricultural systems is necessary. This transformation must be achieved through improved agronomic practices combined with the use of varieties of crops with reliable yields under more adverse conditions (Foley *et al.* 2011).

One such crop is pigeonpea [*Cajanus cajan* (L.) Millsp.], a sub-tropical and tropical grain legume that originated in the northern region of the Indian sub-continent, spreading to East Africa at least 4000 years BCE, and then to Southeast Asia, West Africa, Latin America, and the Caribbean. The seed is eaten as a green vegetable and dry pulse and is an important source of protein, vitamin B, carotene, and ascorbic acid (Odeny 2007, Choudhary *et al.* 2013). The pods and foliage of the plant are used as livestock forage and fodder, the crop is cultivated as a green manure, and its woody stem is used as fuel and construction material (Mallikarjuna *et al.* 2011). Pigeonpea is an important income generator, particularly in Tanzania, Malawi, and Myanmar as an export crop to India (Odeny 2007).

Pigeonpea is generally planted by smallholder farmers in low input, rain-fed conditions. The crop is well suited to a wide range of agricultural systems, including intercropping and no-till. Cultivation improves soil fertility through biological nitrogen fixation as well as through the solubilization of soil-bound phosphorus (Mallikarjuna *et al.* 2011, Choudhary *et al.* 2013), increasing the yield of intercropped cereals, other pulses, and vegetables (Saxena 2005, Odeny 2007), and has been shown to enhance the control of *Striga* (Odeny 2007). Pigeonpea is more heat tolerant than the majority of grain legume crops (Figure 1) and is regarded as drought-resistant. These traits are associated with the ability to maintain or regulate transpiration under high temperatures and/or low soil moisture, for example through adjustment of leaf osmotic pressure (Subbarao *et al.* 2000), maintenance of photosynthetic function under stress

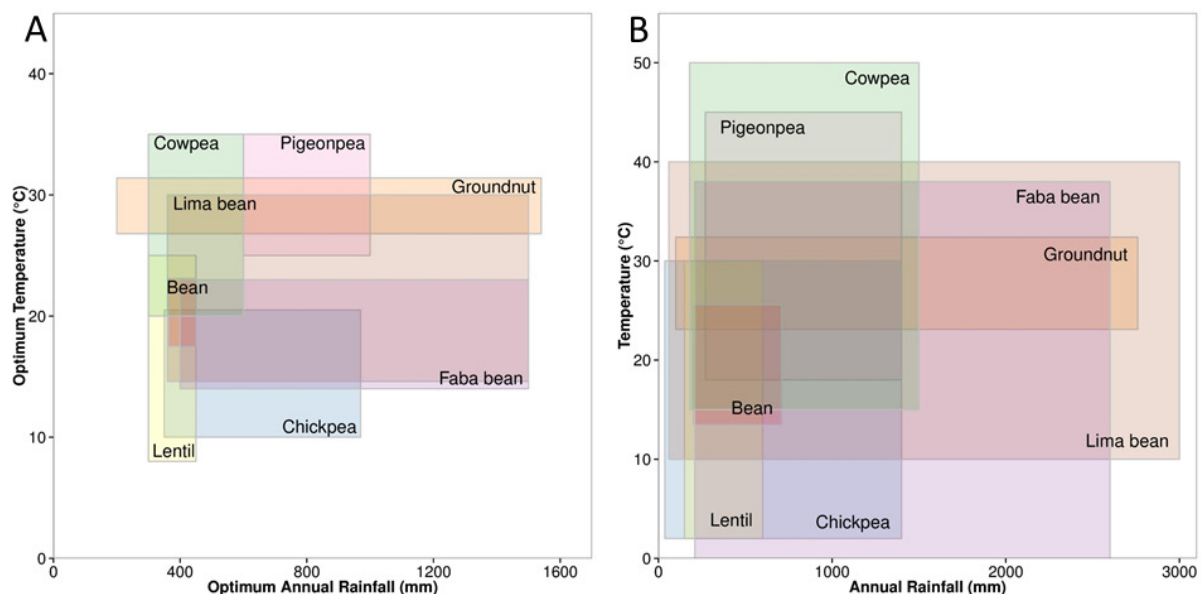


Figure 1. Climatic niches for temperature and precipitation for major grain legumes. (A) optimal range and (B) minimum and maximum observed range (Bogdan 1977, van der Maesen 1989, Valenzuela and Smith 2002, Odeny 2007, FAO 2010a, Sardana *et al.* 2010, Saxena *et al.* 2010).

(Lopez *et al.* 1987), and deep root systems (Flower & Ludlow 1987).

Due to its high nutritive value and agronomic traits, pigeonpea can play an increasing role in low input production systems in India, East Africa, and elsewhere in the dry sub-tropics and tropics (Saxena 2005, Odeny 2007). Concerted breeding efforts for this purpose have resulted in a number of promising advances, particularly the creation of early maturing varieties, and developments toward diverse high yielding hybrids (Saxena 2005, Saxena *et al.* 2013, Saxena & Sawarganokar 2014). However, crop yield in most production regions is well below its potential and has been stagnant for a number of decades, with increased production during this time largely due to an expansion of harvested area (Jones *et al.* 2002, Saxena 2005, Odeny 2007). In order to increase pigeonpea yield and adaptability, current breeding priorities include photoperiod insensitivity, resistance to biotic pressures, and tolerance to abiotic stresses, notably waterlogged and mineral deficient soils, cold and heat stress, salinity, and drought (Mligo & Craufurd 2005, Saxena 2005, Odeny 2007, Upadhyaya *et al.* 2007, Choudhary *et al.* 2011).

The long-term viability of major food crops, particularly in light of the increasing need for sustainable production techniques, is dependent upon the use of diverse genetic resources to maintain productivity and adapt to changing climatic conditions and emerging pest and disease

Table 1. Published confirmed (C) and potential (P) uses of pigeonpea CWR in crop improvement.

Taxon	Trait
<i>C. acutifolius</i>	Cytoplasmic male sterility (C ^a), Pod borer resistance (C ^b), High seed weight (C ^c), Sterility mosaic disease resistance (P ^c), Pod fly resistance (P ^d)
<i>C. albicans</i>	High seed protein (C ^c , P ^e), Pod borer resistance (P ^b), Pod fly resistance (P ^d), Pod wasp resistance (P ^d), Broad pods (P ^e), More seeds per pod (P ^e), Good forage source (P ^e), Sterility mosaic disease resistance (P ^f), Salt tolerance (P ^g)
<i>C. cajanifolius</i>	Nuclear male sterility (C ^a), Cytoplasmic male sterility (C ^h), High seed protein (P ^e)
<i>C. cinereus</i>	More seeds per pod (P ^e)
<i>C. crassus</i>	High seed protein (P ^e), Good forage source (P ^e), Sterility mosaic disease resistance (P ^f)
<i>C. lineatus</i>	Cleistogamy (C ⁱ), Cytoplasmic male sterility (C ^j), Pod fly resistance (P ^d), Sterility mosaic disease resistance (P ^f)
<i>C. mollis</i>	More seeds per pod (P ^e), High seed protein (P ^e), Good forage source (P ^e)
<i>C. platycarpus</i>	<i>Phytophthora</i> blight resistance (C ^{f,k}), Sterility mosaic disease resistance (P ^c), Pod borer resistance (P ^c), Early flowering (P ^c), High seed protein (P ^e), Cyst nematode resistance (P ^f), Salt tolerance (P ^{g,l}), Aluminum toxicity resistance (P ^g), Annuality (P ^m), Photoperiod insensitivity (P ⁿ), High flower and pod setting (P ⁿ)
<i>C. scarabaeoides</i>	Pod borer resistance (C ^{b,f}), Sterility Mosaic Disease Resistance (C ^c), Protein improvement (C ^o), Dwarfism (C ^p), Cytoplasmic male sterility (C ^q), Pod fly resistance (P ^d), Pod wasp resistance (P ^d), Early flowering (P ^{e,r}), Salt tolerance (P ^g), Aluminum toxicity resistance (P ^g), Drought tolerance (P ^r)
<i>C. sericeus</i>	High seed protein (C ^c), Cytoplasmic male sterility (C ^s), Pod borer resistance (P ^b), Pod fly resistance (P ^d), Salt tolerance (P ^g), Sterility mosaic virus resistance (P ^{m,f}), <i>Phytophthora</i> blight resistance (P ^m)

^aMallikarjuna and Saxena 2005, ^bMallikarjuna *et al.* 2007, ^cMallikarjuna *et al.* 2011, ^dSharma *et al.* 2003, ^eUpadhyaya *et al.* 2013b, ^fSaxena 2005, ^gChoudhary *et al.* 2011, ^hSaxena *et al.* 2005, ⁱSaxena *et al.* 1998, ^jSaxena *et al.* 2010, ^kMallikarjuna *et al.* 2005, ^lSubbarao *et al.* 1990, ^mobservation by authors, ⁿMudaraddi *et al.* 2013, ^oReddy *et al.* 1979, ^pReddy 1990, ^qSaxena & Kumar 2003, ^rUpadhyaya 2006, ^sAriyanayagam *et al.* 1995.

pressures (Xiao *et al.* 1996, Guarino & Lobell 2011, McCouch *et al.* 2013). Due to the genetic bottleneck effect caused by domestication and crop improvement, pigeonpea cultivars possess only a small portion of the overall genetic diversity present within the genepool (Kassa *et al.* 2012), which also includes traditional farmer varieties and wild related species (Vincent *et al.* 2013). Crop wild relatives (CWR) of pigeonpea have contributed valuable genetic resources for pest and disease resistance, improved nutritional quality, desirable plant architecture, and breeding efficiency. They are considered to possess superior levels of resistance to diseases such as *Fusarium* wilt and *Phytophthora* blight, insect pests such as pod borer, pod fly, and pod wasp, and tolerance to abiotic stress, in comparison to the cultivated species (Table 1).

Increasing awareness of the extent of habitat destruction, invasive species, and other threats to the habitats of the CWR of major crops has given urgency to efforts to identify important species, determine their distributions, and to ensure their conservation for the long term and thus their availability to plant breeders (Jarvis *et al.* 2008, FAO 2010b, Khoury *et al.* 2010). Genetic resource conservation planning efforts have benefitted from advancements in geographic information systems technologies, which have enabled high resolution species distribution modeling in order to inform collecting priorities (Jarvis *et al.* 2005), recognition of important gaps in *ex situ* collections (Ramírez-Villegas *et al.* 2010), and the identification of populations that may possess particularly valuable traits for crop improvement (Tapia *et al.* 2014).

Given the importance of pigeonpea in low input production systems in regions facing food and nutritional insecurity and the capacity for enhancement of the crop through breeding, the aim of this research is to contribute to ensuring the conservation and availability of a broad range of diversity of CWR genetic resources of potential value to present and future crop improvement objectives. Therefore, we analyzed the comprehensiveness of *ex situ* conservation of pigeonpea CWR through a series of questions: (a) what constitutes a potentially useful wild relative of pigeonpea?, (b) where are these species encountered in the wild?, (c) what is the state of conservation and availability of these species to plant breeders? If suboptimal, what are the highest taxonomic and ecogeographic priorities for further collecting? And finally, (e) what CWR resources possess high potential for contribution of traits of value for crop breeding objectives?

MATERIALS AND METHODS

Identification of target CWR species and occurrence data compilation

We identified potentially useful CWR at the species level based upon a genepool concept (Harlan and de Wet 1971) for pigeonpea, which focused on those wild species capable of hybridization with the crop (i.e., members of the primary or secondary genepools), as these species possess the greatest potential for successful introgression of traits (Vincent *et al.* 2013). Taxa in the tertiary genepool with published evidence of confirmed or potential use in crop improvement (Table 1) were also included.

Occurrence records for pigeonpea CWR were acquired from online biodiversity, herbarium,

and germplasm databases; via communications with herbarium and genebank managers, and other crop researchers; and through direct recording of provenance data during visits to selected herbaria (Supplementary Table 1). Germplasm data were obtained from repositories that provide straightforward access to genetic resources and associated data to the global research community through online information systems. Occurrence data were compiled in a standardized format and taxonomically verified following GRIN Taxonomy for Plants (2012) and The Plant List (2010) as references. Existing coordinates were cross-checked to country and verified as occurring on land (Hijmans *et al.* 1999), and records with locality information but no coordinates were geo-referenced using the Google Maps Geocoder (2013) application programming interface. Occurrence data were mapped, iteratively evaluated for correctness with pigeonpea CWR experts, and subsequently further processed in order to form a final dataset of maximized taxonomic and spatial accuracy.

Species potential distribution modeling

A potential distribution model for each species was calculated using the maximum entropy (Maxent) algorithm (Phillips *et al.* 2006), with a set of ecogeographic variables and unique species presence records as inputs. We chose Maxent due to its wide application in predicting species distributions (Elith *et al.* 2006, Phillips & Dudik, 2008, Costa *et al.* 2010). We performed modeling at a resolution of 2.5 arc- minutes ($\sim 5\text{km} \times 5\text{km}$ cell size at the equator), employing 10,000 background points for model training over the combined distributional range of the pigeonpea CWR. Ecogeographic inputs included altitude and nineteen bioclimatic variables from the WorldClim database (Hijmans *et al.* 2005), and seven major edaphic drivers of plant species distributions with consistent data coverage throughout the range of the pigeonpea CWR species, obtained from ISRIC- World Soil Information (Hengl *et al.* 2014) (Supplementary Table 2). For the edaphic variables we calculated a weighted mean across 0-5, 5-15, 15-30, 30-60, and 60-100 cm soil depth values in order to derive a single data value for 0-100 cm. We then resampled the 1 km resolution data to form 2.5 arc-minutes resolution inputs aligned with the WorldClim datasets.

In order to refine and test the stability of the distribution models for each species, we analyzed Maxent results across three ecogeographic input variations: (a) the full set of nineteen bioclimatic variables (Ramírez-Villegas *et al.* 2010); (b) the bioclimatic variables, altitude, and the additional set of seven edaphic variables, totaling 27 input variables; and (c) a species-

specific derivation of the most important drivers of distribution based upon presence data, further refined by removing highly correlated variables. For the ecogeographic variables in the species-specific method, we utilized a non-linear iterative partial least squares (NIPALS) algorithm to perform a principal-component analysis (PCA), as NIPALS can handle data arrays in which the number of observations is less than the number of input variables, and identified those variables with the greatest contribution (>0.7 or <-0.7) to the first two principal components per species based upon occurrence data points. We then used a variance inflation factor (VIF) to identify the variables with a low degree of collinearity (see Supplementary Table 3 for a list of variables utilized per species). A calibrated area under the ROC curve (cAUC) was obtained to assess the predictive performance of each model (Hijmans 2012). The three modeling methods were evaluated with a correlation coefficient against a null model, and the species-specific variables method showed the least spatial sorting bias among methods (spearman's rho for the 19 variables was 0.53; for 27 variables was 0.56; and for the species-specific method was 0.37), and the differences in median AUC distributions across species for each method were found to be statistically significant ($p = 0.0002$) through a Kruskal-Wallis non-parametric analysis of variance test. Potential distribution models based upon the species-specific variables method were therefore utilized in subsequent analyses.

Potential distribution models were constrained per species by a native range defined at the country level as listed in GRIN (2012) and van der Maesen (1986), and were clipped by measuring the shortest distance between the receiver operating characteristic curve (ROC-curve) and the top-left corner of the plot (Liu *et al.* 2005). We limited the spatial analysis to the native distributions of taxa in order to focus prioritization recommendations on those regions with species with long-term adaptation to specific ecogeographic conditions.

Adapted from Ramírez-Villegas *et al.* (2010), Maxent models were produced using the cross-validation option ($k=5$) and were subjected to a four-fold assessment process including: (a) the 5-fold average area under the ROC curve of test data (ATAUC), (b) the standard deviation of the test AUC of the 5 different folds (STAUC), (c) the proportion of the potential distribution coverage with standard deviation above 0.15 (ASD15), and (d) the cAUC value. Models with ATAUC above 0.7, STAUC below 0.15, ASD15 below 10%, and cAUC exceeding 0.40 were considered accurate and stable. For species where the Maxent model did not pass the cross-validation, potential distributions were mapped with a circular buffer of 50 km (CA50)

surrounding each geo-referenced record (Hijmans *et al.* 2001).

Analysis of current *ex situ* conservation and further collecting needs for CWR

We adapted a gap analysis methodology proposed by Ramírez-Villegas *et al.* (2010), combining three metrics used to assess the urgency of further collecting in order to fill gaps in *ex situ* conservation of CWR. The total sample representation of each species in genebank collections was estimated via a sampling representativeness score (SRS), calculated as the number of germplasm samples (G) divided by the total number of samples (G + herbarium samples (H)) (i.e., all other records aside from available genebank accessions).

The sufficiency of geographic coverage of germplasm collections of each species was estimated through a geographic representativeness score (GRS), calculated as the share of the combined total area of CA50 placed around each existing germplasm collection point compared to the overall potential geographic distribution of the species.

The comprehensiveness of ecological coverage of germplasm collections of each species was estimated through an ecological representativeness score (ERS), calculated by estimating the distinct ecosystem classifications (Olson *et al.* 2001) represented in the CA50 of existing germplasm collection points compared to the diversity of ecosystems in which the overall potential geographic distribution model of the species occurs.

A final priority score (FPS) for further collecting for *ex situ* conservation was assigned to each species by averaging the three gap analysis metrics (SRS, GRS, and ERS). FPS scores were further classified into four categories of urgency for collecting: high priority species (HPS) for taxa whose $0 < FPS \leq 2.5$ or when no germplasm accessions currently exist; medium priority species (MPS) when $2.5 < FPS \leq 5$; low priority species (LPS) when $5 < FPS \leq 7.5$; and ‘no further collecting recommended’ (NFCR) when $7.5 < FPS \leq 10$. We produced collecting priorities maps for all species, displaying the geographic areas that have not yet been collected from within the potential distributions of taxa.

The ecogeographic data preparation, species distribution modeling, and gap analysis were written and performed in R v2.15.1 (R Core Team 2013), utilizing packages *maptools* (Bivand & Lewin-Koh 2014), *rgdal* (Bivand *et al.* 2014), *SDMTools* (van der Wal *et al.* 2014), *raster*

(Hijmans 2014), sp (Pebesma & Bivand 2005, Bivand *et al.* 2013), dismo (Hijmans *et al.* 2013), and plsdepot (Sanchez 2012). Resulting spatial files were mapped in ArcMap v.10 (ESRI 2011). Collecting priorities spatial files were analyzed using the Zonal Statistics tool in ArcMap to list the countries prioritized for further collecting for *ex situ* conservation.

In order to validate and/or expose deficiencies in our findings, we subjected the gap analysis numerical and spatial results to an evaluation performed by four crop experts with experience in the distribution and/or conservation status of CWR of pigeonpea: Muluaalem Kassa, Cereal Research Centre, Agriculture and Agri-Food Canada; Sally Norton, Australian Grains Genebank, Australia; Hari Upadhyaya, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT); and Jos van der Maesen, Naturalis Biodiversity Center, the Netherlands). These experts were first asked to provide an evaluation of the sufficiency of germplasm collections per species based only upon their knowledge of total accessions, and geographic and environmental gaps. Such an assessment [comparable expert priority score (EPS)] was considered directly comparable to the FPS of the gap analysis results. A second evaluation score (contextual EPS) based on the entirety of expert knowledge, including threats to species *in situ* and prioritization by usefulness in crop improvement, was collected in order to provide additional information to collecting prioritization efforts. In both cases, an EPS between 0 and 10, aligned with the gap analysis results prioritization scale, was requested. After these steps, experts were shown the gap analysis results and asked to comment on assessed quantitative results, occurrence data, potential distribution models, and maps of collecting priorities. Following these contributions by experts, input occurrence data were further refined by eliminating clearly incorrect points and adjusting country-level native areas, and the potential distribution modeling and gap analyses were re-run in order to improve the quantitative and spatial outputs. Expert metrics displayed in the results pertain to the final evaluation of improved gap analysis outputs.

A multiple factor analysis (MFA) was used in order to compare the various forms of expert evaluation inputs with the gap analysis results, and an expert evaluation index was created, which estimated the degree of accord between all experts and the gap analysis results for each species, with a scale from 0 (disagreement) to 100 (agreement). Analyses were performed using R package FactoMineR (Husson *et al.* 2009).

Identification of CWR with potential traits of use in breeding for abiotic stress tolerance

We utilized ecogeographic information in combination with species presence data in order to identify populations of species with outstanding adaptations to climatic and/or edaphic conditions of interest to pigeonpea breeding objectives. We assessed the relative importance of the 27 ecogeographic variables (Supplementary Table 2) in explaining the total variation among pigeonpea CWR through a PCA, utilizing all occurrence data points found within the native areas of the species. We created a hierarchical cluster of principal components (HCPC) in order to identify ecogeographic clusters for the species using R package FactoMineR.

Boxplots for each of the 27 ecogeographic variables were created based upon CWR species occurrence data points, displaying the median and variance parameters per species per variable. Comparable ecogeographic variable data for the pigeonpea crop was extracted from area of cultivation maps (Monfreda *et al.* 2008) at a resolution of 5 arc-minutes, with a random sample of 1000 points weighted by harvested area, taken from the major cultivation areas in Asia, Africa, and Latin America. As both the CWR and the crop displayed outlier occurrence points that could potentially contribute to an overinflated ecogeographic niche concept, for further comparative analyses focused on breeding objectives for the crop we restricted the ecogeographic niche per species to the central 90% of variation (i.e., 10% outliers were excluded). Ecogeographic niches for CWR and the pigeonpea crop were mapped in R package ggplot2 (Wickham 2009).

RESULTS

Wild relatives of pigeonpea

The genus *Cajanus* Adans. is composed of 32-34 taxa divided into three clades: Indian, Australian, and *Scarabaeoides* (van der Maesen 1986, Kassa *et al.* 2012). No wild conspecific to the cultivated species exists, and thus there are no wild taxa falling within the primary genepool of pigeonpea as defined by Harlan and de Wet (1971). The secondary genepool is comprised of *Cajanus cajanifolius* (Haines) Maesen, in the Indian clade, the putative progenitor of the crop (Kassa *et al.* 2012), *C. acutifolius* (F.Muell.) Maesen, *C. albicans* (Wight & Arn.) Maesen, *C. cinereus* (F.Muell.) F.Muell., *C. confertiflorus* F.Muell., *C. lanceolatus* (W.Fitzg.) Maesen, *C. latisepalus* (Reynolds & Pedley) Maesen, *C. lineatus* (Wight & Arn.) Maesen, *C. reticulatus* (Dryand) F.Muell., *C. scarabaeoides* (L.) Thouars, *C. sericeus* (Baker) Maesen, and

C. trinervius (DC.) Maesen (Saxena *et al.* 2005, Mallikarjuna *et al.* 2011). Three additional species from the tertiary gene pool [*C. crassus* (King) Maesen, *C. mollis* (Benth.) Maesen, and *C. platycarpus* (Benth.) Maesen] have been the subject of publications of confirmed or potential uses in crop improvement and were therefore also included in the analysis (Table 1, Supplementary Table 3). *Cajanus volubilis* (Blanco) Blanco was recorded in Wanjari *et al.* (1999) as contributing sterility traits, but we believe that the material studied in this reference was actually *C. crassus*, therefore *C. volubilis* was not included in this analysis.

A total of 3171 occurrence records for the fifteen CWR were gathered for use in potential distribution modeling and in the gap analysis, including 377 germplasm accession records sourced from six genebanks, and 2794 herbarium and other occurrence reference records sourced from 17 providers (Supplementary Table 1). Records per species ranged from 15 (*C. mollis*) to 594 (*C. acutifolius*). Of these, 1068 records containing distinct cross-checked coordinates were used to model species potential distributions and to locate the original collecting site of existing germplasm accessions (Supplementary Table 3).

CWR species distributions

Potential distribution models performed in Maxent passed the four-fold cross-validation for eleven out of the fifteen CWR. Models for *C. albicans*, *C. cajanifolius*, *C. mollis* and *C. platycarpus* failed the cross-validation due generally to insufficient and dispersed presence records, and were instead mapped by creating CA50 buffers around their occurrences. Native distributions of pigeonpea CWR occur from South Asia through Southeast Asia into northern Australia, as well as on the eastern coast of Madagascar. Species diversity is richest in southern India and in northern Australia, with up to six modeled potential species distributions overlapping in a single area (Figure 2).

Analysis of current *ex situ* conservation and further collecting needs for CWR

Twelve out of fifteen species were assigned high priority for further collecting due to the average of total samples, geographic, and ecological gaps in their *ex situ* germplasm collections (Figure 3, Supplementary Table 3, Supplementary Figure 1). These high priority species included taxa with narrow distributions (*C. cajanifolius*) as well as those with large ranges (e.g., *C. cinereus*, *C. crassus*, and *C. scarabaeoides*). *Cajanus albicans* was assessed as medium priority due to being modeled as relatively well represented *ex situ* in regard to ecosystem

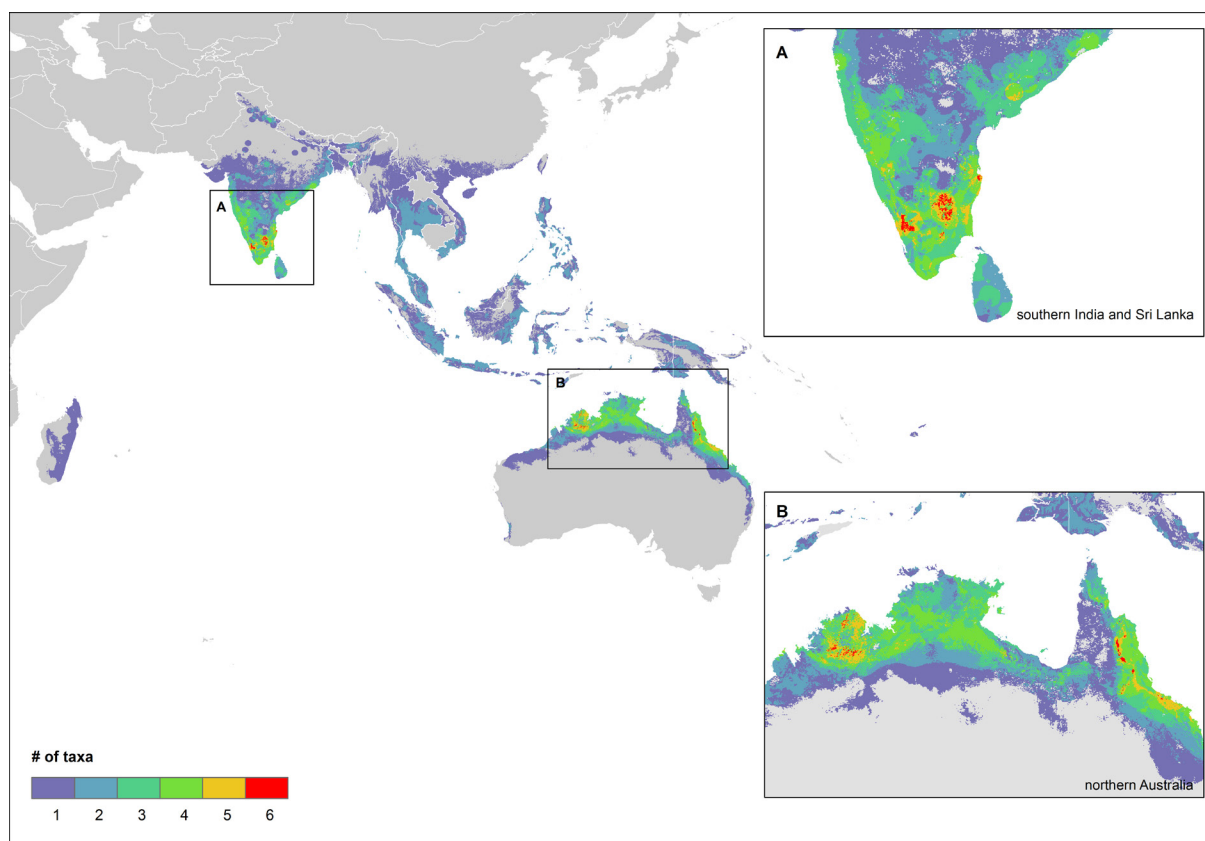


Figure 2. Richness map for assessed pigeonpea CWR potential distribution models, including high species richness areas in (A) southern India and (B) northern Australia.

coverage, and *C. mollis* and *C. platycarpus* as low priority for the same reason, plus high sampling representativeness scores due to having a disproportionately large number of germplasm samples in comparison to herbarium records. The failure of cross-validation of the Maxent models for these species as well as for *C. cajanifolius* resulted in CA50 buffer potential distributions that are likely to be underestimates of the full range of the taxa, especially given the relatively dispersed distributions of available presence records. Thus, the gap analysis assessments for these species likely overestimated the comprehensiveness of their coverage in *ex situ* repositories. The mean final priority score across all CWR was 2.05 ± 1.94 .

Further collecting priorities for the pigeonpea CWR were identified in 20 countries, all of which contained gaps for high priority species (Supplementary Figure 2, Supplementary Table 4). As with species richness, the regions identified for further collecting of the greatest number of species occurred in southern India and in northern Australia (Figure 4). Occurrence data, potential distribution models, and collecting priorities maps for all species are available in an interactive map format at <http://www.cwrdiversity.org/distribution-map/>.

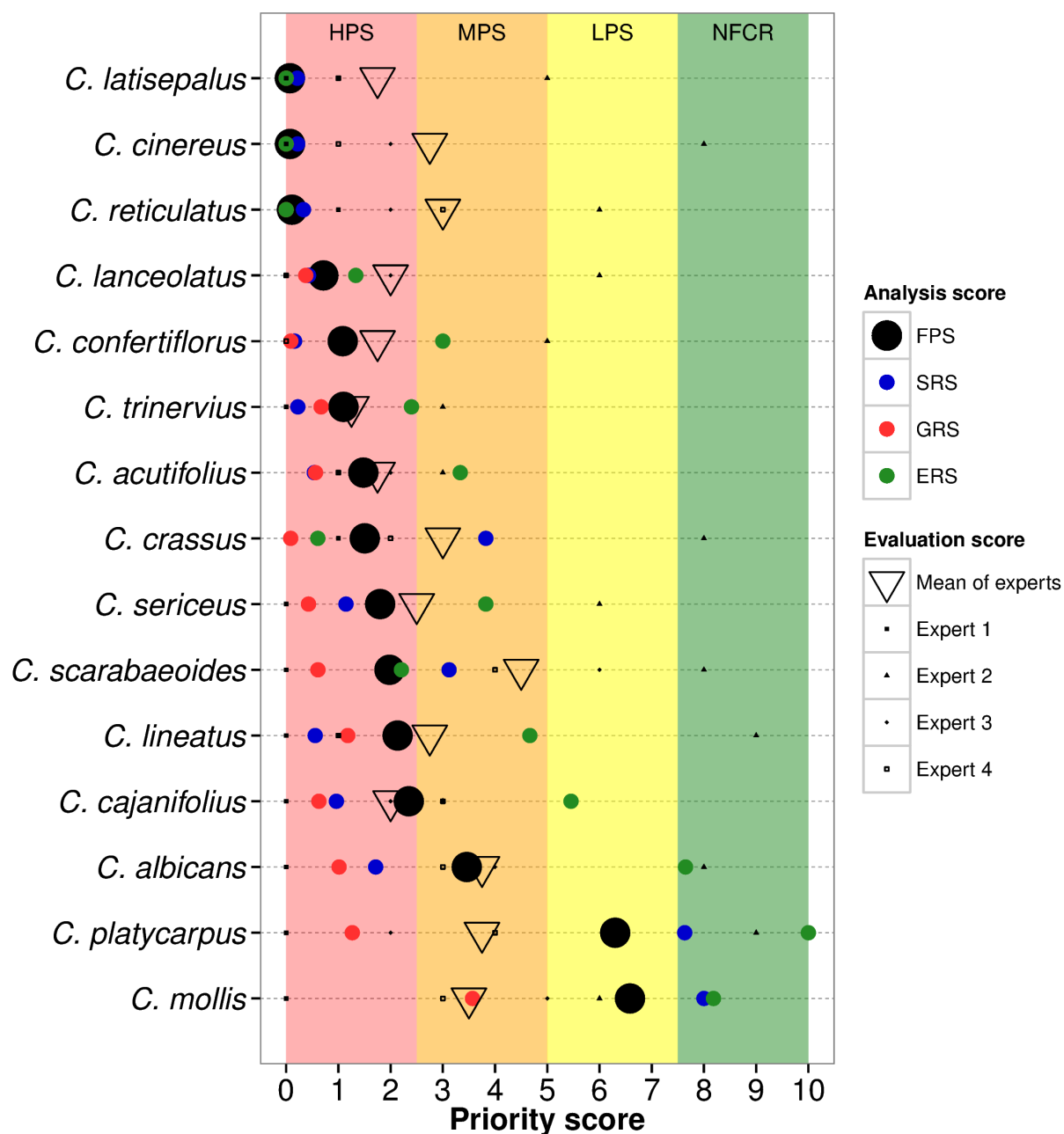


Figure 3. Gap analysis results and comparable expert assessments per species. CWR are listed by descending priority for further collecting by priority categories [high priority species, HPS (red); medium priority species, MPS (orange); low priority species, LPS (yellow); and no further collecting recommended, NFCR (green)]. The black circle represents the final priority score (FPS) for the species, which is the mean of the sampling representativeness score (SRS), geographic representativeness score (GRS), and ecological representativeness score (ERS).

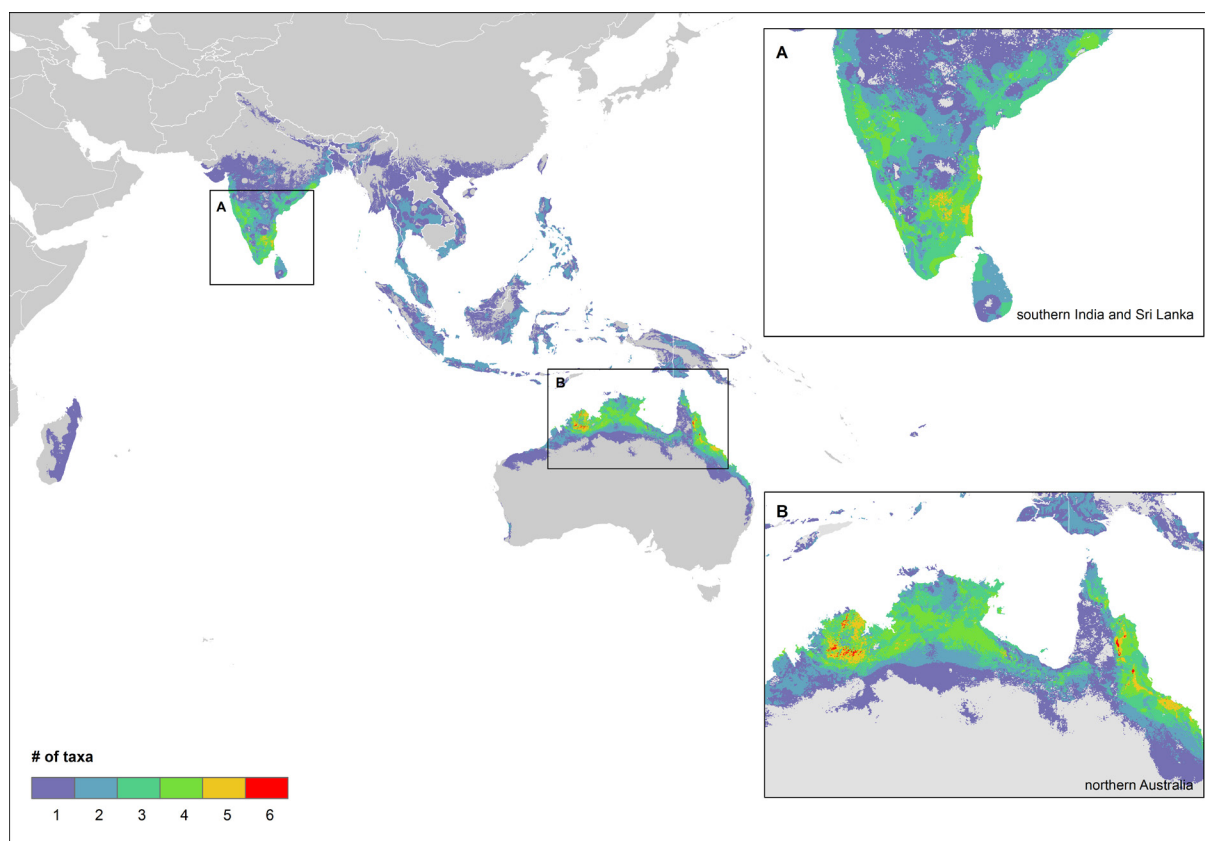


Figure 4. Further collecting priorities hotspots map for high priority (HPS) pigeonpea CWR. The map displays areas within the potential distributions of HPS species that have not been previously collected for *ex situ* conservation, including areas of high species richness in (A) southern India and (B) northern Australia.

The average of the directly comparable expert evaluation priority scores (EPS) across the four experts correlated with the gap analysis results for pigeonpea CWR, with a mean EPS across all experts and all CWR of 2.67 ± 0.9 . The assignment of lower priority for further collecting in the mean score across experts in comparison to the gap analysis, with seven species assigned as HPS and eight as MPS, was highly influenced by one expert's determination of species at a lower priority level than the other three experts. This trend was consistent across species (Figure 3; Supplementary Figure 3A, Supplementary Table 3). The contextual expert priority score per species did not vary widely from the comparable score, with a mean across all experts and species of 2.3 ± 0.89 . The mean contextual score gave slightly higher priority to species for further collecting than did the comparable score, due to knowledge of threats to taxa *in situ* and/or to the generally high value given to pigeonpea CWR in regard to their potential contributions to crop improvement (Supplementary Figure 3B).

The multiple factor analysis revealed relatively strong agreement among the experts and the

quantitative and spatial evaluation variables and thus confidence in the expert evaluation index (Supplementary Figure 3C). Those taxa with the highest accord between all experts and variables and the gap analysis results included Indian species *C. trinervius* and *C. sericeus*, and most of the Australian species (*C. cinereus*, *C. acutifolius*, *C. confertiflorus*, *C. lanceolatus*, and *C. latisepalus*). *Cajanus scarabaeoides* was given a very low index score, and assigned least priority of all species for further collecting by the experts, due to the relatively large total number of *ex situ* germplasm accessions held for the species, whereas the gap analysis assessed the species as high priority due to large geographic and ecological gaps in *ex situ* collections in comparison to the total potential distribution. The taxa with relatively few and dispersed occurrence records and resulting CA50 potential distribution models (*C. albicans*, *C. cajanifolius*, *C. mollis* and *C. platycarpus*) were also among those species receiving the lowest index scores (Supplementary Figure 3D). The evaluations thus served to highlight those species with greatest need of further investigation in regard to their distributions, and at the same time confirmed the robustness of the Maxent models in creating valid depictions of the general range of those CWR with sufficient data availability. Furthermore, the evaluations were useful in identifying erroneous occurrence records for the species, and in highlighting factors contributing to sampling bias in existing collections (e.g., proximity to roads or to research sites), which may affect species distribution modeling.

Identification of CWR with potential traits of use in breeding for abiotic stress tolerance

Strong linear relationships were found between ecogeographic variables within the study area, justifying the application of the PCA, with 70.3% of variance explained through the first three principal components. The first component (37.9% of variation) was generally positively correlated with high and variable temperatures, soil bulk density and pH, and negatively with precipitation and soil organic carbon. The second component (20.8% of variation) was determined by extreme temperature variables. The third component (11.6% of variation), was positively related with precipitation and finer soils (Supplementary Figure 4A; Supplementary Table 2).

Occurrence data were segregated into three ecogeographic clusters. The first cluster, corresponding to lowland areas of Southeast Asia and southern India, was characterized generally by highly variable temperatures and finer soils. The second cluster corresponded to more temperate and/or highland regions in South Asia as well as the eastern coast of Australia,

Table 2. Agronomic traits prioritized in pigeonpea breeding objectives potentially associated with ecogeographic niches of CWR.

Taxon	Gene pool	Gap analysis priority	Mean expert priority	Eco geographic cluster	Potential traits of value for major breeding objectives for abiotic stress tolerance in pigeonpea
<i>C. acutifolius</i>	2	HPS	HPS	3	Heat, drought
<i>C. albicans</i>	2	MPS	MPS	1,3,2	
<i>C. cajanifolius</i>	2	HPS	HPS	3,2	Heat
<i>C. cinereus</i>	2	HPS	MPS	3	Heat, drought
<i>C. confertiflorus</i>	2	HPS	HPS	2,3	Cold
<i>C. crassus</i>	3	HPS	MPS	1,3	Temperature variation/ seasonality, high precipitation, waterlogging
<i>C. lanceolatus</i>	2	HPS	HPS	3	Heat, drought
<i>C. latisepalus</i>	2	HPS	HPS	3	Heat, drought
<i>C. lineatus</i>	2	HPS	MPS	1,3	High precipitation, waterlogging, drought
<i>C. mollis</i>	3	LPS	MPS	2	Cold
<i>C. platycarpus</i>	3	LPS	MPS	3,2	Heat, temperature variation/seasonality, cold
<i>C. reticulatus</i>	2	HPS	MPS	3,2	
<i>C. scarabaeoides</i>	2	HPS	MPS	1,3,2	Heat, temperature variation/ seasonality
<i>C. sericeus</i>	2	HPS	HPS	3,1,2	High precipitation, waterlogging, drought
<i>C. trinervius</i>	2	HPS	HPS	2,1	Cold, high precipitation

Genepool 2 refers to the secondary genepool, and 3 to the tertiary. Priority categories for the CWR of pigeonpea included high (HPS), medium (MPS), and low priority species (LPS) for further collecting for *ex situ* conservation.

and was characterized by dry conditions and colder temperatures. The third cluster, corresponding more generally to India, the Mekong region, and northern Australia, was characterized by low precipitation and low soil carbon (Supplementary Figure 4B). The exploration of germplasm through the lens of ecogeographic clusters may facilitate the identification of populations of species with valuable traits, in this case for temperature stress resistance and waterlogging tolerance; cold tolerance; and tolerance to drought and low soil fertility, respectively. The great majority of presence records of Australian species *C. acutifolius*, *C. cinereus*, *C. latisepalus*, and *C. lanceolatus*, fell within one cluster, while the South and Southeast Asian species generally contained populations falling into two or three clusters (Table 2, Supplementary Figure 4C).

The assessment of climatic and edaphic niches of CWR species based upon occurrence data points revealed large differences in adaptation to temperature, precipitation, and soil characteristics variables (Supplementary Figure 5), including populations of species tolerant of

low and high temperatures, as well as very low and very high rainfall. Such variables for many species fell well outside the ecogeographic niche of the cultivated taxon, particularly for low temperatures and high precipitation (Figure 5A).

For adaptation to high temperatures, populations of northern Australian species such as *C. latisepalus*, *C. cinereus*, *C. acutifolius*, and *C. lanceolatus* stood out, as did *C. platycarpus* and *C. cajanifolius* (Table 2, Supplementary Figure 5B,F,K). Taxa with large spreads in their adaptation to temperature generally were those more relatively widespread species such as *C. scarabaeoides*, *C. crassus*, and potentially *C. platycarpus*. Species with populations of notable adaptation to low temperatures included *C. mollis*, *C. trinervius*, *C. confertiflorus*, and again *C. platycarpus* (Supplementary Figure 5B,G,L). Scant occurrence information was available for a number of these species and further exploration is needed to determine the full range of the taxa, which may result in the identification of populations with even greater tolerance to extreme temperatures, e.g., from populations at higher elevations.

Pigeonpea CWR occurring in areas of notably high annual and/or seasonal precipitation included the central and southern Indian species *C. lineatus*, *C. sericeus*, and *C. trinervius*, as well as *C. crassus* (Table 2, Supplementary Figure 5M,N,Q). As populations of most of these species are adapted to soils with relatively high clay content, these may also represent candidates for traits for waterlogging tolerance (Figure 5B, Supplementary Figure 5W).

Populations of CWR encountered in regions of very low annual and/or seasonal precipitation included those from Australian species *C. latisepalus*, *C. cinereus*, *C. acutifolius*, and *C. lanceolatus*, as well as *C. sericeus* and *C. lineatus*, among others (Table 2, Supplementary Figure 5M,O,R). The pigeonpea crop was also modeled as being tolerant to very low rainfall regions. Despite such tolerance in pigeonpea, we identified eleven CWR species with distributions occurring in regions with less annual precipitation than the driest areas modeled within the middle 90% of occurrence data inputs in regard to the area of cultivation of the crop (i.e., <500 mm). Maps of potential distributions for a selected number of these CWR that are not currently represented in *ex situ* genebanks are displayed in Figure 6, and the differentiation of the occurrence data of these species in the PCA is shown in Supplementary Figure 4D.

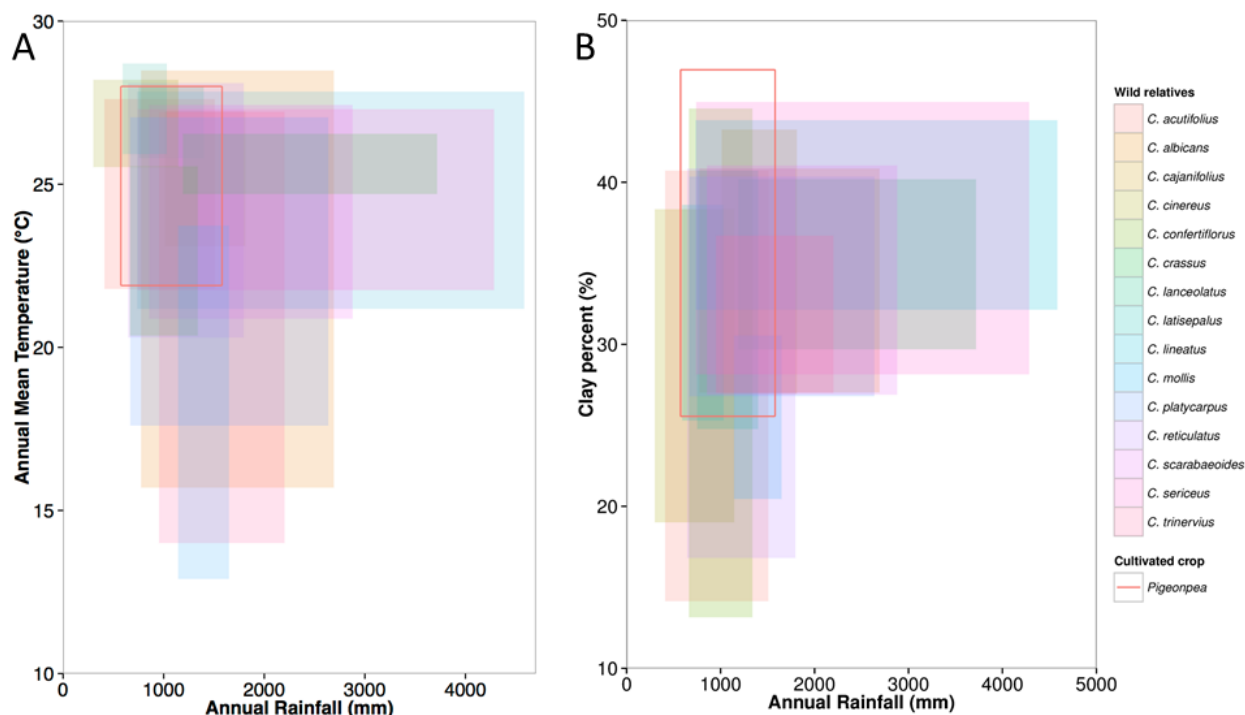


Figure 5. Ecogeographic niches of pigeonpea CWR for (A) annual mean temperature and precipitation and (B) percent clay and annual precipitation. Niches per species represent the middle 90% of occurrence points, i.e., 10% outliers were excluded. For niches per ecogeographic variable per species, see Supplementary Figure 5.

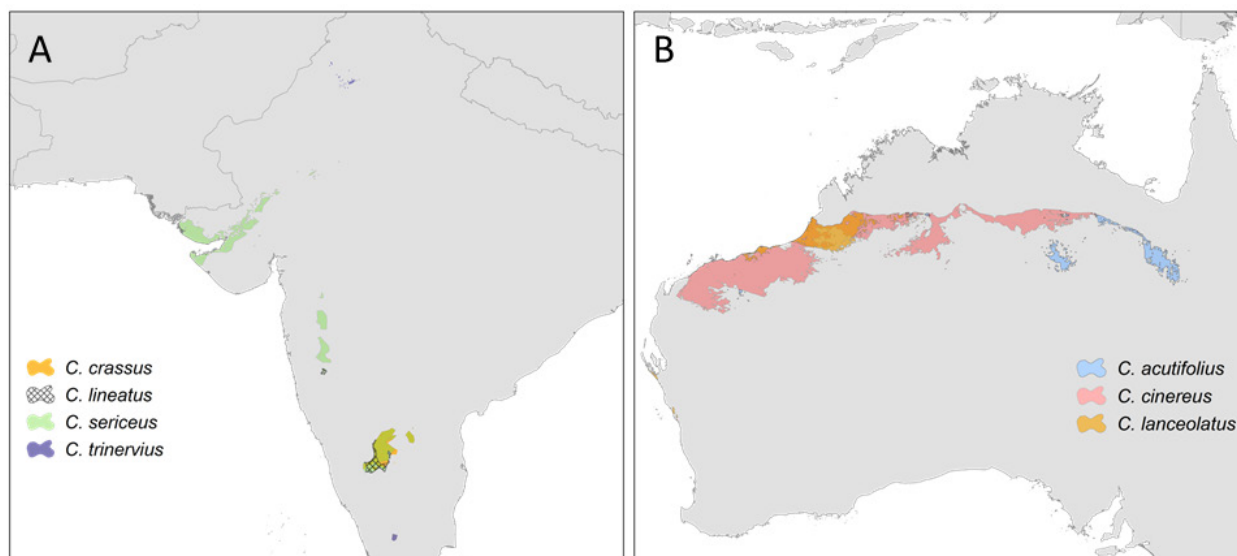


Figure 6. Potential distributions of selected CWR in (A) south Asia and (B) Australia that are not currently represented in germplasm collections and which occur in regions with <500 mm annual precipitation.

DISCUSSION

With 80% of the CWR of pigeonpea in this study assessed as high priority for further collecting for *ex situ* conservation, agreement from expert evaluators of medium to high importance for all species, and with significant geographic gaps in *ex situ* collections for virtually all species, it is clear that further conservation action is needed to safeguard the wild genetic resources of this important grain legume. Included in this list of priorities are species with very few germplasm accessions accessible to the global community in genebank information systems, including *C. confertiflorus*, *C. trinervius*, *C. latisepalus*, and the putative progenitor *C. cajanifolius*. Such taxa represent the highest level of priority for further collecting to fill gaps in germplasm collections at the species level. As the species diversity gaps in genebank collections largely align with the geographic distribution of species richness of pigeonpea CWR, hotspots in India and in northern Australia represent particularly high value regions for efficient collecting of the taxa (Figure 4). These areas may also be of interest for encountering genetic variation created through hybridization between CWR species.

Additional unrepresented distributions of high priority species such as *C. crassus* and *C. scarabaeoides* occur outside these regions, thus targeted collecting throughout the geographic distributions of the species is necessary in order to form germplasm collections that are comprehensive at the population level. Non-native distributions of widespread species, particularly *C. scarabaeoides*, may also be considered for further collecting in the search for useful traits for crop improvement. As techniques for the efficient utilization of wide diversity of plant genetic resources improve, the collection, conservation, and availability of more distant relatives of the crop may also become more worthwhile. We assessed the representation of the other 17-19 species within genus *Cajanus* in *ex situ* conservation as currently minimal, with only 41 accessions from seven species listed in germplasm repositories.

Cajanus scarabaeoides and *C. platycarpus* were identified as exhibiting potential adaptation to climatic extremes, and are represented by some diversity of accessions conserved *ex situ*. Existing collections should therefore be further screened for adaptive traits. Other species identified as having potentially useful adaptations in contribution to major abiotic stress tolerance breeding objectives for pigeonpea are represented by very few germplasm accessions, especially *C. confertiflorus*, *C. trinervius*, *C. cajanifolius*, *C. latisepalus*, *C. lanceolatus*, and *C. cinereus*. The climatic extremes of potential distribution models, such as those displayed in

Figure 6, may represent particular areas of interest for exploration. As Maxent models are based upon known presence points for species and are thus subject to sampling bias, they may not fully capture the possibility of occurrence of populations of CWR species in unique climates (Araújo & Guisan 2006, Loiselle *et al.* 2008, Kramer-Schadt *et al.* 2013). Further field exploration of climatic extremes beyond the edges of the distributions created through these methods may therefore lead to the discovery of new populations with particularly valuable adaptations to abiotic stress (Williams *et al.* 2009).

Despite sizable existing germplasm collections, species such as *C. scarabaeoides* and *C. albicans* were categorized as medium or high priority for further collecting due to geographic and ecological gaps in the collections. As the cost of conserving and investigating germplasm *ex situ* is significant, particularly for CWR, a further assessment informed by genotypic diversity analyses of what constitutes sufficient germplasm collections for pigeonpea CWR is warranted. Given adequate resources, further collecting should be considered for these species, as extremely valuable traits sourced from CWR of native Southeast Asian crops such as rice have been found in only a few populations despite screening of a large number of accessions (Brar & Khush 1997), and accessions of individual CWR species of pigeonpea have been shown to possess notable variation in traits such as resistance to insect pests (Sharma *et al.* 2003), seed protein content, and days to maturity (Upadhyaya *et al.* 2013a).

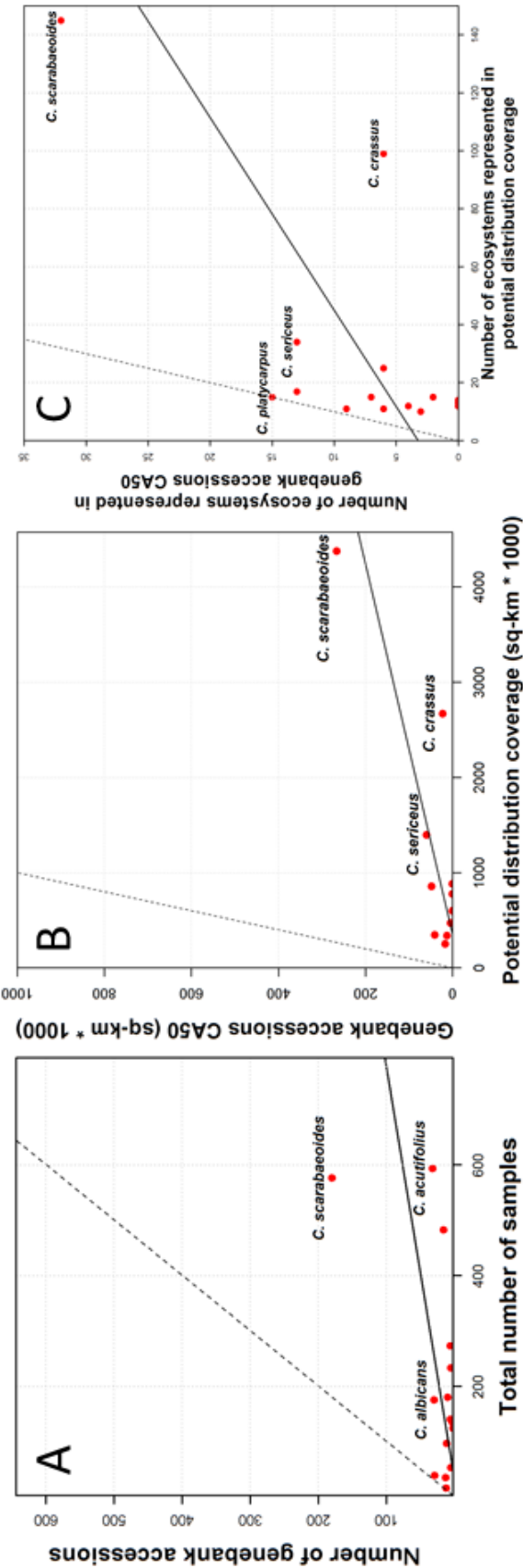
The regions of distribution of pigeonpea CWR species occur in areas undergoing habitat change due to conversion to agriculture, logging, urbanization, mining, invasive species, and climate change, among other factors (Sodhi *et al.* 2004, Sodhi *et al.* 2009, Upadhyaya *et al.* 2013a, Sahai *et al.* 2014). It is clear that the window of opportunity for comprehensive collecting for *ex situ* conservation of pigeonpea CWR will not exist indefinitely.

CONCLUSION

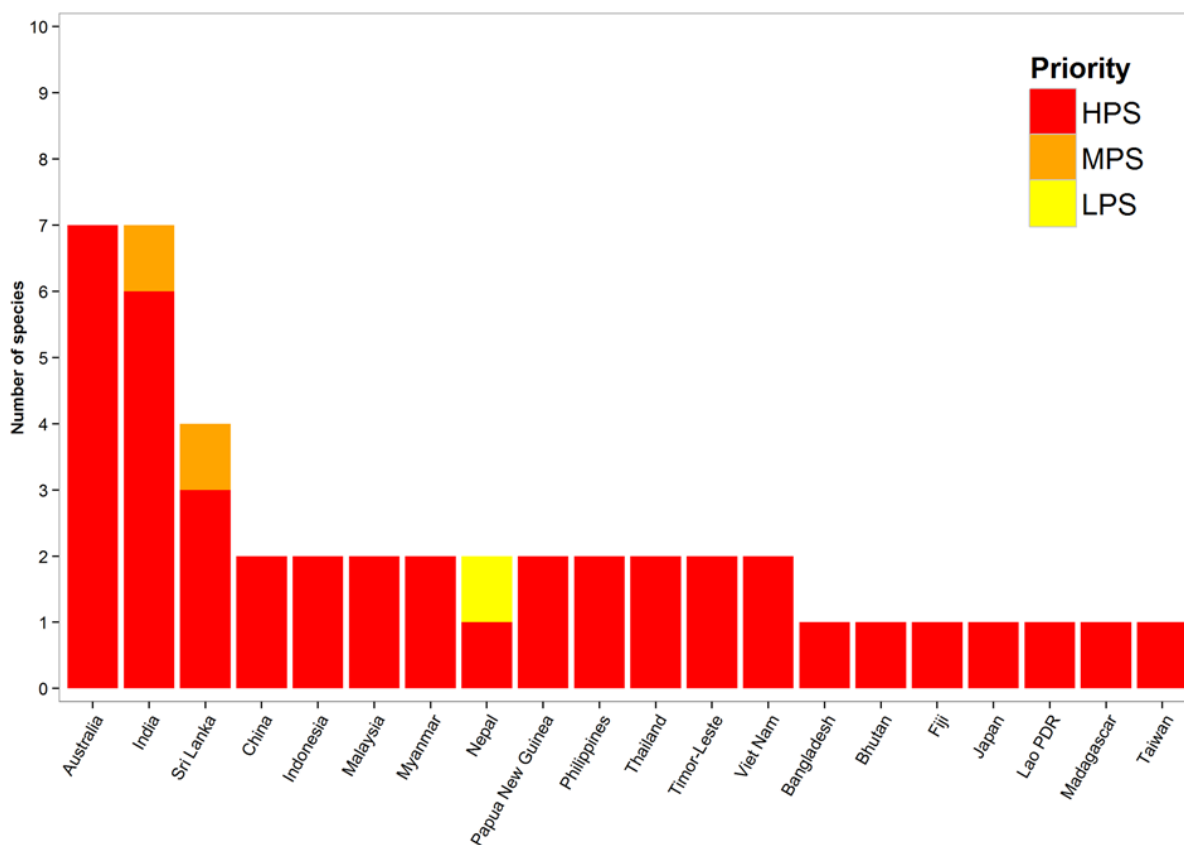
Pigeonpea cultivation is still largely limited to its origins in South Asia and in East Africa. Due to its high nutritive value, agronomic versatility, stress-tolerance, and multiple uses, increasing yield in existing production lands as well as expanding the crop into other areas of Asia and Africa, as well as the Americas, can contribute toward greater agricultural sustainability and improved human nutrition in sub-tropical and tropical regions. While pigeonpea already possesses very favorable agronomic characteristics compared to other major grain legumes, its

productivity can be improved via breeding, and its wild relatives show promise in providing critical adaptive traits for major breeding objectives, including heat and cold tolerance, high precipitation, waterlogging, and drought tolerance. Further collecting for *ex situ* conservation of this diversity, securing long-term funding for this conservation and associated research, ensuring safety-duplication of unique germplasm, and sharing of this diversity with the global research community are critical to this process (FAO 2002, Esquinas-Alcázar 2005). Greater investment in genotypic and phenotypic characterization and evaluation for traits of interest (Mallikarjuna *et al.* 2011, Varshney *et al.* 2011, Upadhyaya *et al.* 2013a) and in breeding programs using CWR, represent equally urgent steps (Tester and Langridge 2010, Guarino & Lobell 2011, Henry 2014). Through such actions the crop research community will contribute to ensuring the long term viability of this important crop, and be better prepared to adapt to the challenges facing present and future grain legume production.

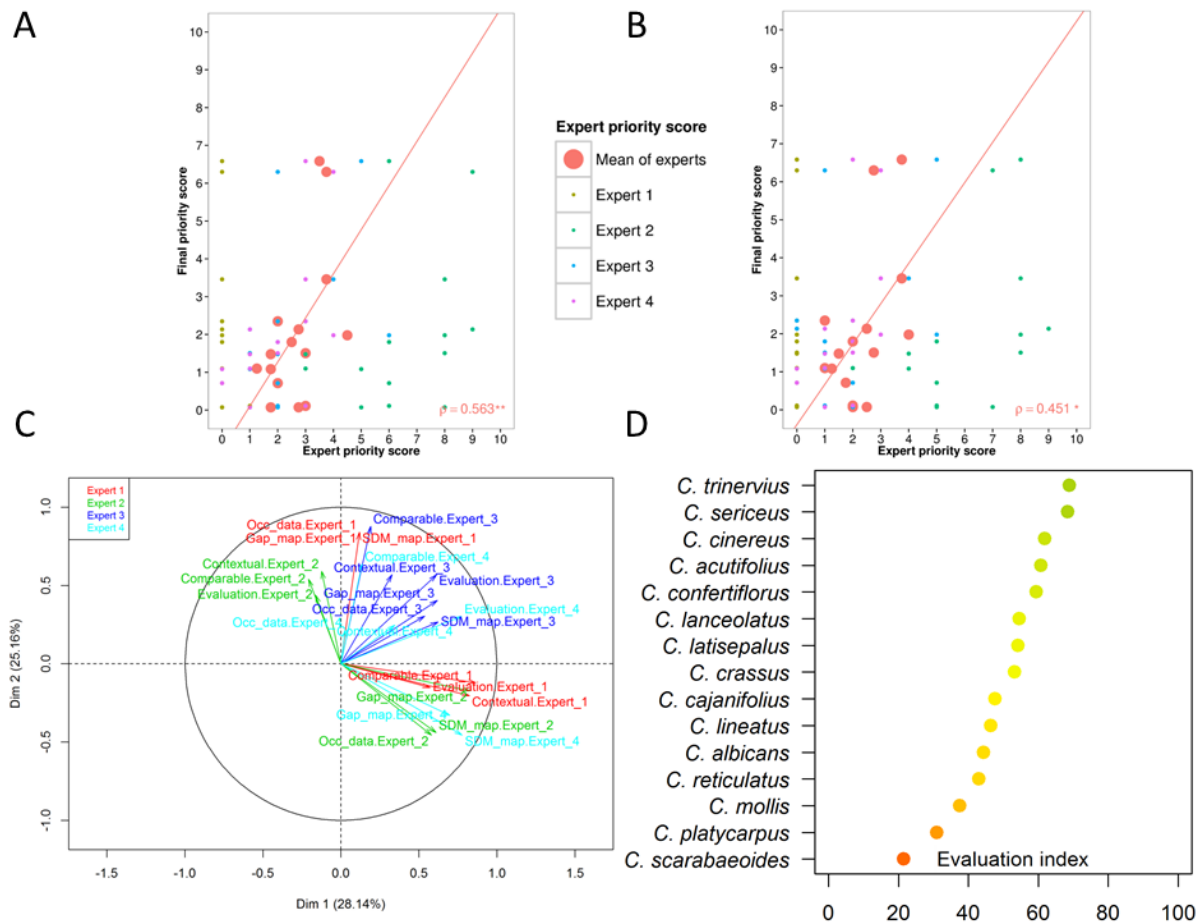
SUPPLEMENTARY INFORMATION



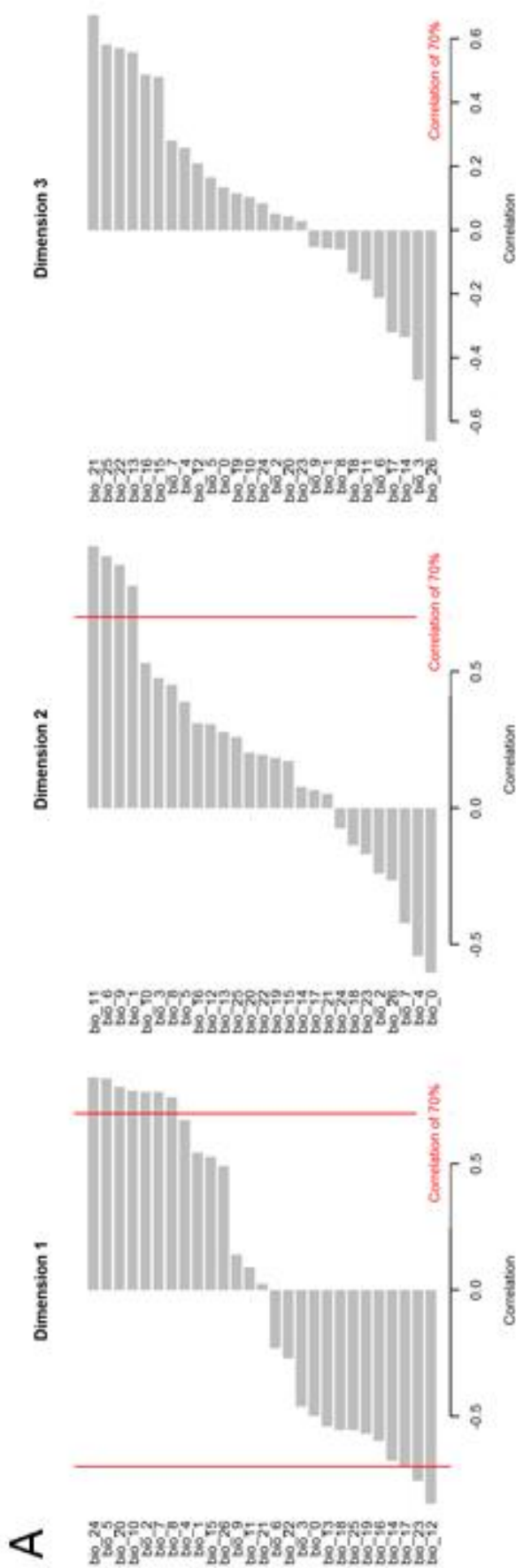
Supplementary Figure 1. Distribution of gap analysis values per CWR species for (A) sampling representativeness score (SRS), (B) geographic representativeness score (GRS), and (C) ecological representativeness score (ERS).



Supplementary Figure 2. Number of CWR species prioritized for further collecting per country. HPS = high priority species, MPS = medium priority species, and LPS = low priority species for further collecting.



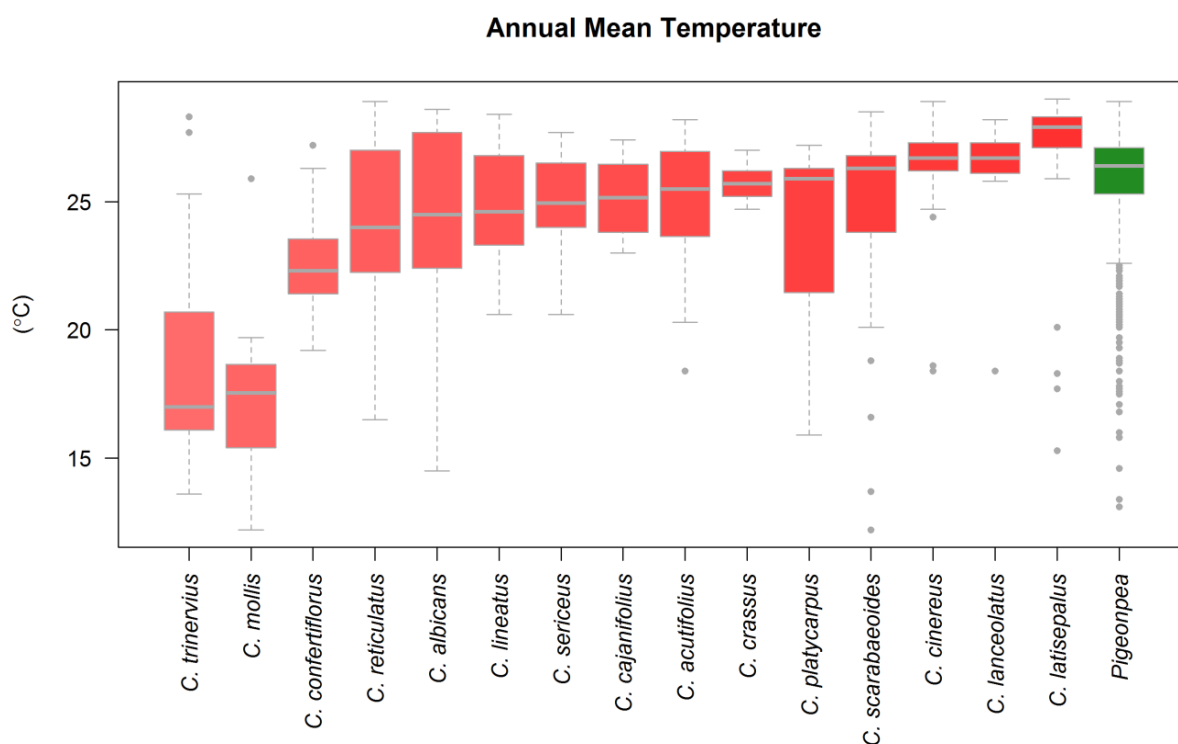
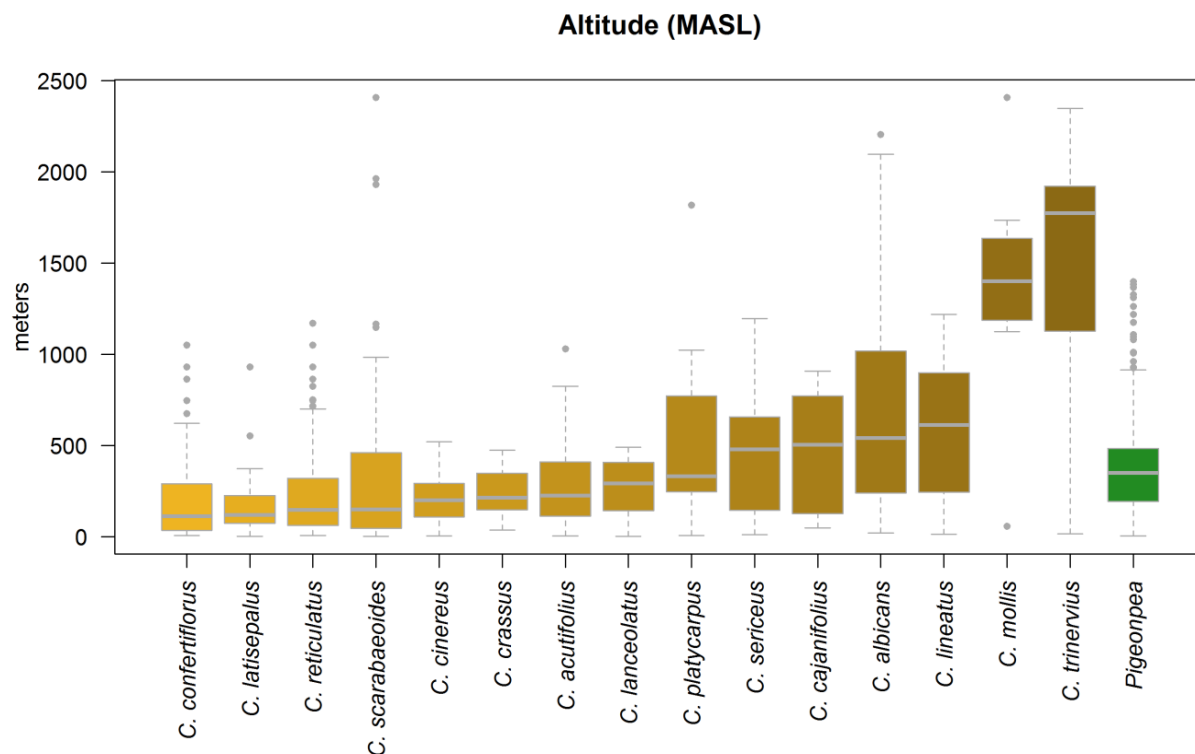
Supplementary Figure 3. Expert evaluation accordance with gap analysis results. (A) correlation between gap analysis results and comparable expert evaluation scores. (B) correlation between gap analysis results and contextual expert evaluation scores. (C) correlation circle of all evaluation variables [comparable expert priority score (Comparable); contextual expert priority score (Contextual); evaluation of gap analysis results score (Evaluation); evaluation of occurrence data (Occ_data); evaluation of potential species distribution models (SDM_map); and evaluation of collecting priorities map (Gap_map)]. (D) combined expert evaluation index score per species.

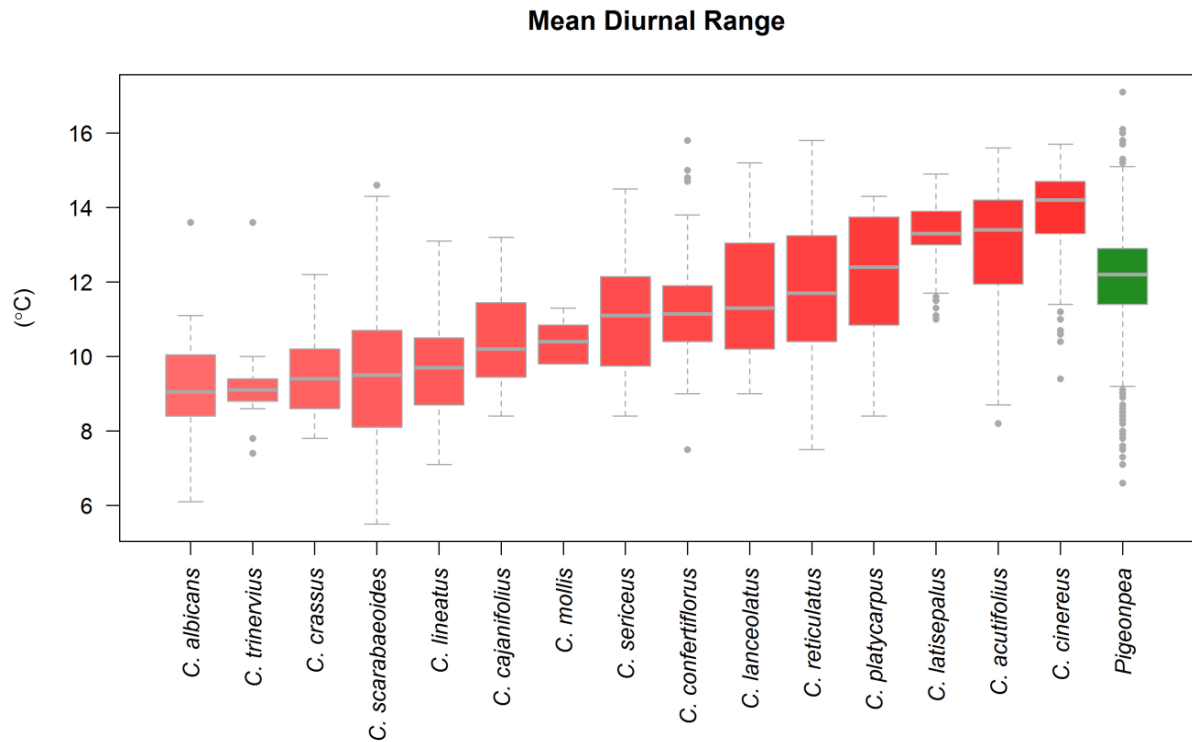


Supplementary Figure 4. Principal component analysis of ecogeographic variables associated with occurrence data for crop wild relative species.
(A) contribution of variables to the first three components.

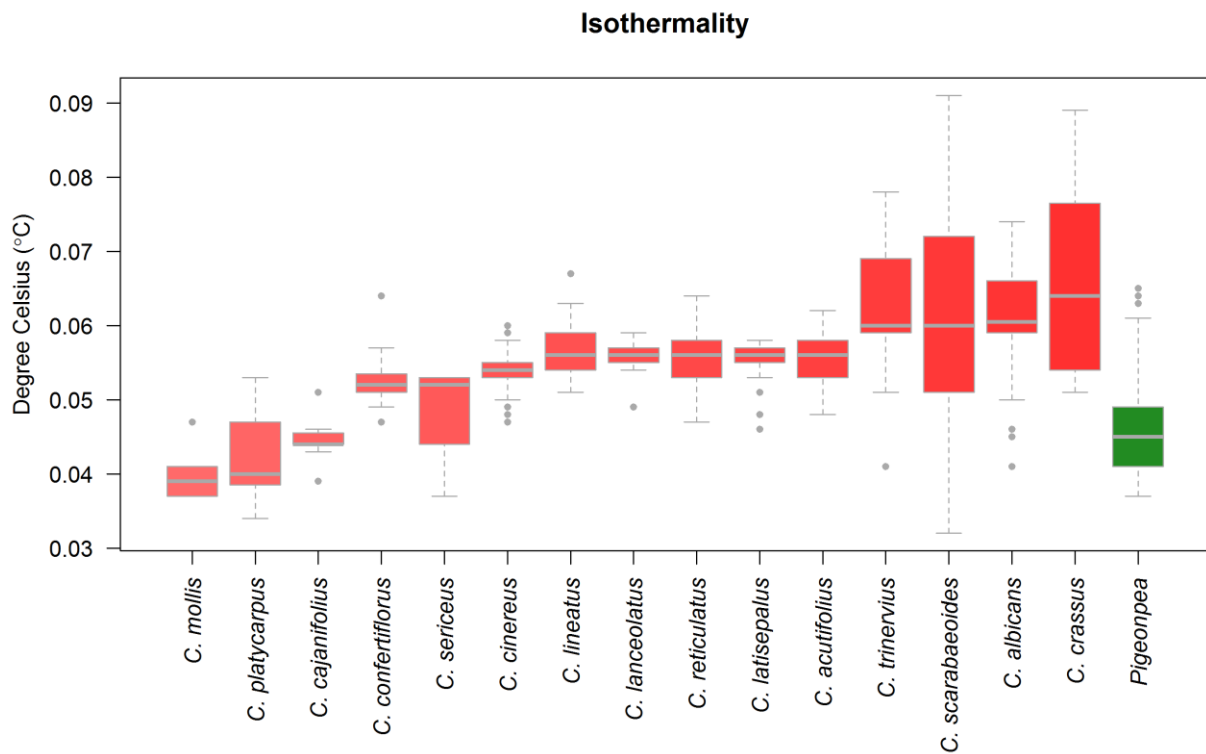
213

Supplementary Figure 5. Ecogeographic niches of crop wild relative (CWR) species based upon their occurrence data presence locations, and the pigeonpea crop, per bioclimatic and edaphic variable. The dark line represents median values, boxplots between 25-75% variation, and open circles outliers.

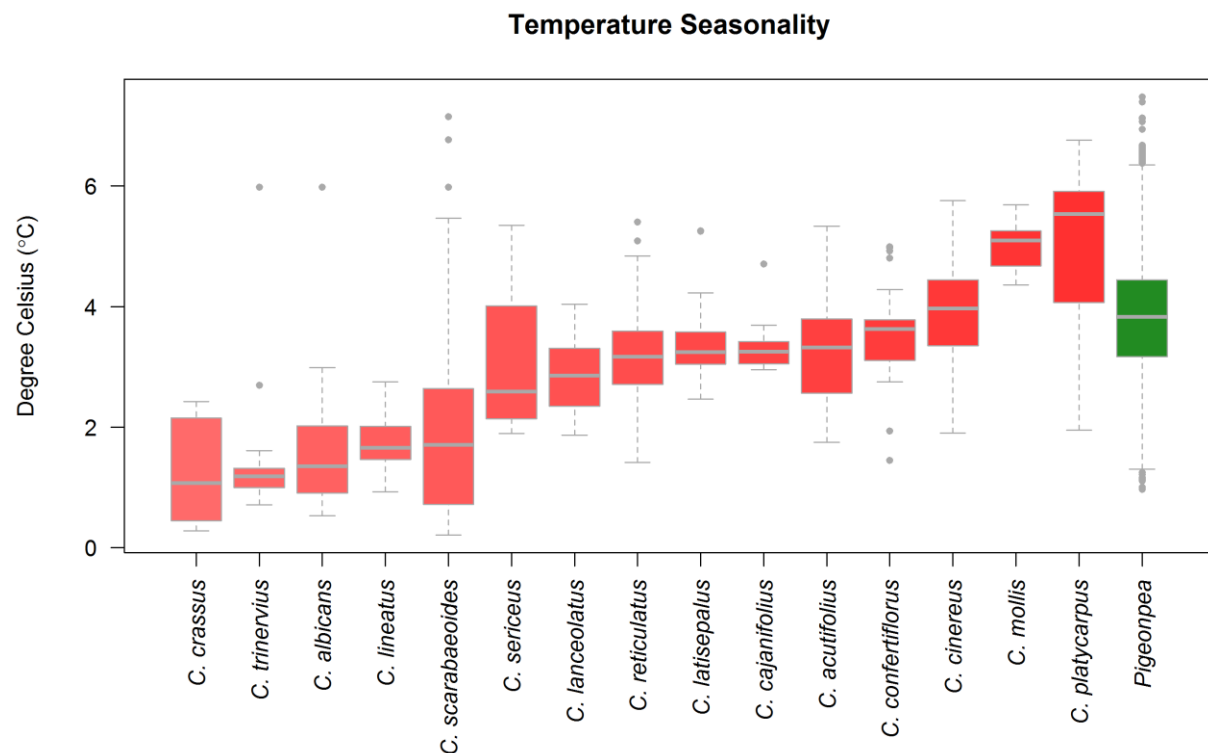




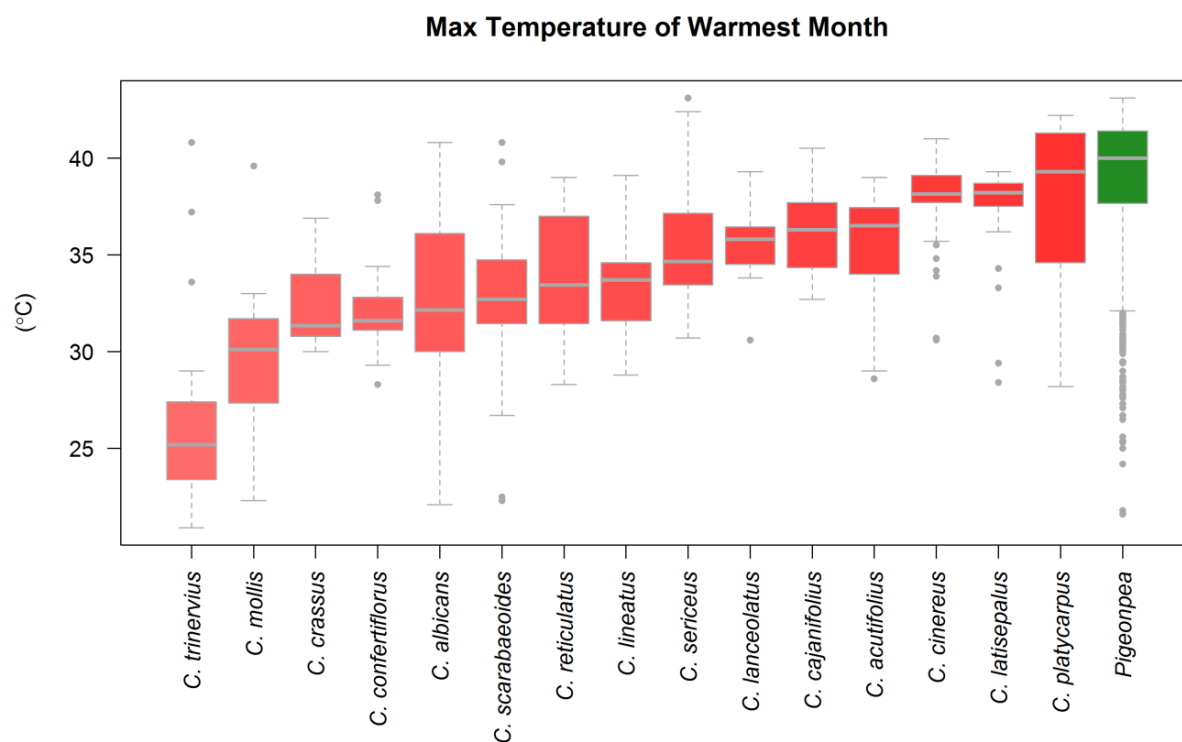
Supplementary Figure 5C. Climatic niches of CWR species and the pigeonpea crop for mean diurnal range.



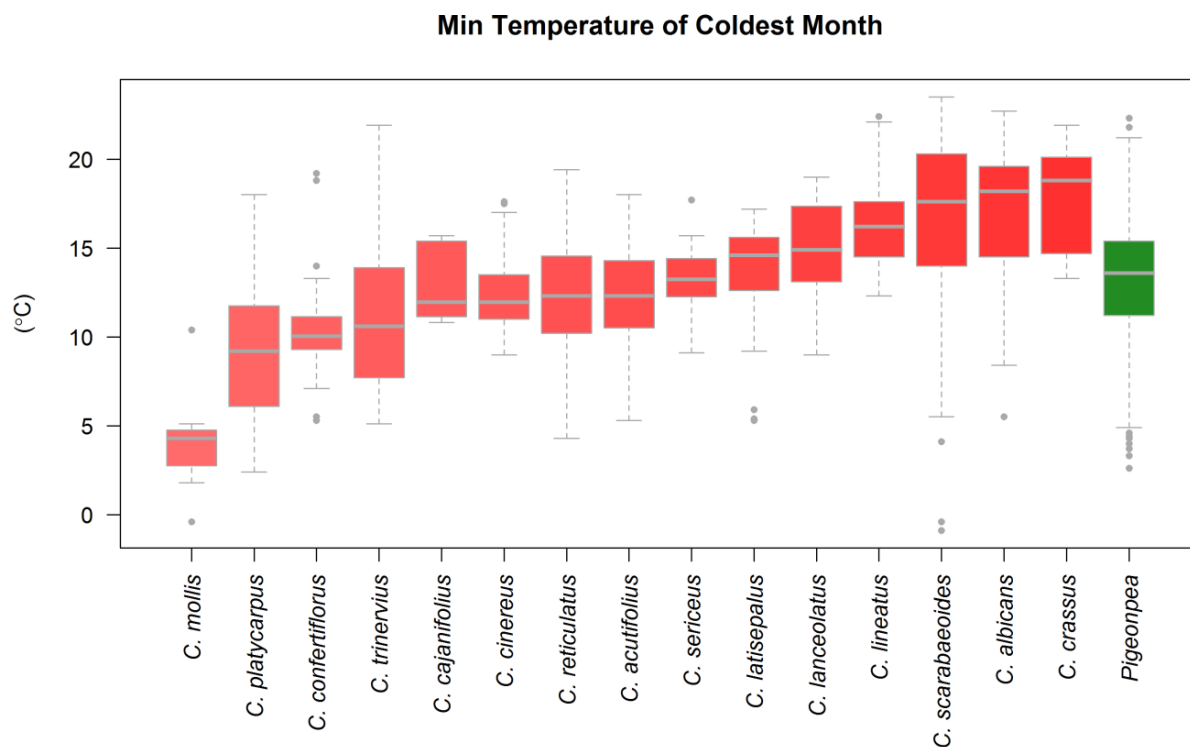
Supplementary Figure 5D. Climatic niches of CWR species and the pigeonpea crop for isothermality.



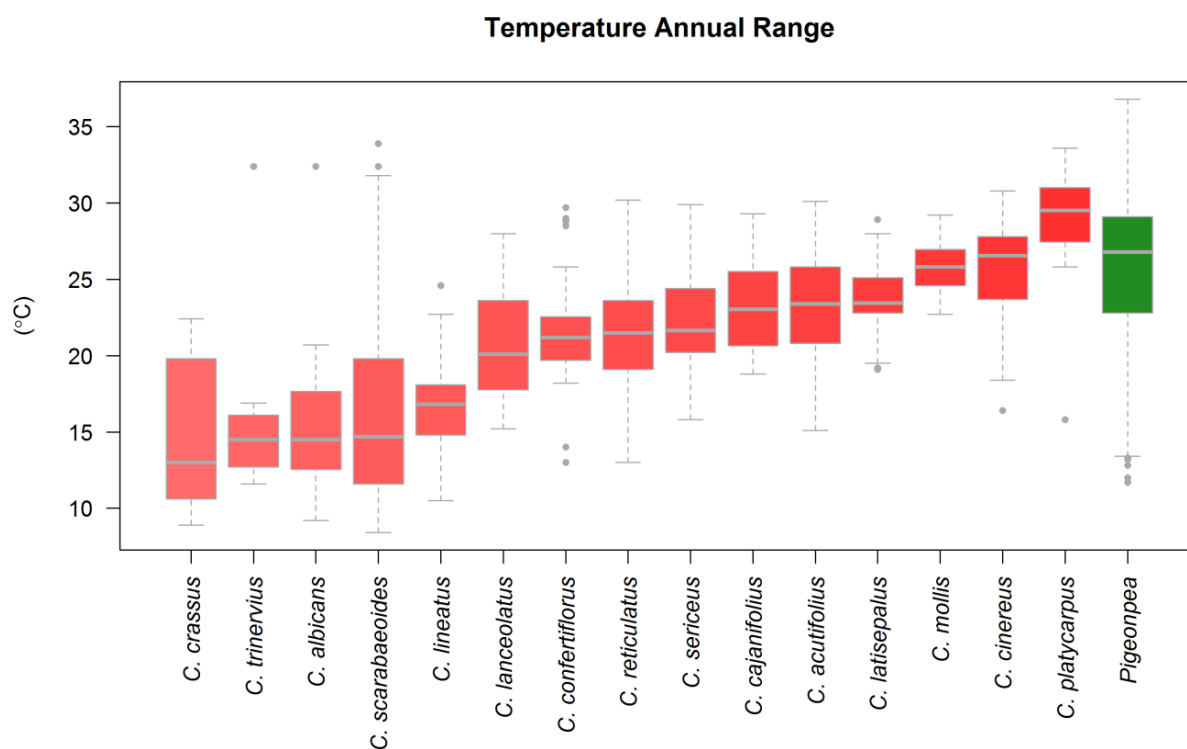
Supplementary Figure 5E. Climatic niches of CWR species and the pigeonpea crop for temperature seasonality.



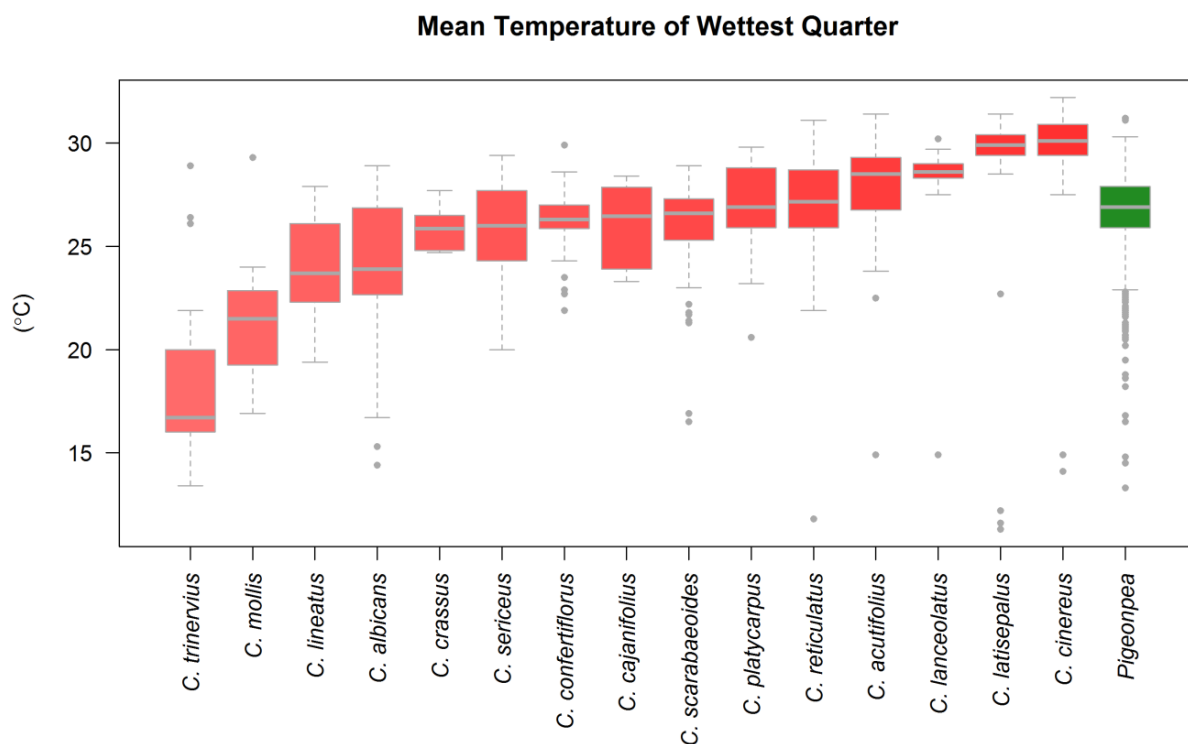
Supplementary Figure 5F. Climatic niches of CWR species and the pigeonpea crop for maximum temperature of the warmest month of the year.



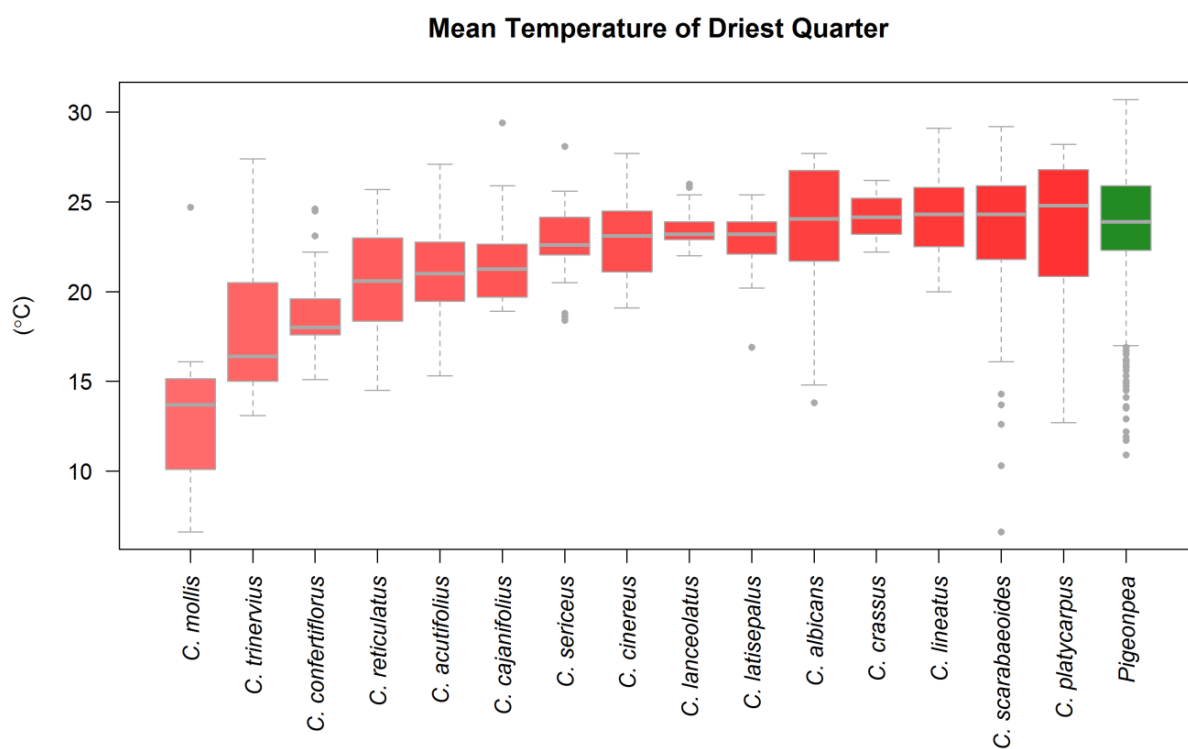
Supplementary Figure 5G. Climatic niches of CWR species and the pigeonpea crop for minimum temperature of the coldest month of the year.



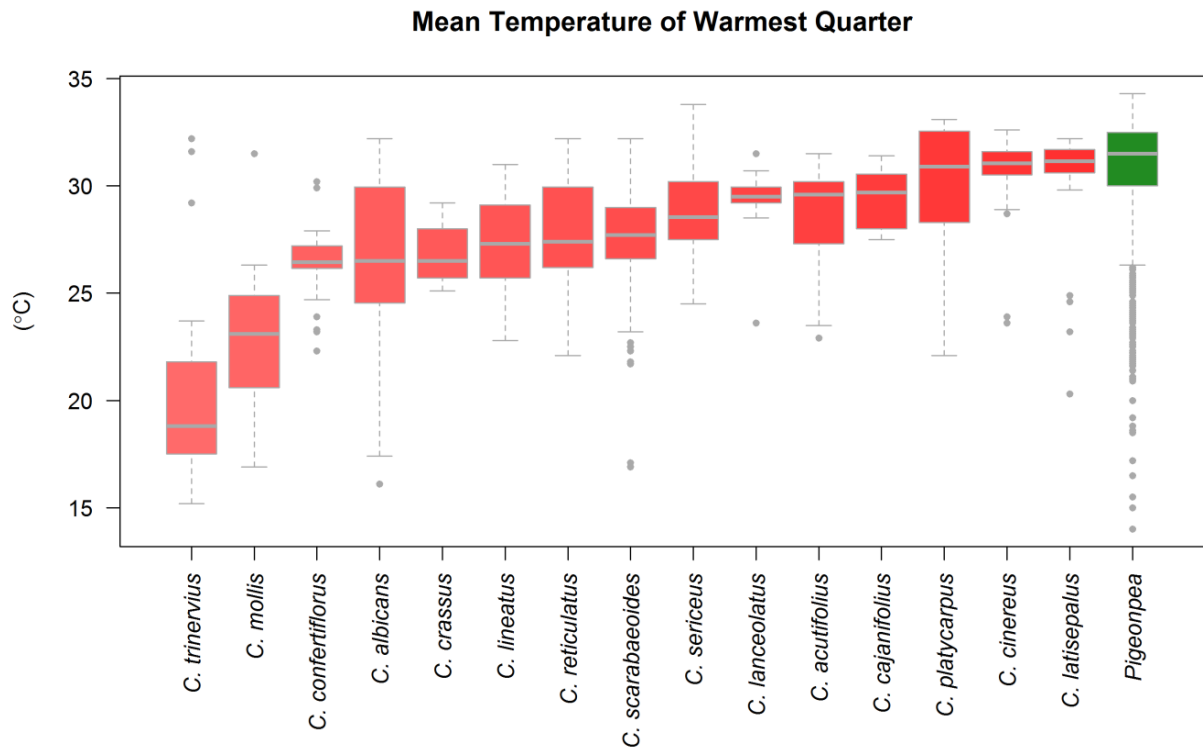
Supplementary Figure 5H. Climatic niches of CWR species and the pigeonpea crop for temperature annual range.



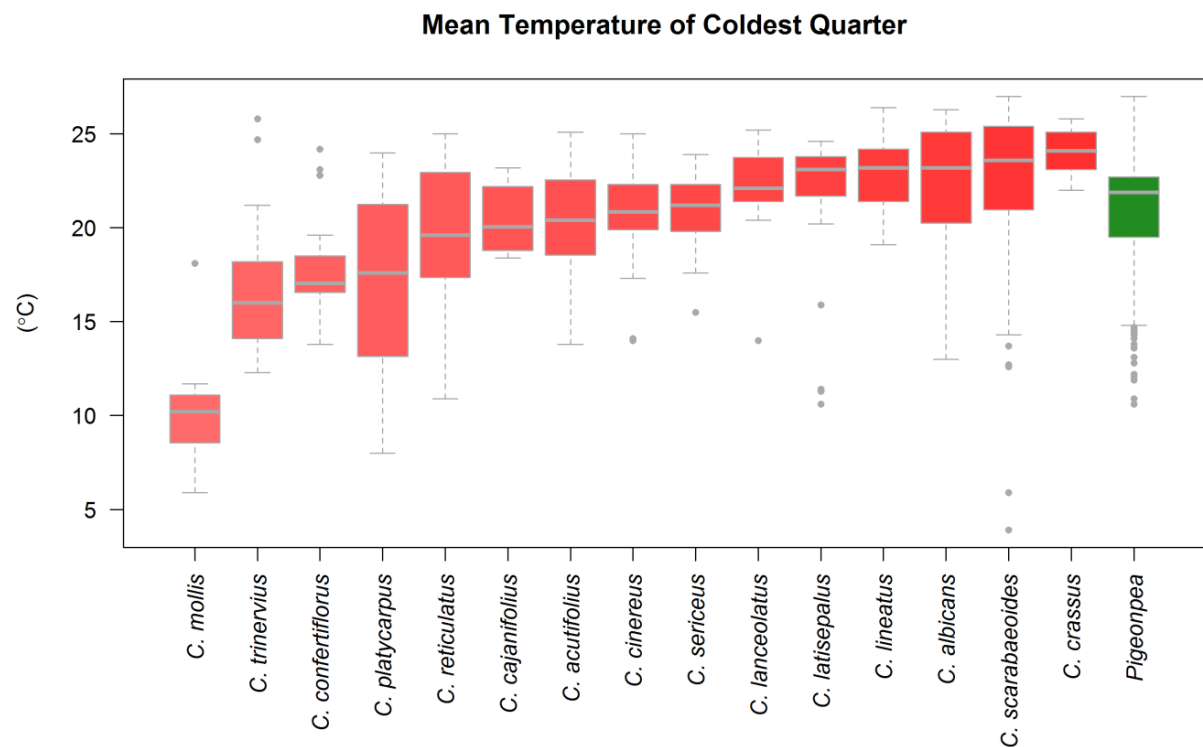
Supplementary Figure 5I. Climatic niches of CWR species and the pigeonpea crop for mean temperature of the wettest quarter of the year.



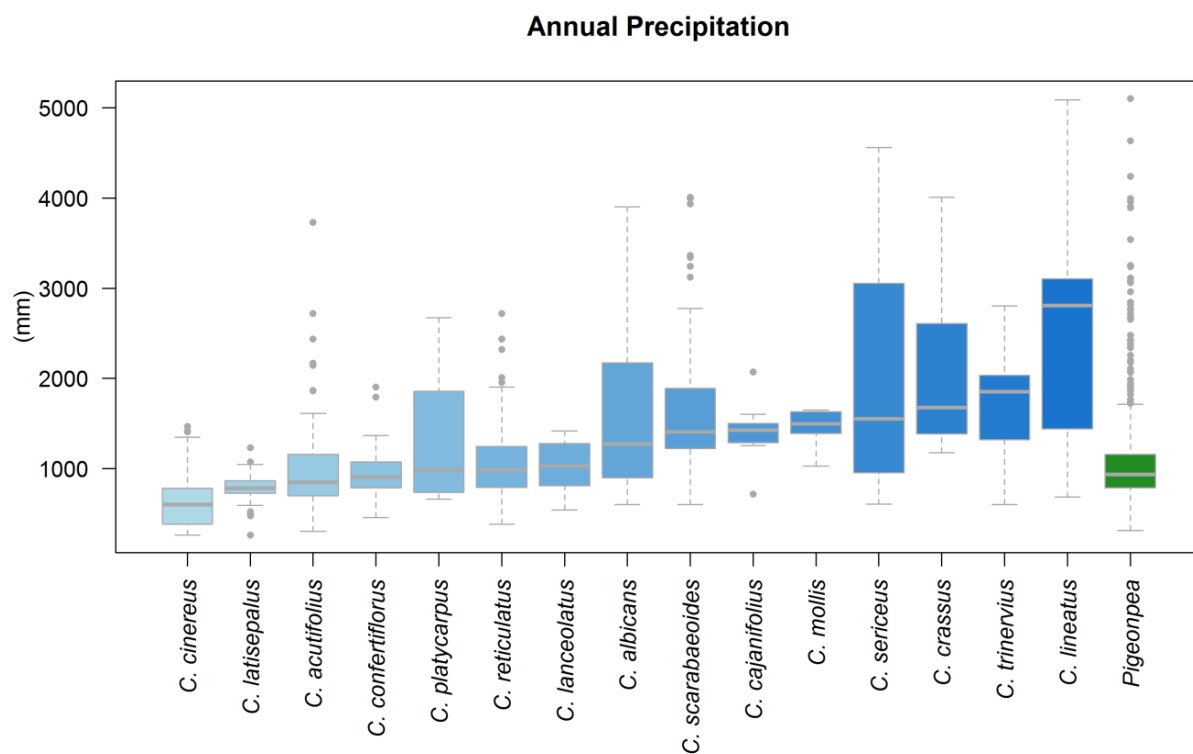
Supplementary Figure 5J. Climatic niches of CWR species and the pigeonpea crop for mean temperature of the driest quarter of the year.



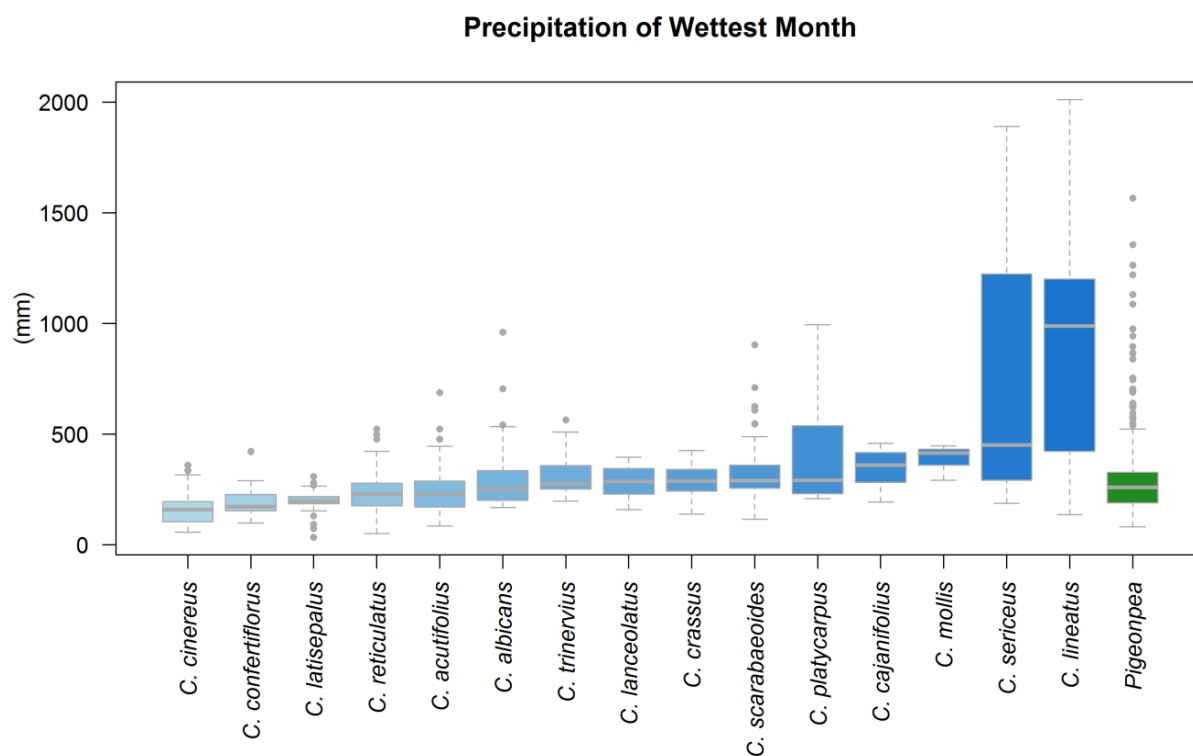
Supplementary Figure 5K. Climatic niches of CWR species and the pigeonpea crop for mean temperature of the warmest quarter of the year.



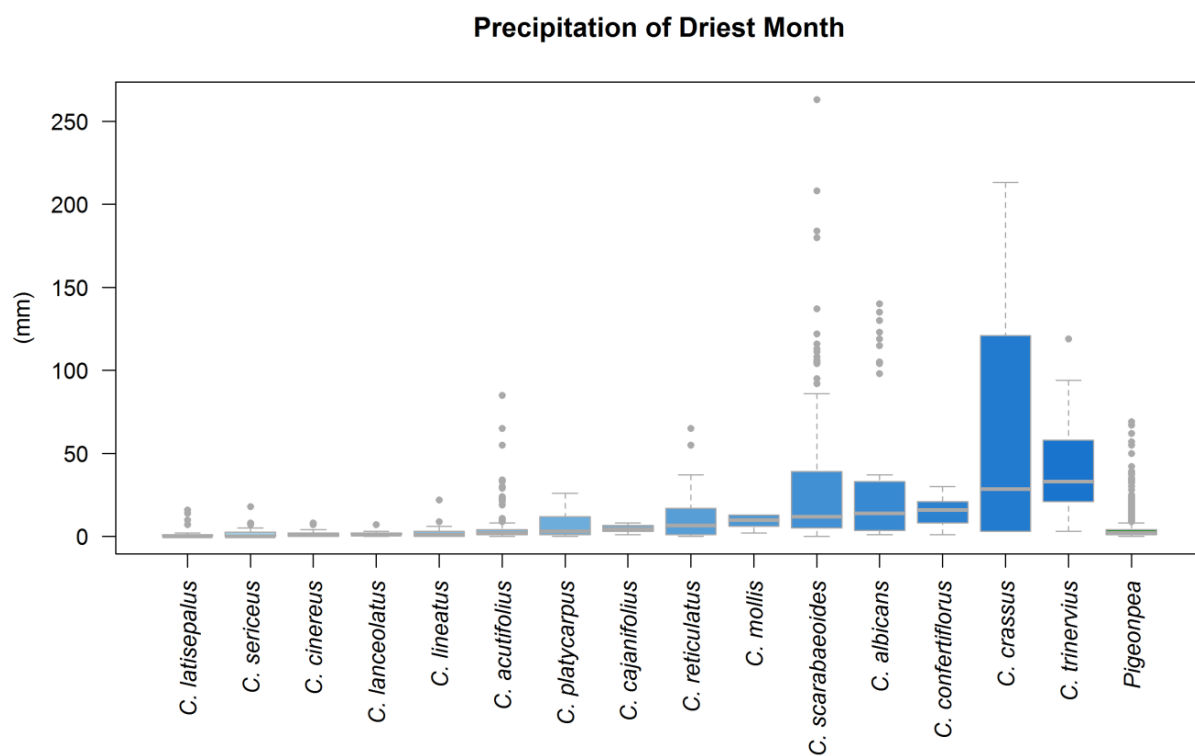
Supplementary Figure 5L. Climatic niches of CWR species and the pigeonpea crop for mean temperature of the coldest quarter of the year.



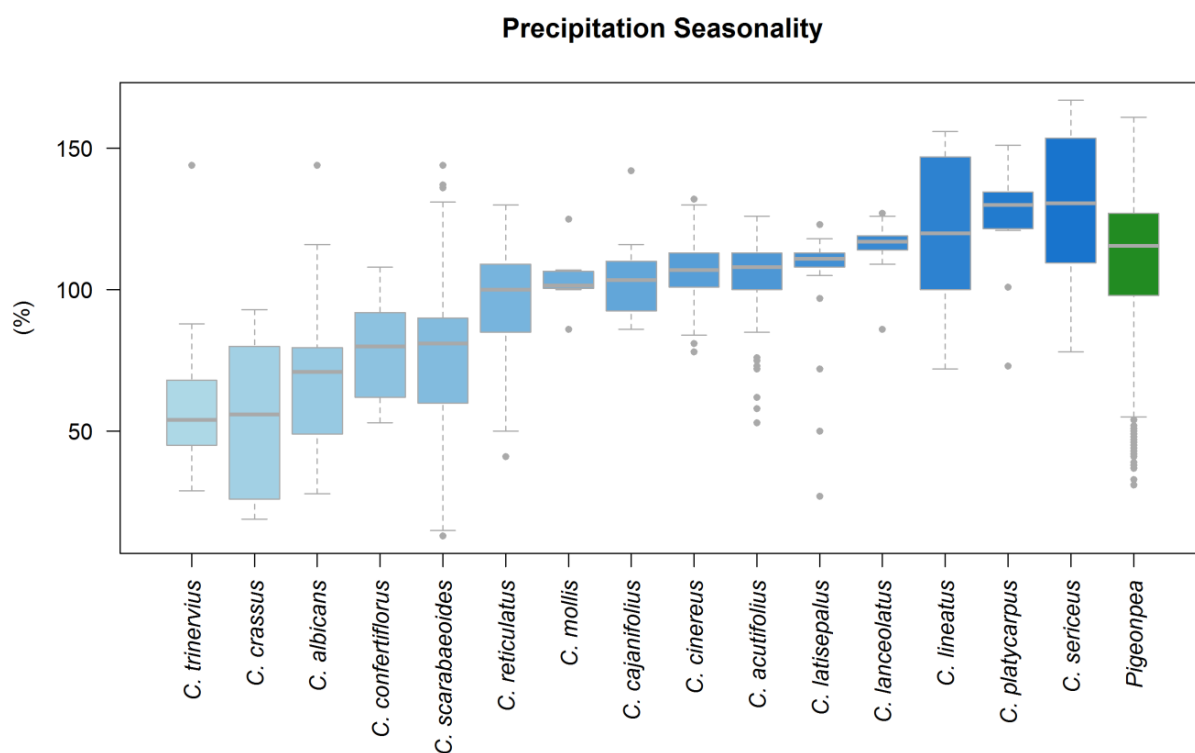
Supplementary Figure 5M. Climatic niches of CWR species and the pigeonpea crop for annual precipitation.



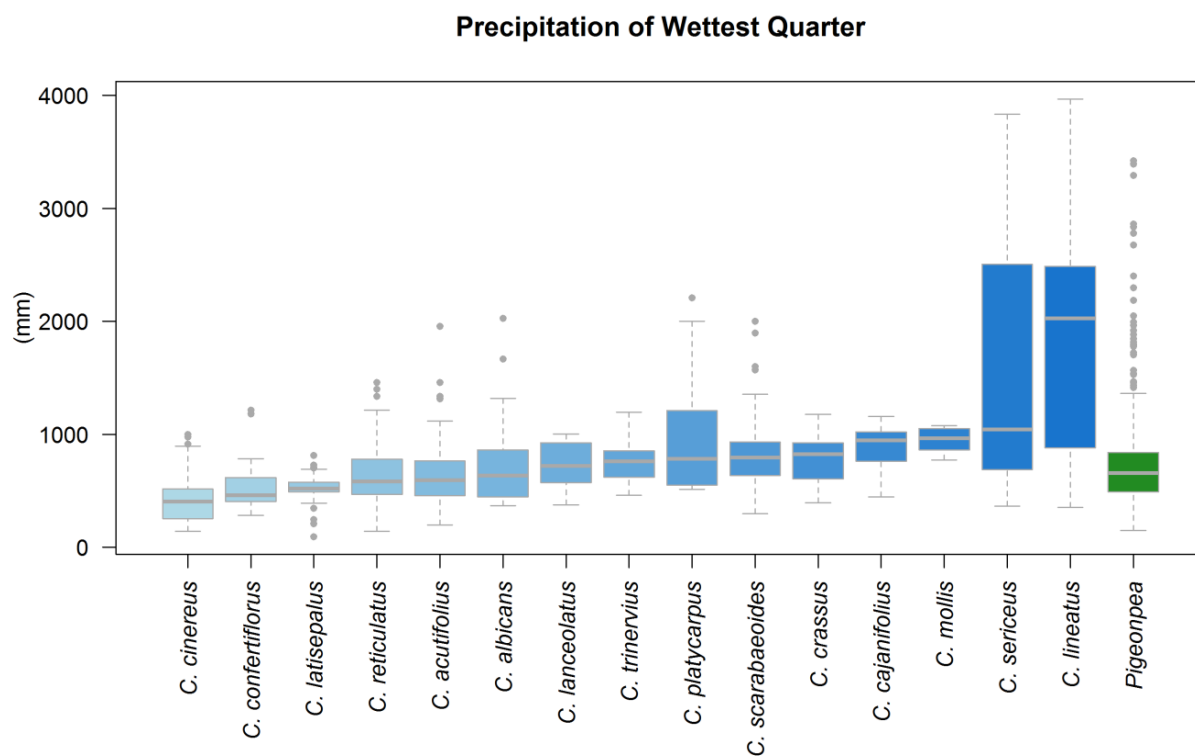
Supplementary Figure 5N. Climatic niches of CWR species and the pigeonpea crop for precipitation of the wettest month of the year.



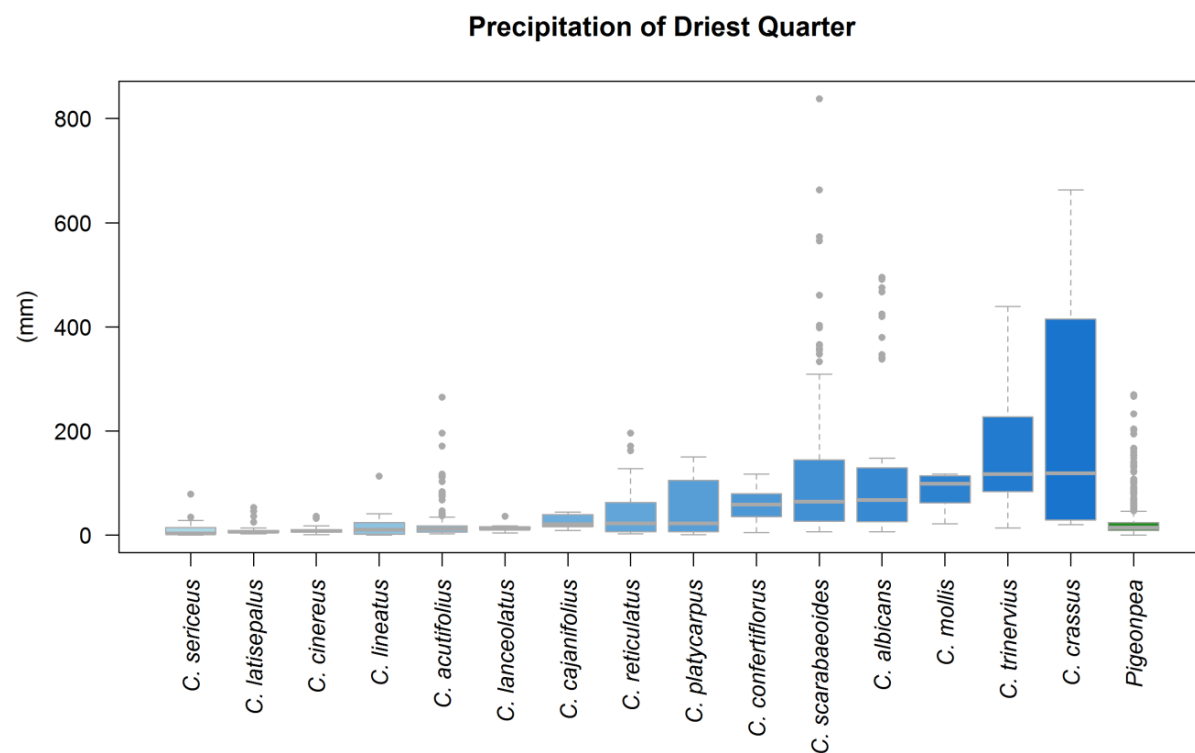
Supplementary Figure 5O. Climatic niches of CWR species and the pigeonpea crop for precipitation of the driest month of the year.



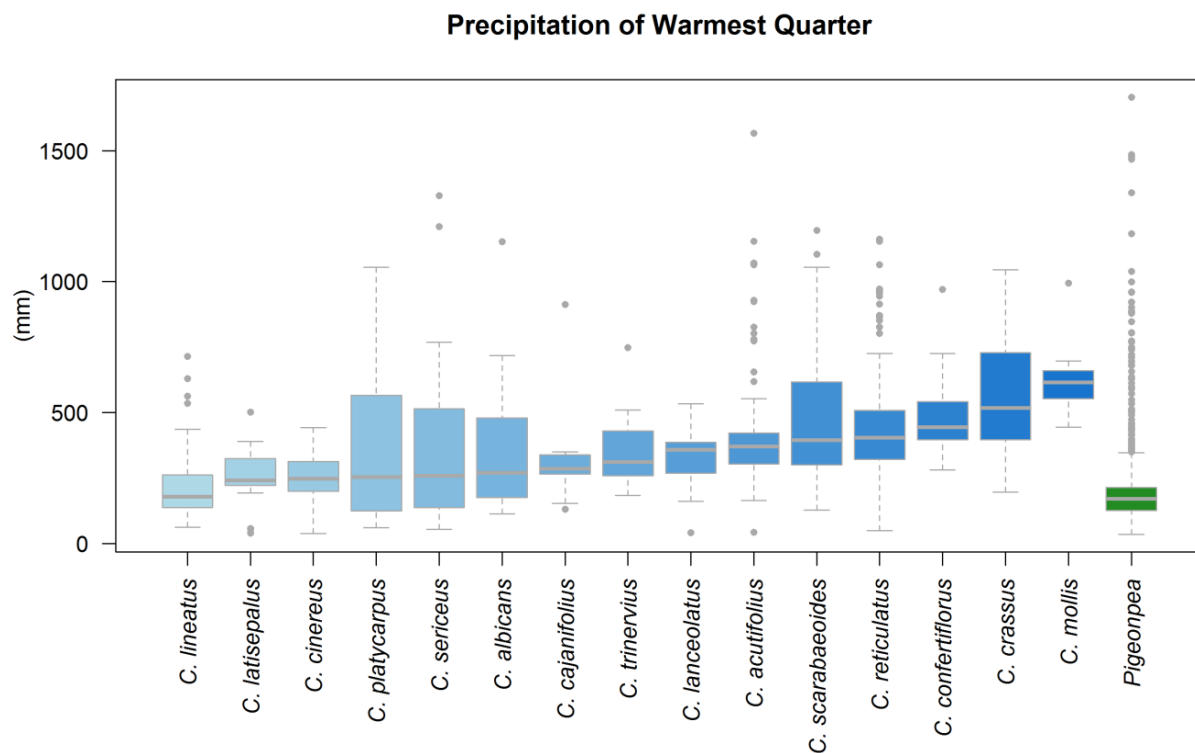
Supplementary Figure 5P. Climatic niches of CWR species and the pigeonpea crop for precipitation seasonality.



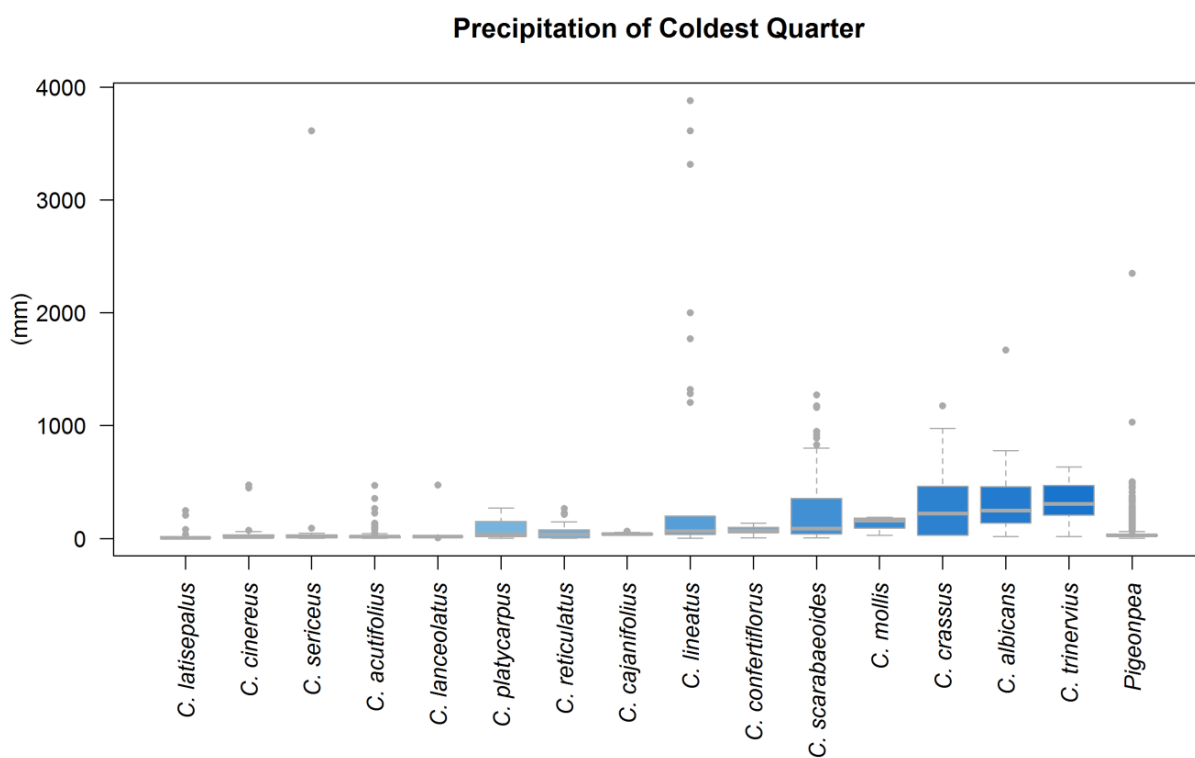
Supplementary Figure 5Q. Climatic niches of CWR species and the pigeonpea crop for precipitation of the wettest quarter of the year.



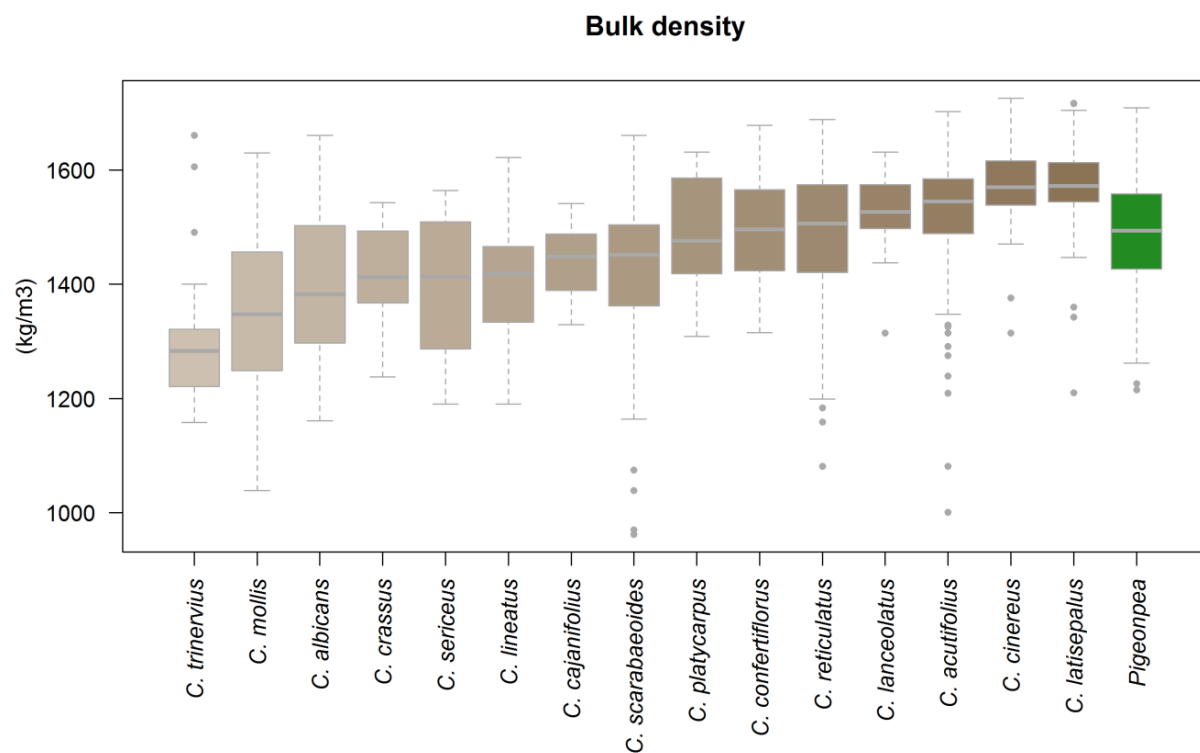
Supplementary Figure 5R. Climatic niches of CWR species and the pigeonpea crop for precipitation of the driest quarter of the year.



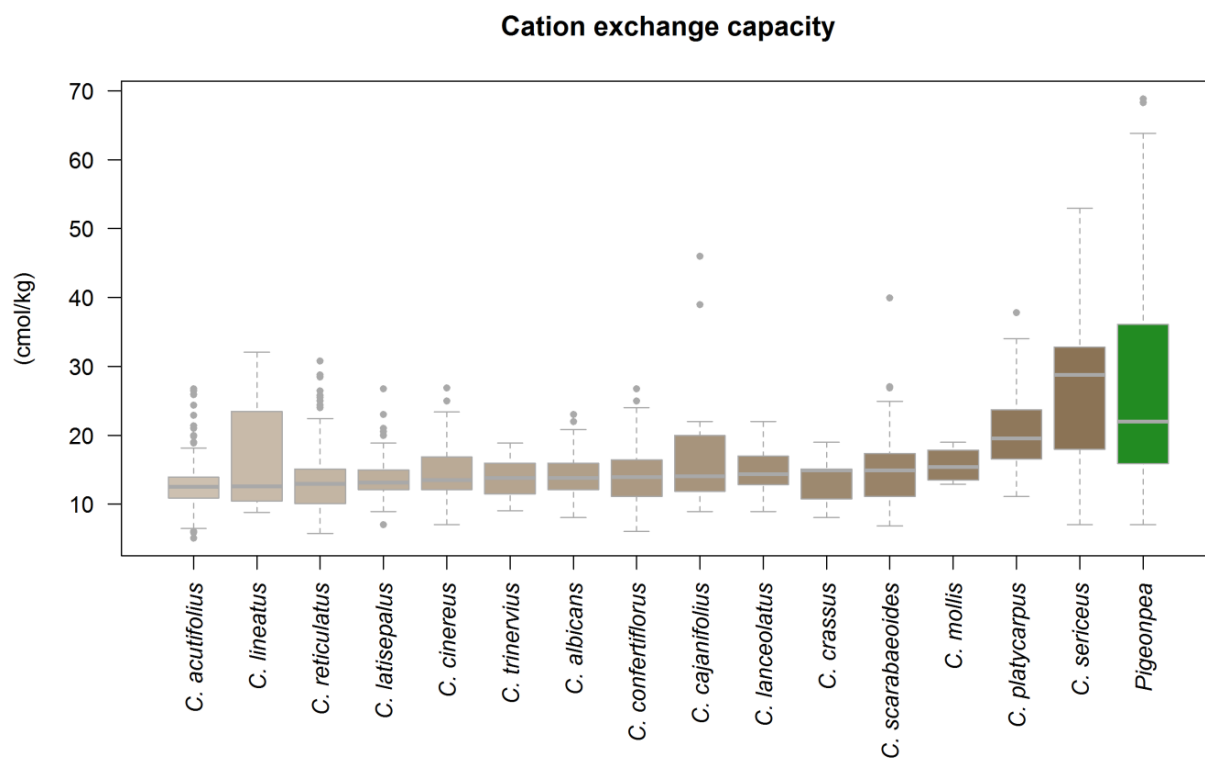
Supplementary Figure 5S. Climatic niches of CWR species and the pigeonpea crop for precipitation of the warmest quarter of the year.



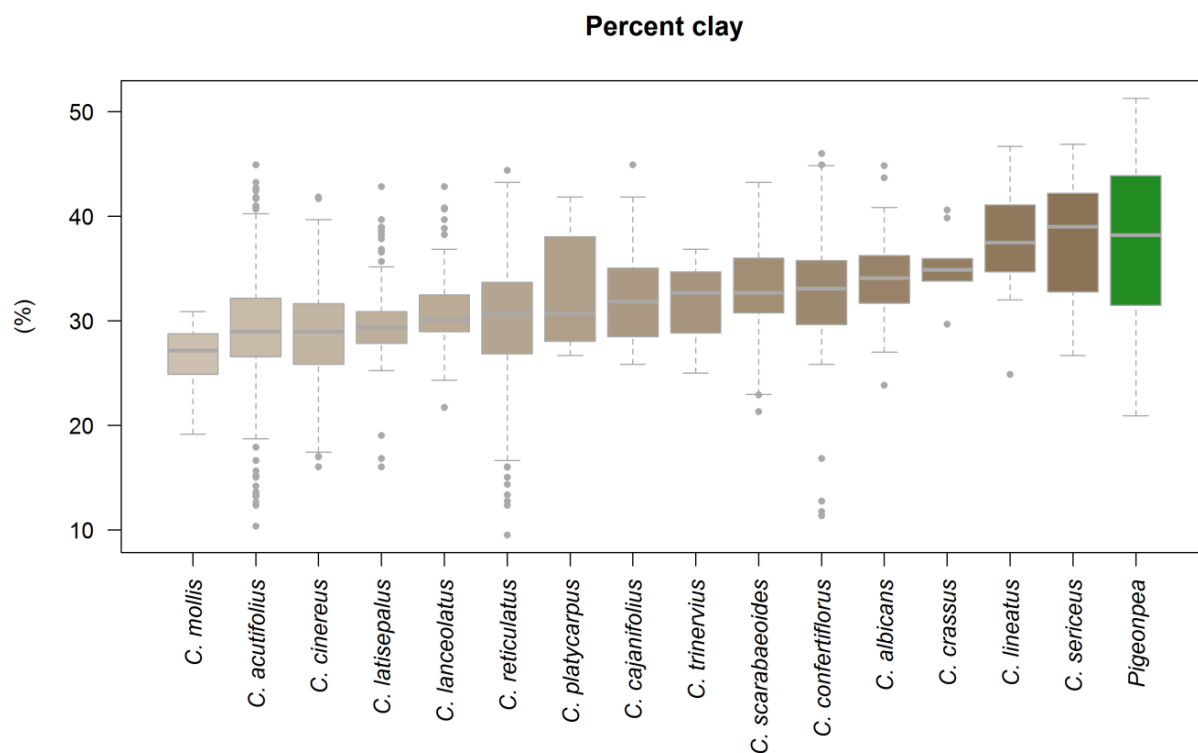
Supplementary Figure 5T. Climatic niches of CWR species and the pigeonpea crop for precipitation of the coldest quarter of the year.



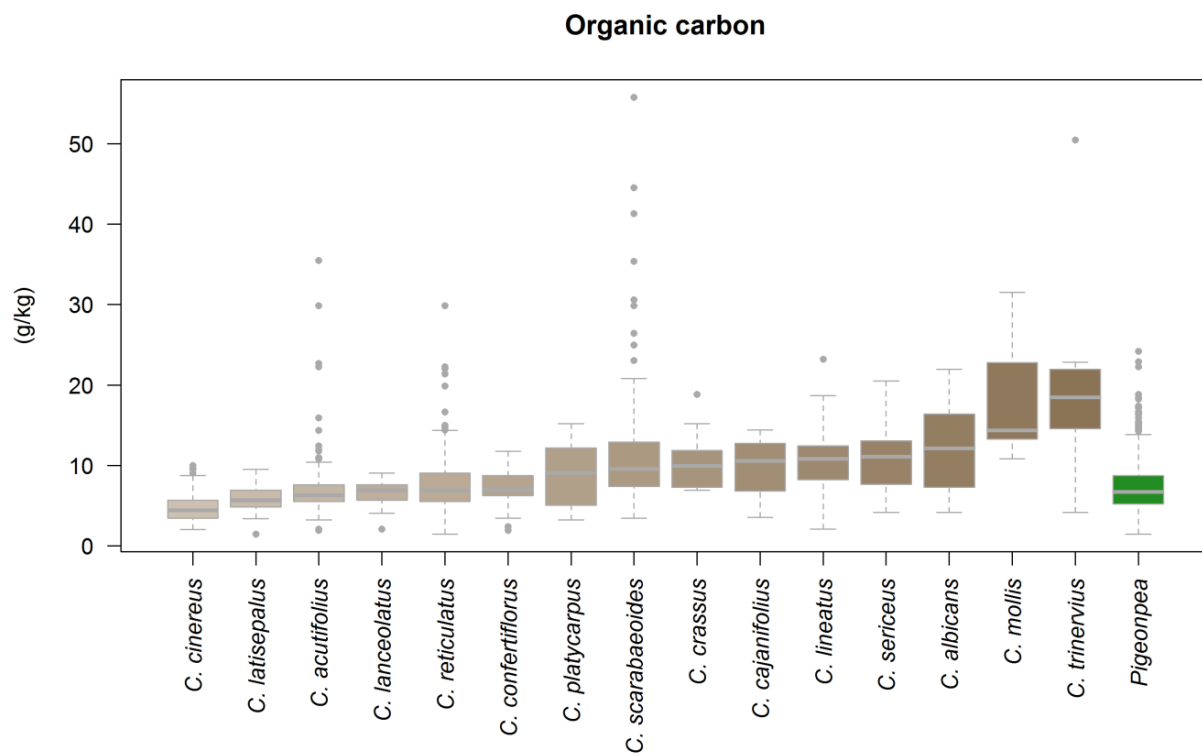
Supplementary Figure 5U. Edaphic niches of CWR species and the pigeonpea crop for bulk density.



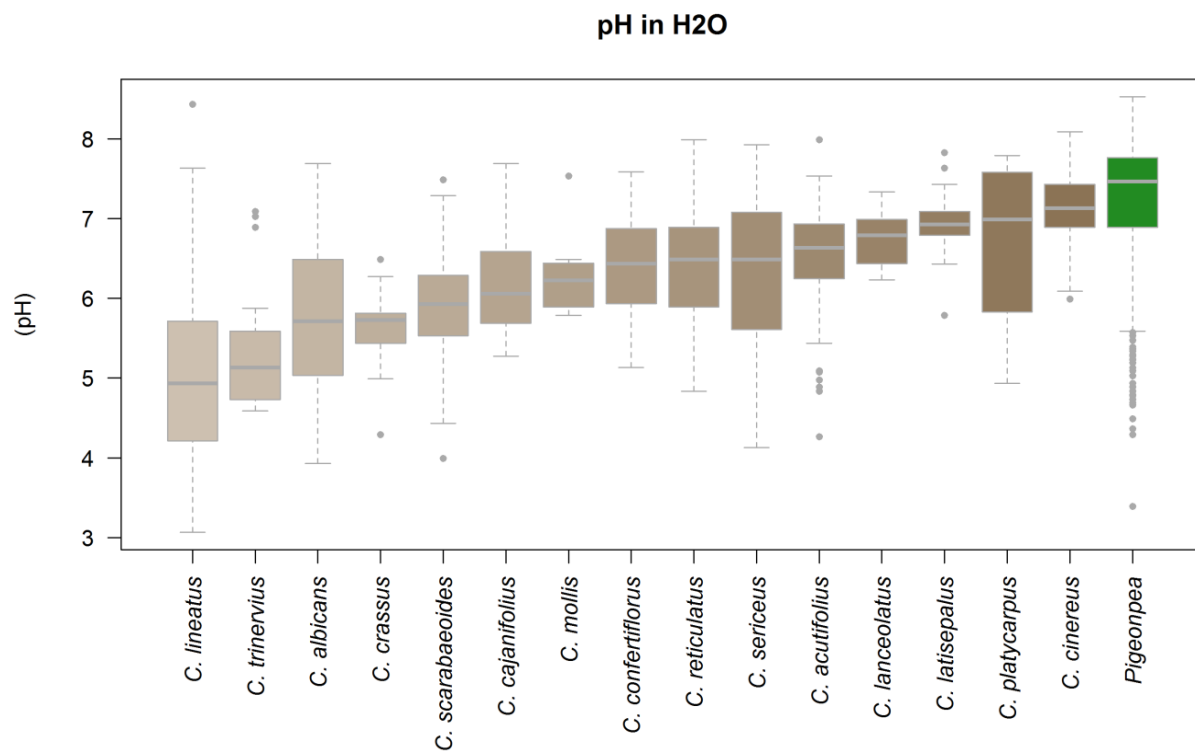
Supplementary Figure 5V. Edaphic niches of CWR species and the pigeonpea crop for cation exchange capacity.



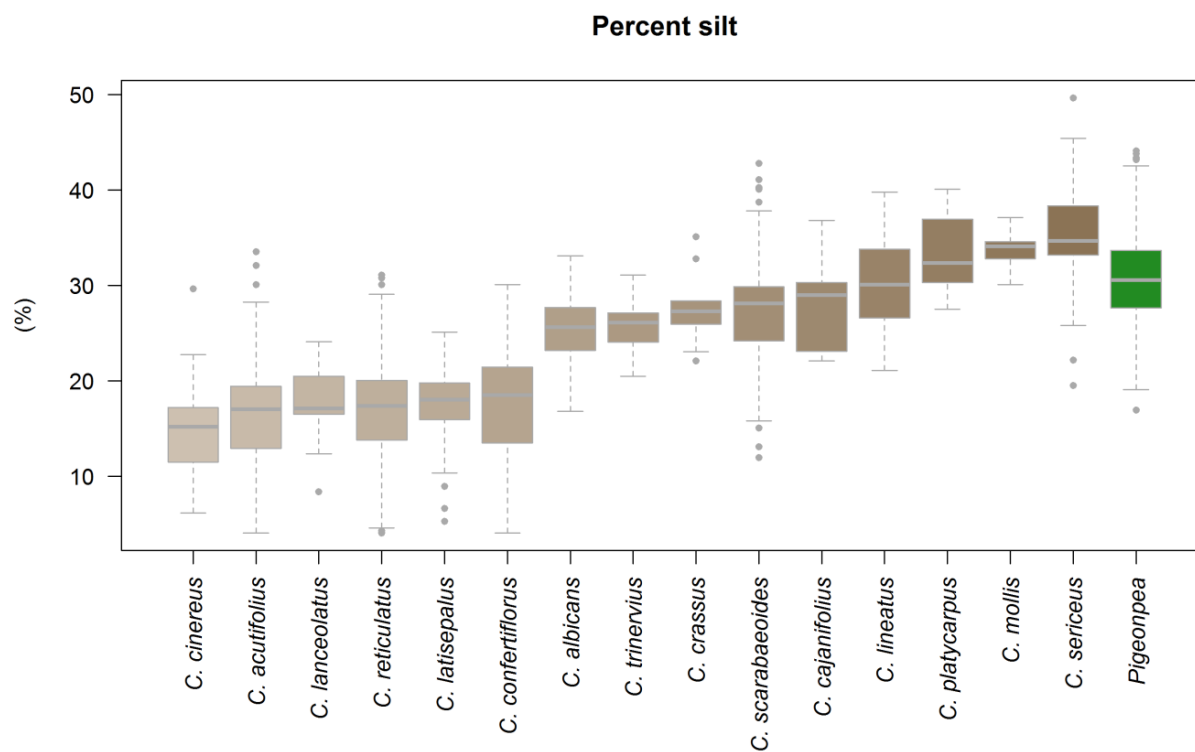
Supplementary Figure 5W. Edaphic niches of CWR species and the pigeonpea crop for percent clay.



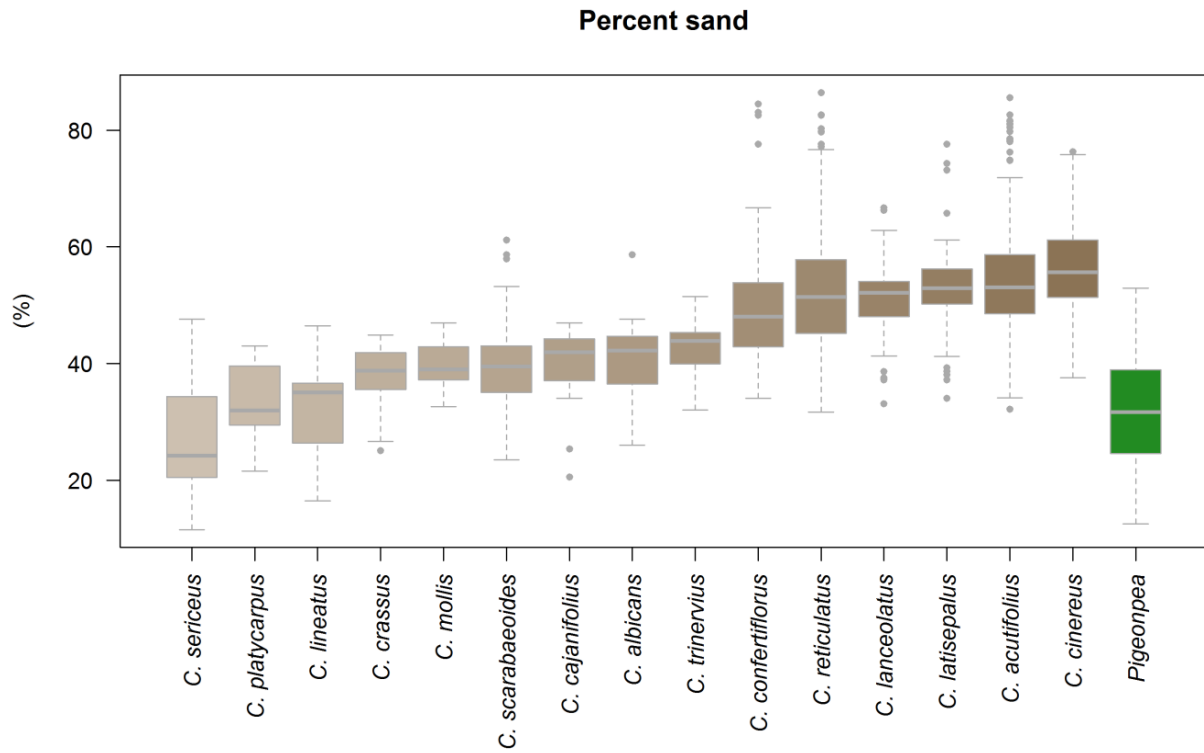
Supplementary Figure 5X. Edaphic niches of CWR species and the pigeonpea crop for organic carbon.



Supplementary Figure 5Y. Edaphic niches of CWR species and the pigeonpea crop for pH.



Supplementary Figure 5Z. Edaphic niches of CWR species and the pigeonpea crop for percent silt.



Supplementary Figure 5AA. Edaphic niches of CWR species and the pigeonpea crop for percent sand.

Supplementary Table 1. Sources of occurrence data for assessed CWR species.

Data provider	Record type	Number of records
Australian Grains Genebank (AGG)	Germplasm accession	201
International Center for Tropical Agriculture (CIAT)	Germplasm accession	75
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)	Germplasm accession	89
International Livestock Research Institute (ILRI)	Germplasm accession	2
Millennium Seed Bank Partnership, Royal Botanic Gardens, Kew	Germplasm accession	9
World Vegetable Center (AVRDC)	Germplasm accession	1
Australia's Virtual Herbarium (AD, BRI, CANB, CBG, DNA, HO, MEL, NSW, PERTH)	Herbarium or other record	1459
Bioversity International	Herbarium or other record	60
California Academy of Sciences Herbarium (CAS)	Herbarium or other record	7
Consortium of Pacific Northwest Herbaria (CPNWH)	Herbarium or other record	1
Harvard University Herbarium (HUH)	Herbarium or other record	70
Missouri Botanical Garden Herbarium (MO)	Herbarium or other record	9
Nationaal Herbarium Nederland (NHN)	Herbarium or other record	114
National Herbarium of New South Wales (NSW)	Herbarium or other record	74
Natural History Museum UK Herbarium (BM)	Herbarium or other record	13
New York Botanical Garden Herbarium (NY)	Herbarium or other record	49
Plants of Taiwan	Herbarium or other record	125
Royal Botanic Gardens, Edinburgh (E)	Herbarium or other record	40
Royal Botanic Gardens, Kew (K)	Herbarium or other record	165
Smithsonian Institution, National Herbarium (US)	Herbarium or other record	61
USDA, National Plant Germplasm System, Germplasm Resources Information Network (USDA NPGS GRIN)	Herbarium or other record	4
van der Maesen (1986)	Herbarium or other record	518
Western Australia Herbarium (PERTH) WAHerb	Herbarium or other record	25

Supplementary Table 2. Ecogeographic variables utilized in the species potential distribution modeling and niche analyses.

Variable number	Variable name	Units
0	Altitude	m
1	Annual mean temperature	°C
2	Mean diurnal temperature range	°C
3	Isothermality	N/A
4	Temperature seasonality (standard deviation)	°C
5	Maximum temperature of warmest month	°C
6	Minimum temperature of coldest month	°C
7	Temperature annual range	°C
8	Mean temperature of wettest quarter	°C
9	Mean temperature of driest quarter	°C
10	Mean temperature of warmest quarter	°C
11	Mean temperature of coldest quarter	°C
12	Annual precipitation	mm
13	Precipitation of wettest month	mm
14	Precipitation of driest month	mm
15	Precipitation seasonality (coefficient of variation)	%
16	Precipitation of wettest quarter	mm
17	Precipitation of driest quarter	mm
18	Precipitation of warmest quarter	mm
19	Precipitation of coldest quarter	mm
20	Bulk density	kg/m ³
21	Cation exchange capacity	cmol/kg
22	Percent clay	%
23	Organic carbon	g/kg
24	pH in H ₂ O	pH
25	Percent silt	%
26	Percent sand	%

Supplementary Table 3. Gap analysis and comparable expert evaluation priorities results per crop wild relative species.

Taxon	GenePool	Total Records	Total Records- Distinct	H Records- Total	H Records- Distinct	G Records- Total	G Records- Distinct	SRS	Ecogeographic Variables	ATAUC	STAUc	ASD15	cAUC	GRS	ERS	FPS	FPCAT	EPS 1	EPS 2	EPS 3	EPS 4	EPS mean
<i>C. acutifolius</i>	2	594	192	562	185	32	7	0.54	1, 2, 9, 17, 18, 20, 23, 24	0.93	0.01	0.28	0.44	0.56	3.33	1.48	HPS	1	3	2	1	1.75
<i>C. albicans</i>	2	175	45	145	27	30	18	1.71	0, 7, 18, 20, 23, 24	0.93	0.10	10.90	0.43	1.01	7.65	3.46	MPS	0	8	4	3	3.75
<i>C. cajanifolius</i>	2	52	13	47	8	5	5	0.96	8, 11, 14, 18, 24	0.78	0.11	0.07	0.32	0.63	5.45	2.35	HPS	0	3	2	3	2
<i>C. cinereus</i>	2	273	132	267	132	6	0	0.22	1, 3, 7, 12, 14, 15, 17, 19, 24	0.93	0.01	1.74	0.44	0.00	0.00	0.07	HPS	0	8	2	1	2.75
<i>C. confertiflorus</i>	2	124	47	122	46	2	1	0.16	2, 8, 12, 15, 18, 20	0.97	0.01	2.55	0.48	0.09	3.00	1.08	HPS	1	5	1	0	1.75
<i>C. crassus</i>	3	34	14	21	11	13	3	3.82	11, 20, 22, 23, 24, 25	0.89	0.02	0.00	0.41	0.09	0.61	1.51	HPS	1	8	1	2	3
<i>C. lanceolatus</i>	2	140	41	134	39	6	2	0.43	2, 3, 14, 15, 19, 20, 24, 25, 26	0.98	0.02	0.70	0.49	0.38	1.33	0.71	HPS	0	6	2	0	2
<i>C. latiseptatus</i>	2	233	86	228	86	5	0	0.21	3, 6, 12, 15, 17, 19	0.96	0.02	0.00	0.47	0.00	0.00	0.07	HPS	0	5	1	1	1.75
<i>C. lineatus</i>	2	180	34	170	26	10	8	0.56	2, 3, 8, 9, 12, 17, 23, 24	0.99	0.01	1.63	0.49	1.18	4.67	2.13	HPS	0	9	1	1	2.75
<i>C. mollis</i>	3	15	8	3	1	12	7	8.00	2, 3, 17, 26	0.67	0.05	0.00	0.18	3.57	8.18	6.58	LPS	0	6	5	3	3.5
<i>C. platycarpus</i>	3	38	12	9	1	29	11	7.63	3, 7, 14, 18, 23	0.75	0.10	0.19	0.30	1.27	10.00	6.30	LPS	0	9	2	4	3.75
<i>C. reticulatus</i>	2	483	228	467	228	16	0	0.33	2, 4, 8, 9, 15, 18, 19, 24	0.95	0.01	0.48	0.45	0.00	0.00	0.11	HPS	1	6	2	3	3
<i>C. scarabaeoides</i>	2	577	162	397	113	180	57	3.12	0, 5, 11, 12, 14, 15, 19, 20	0.95	0.02	0.42	0.50	0.61	2.21	1.98	HPS	0	8	6	4	4.5
<i>C. sericeus</i>	2	96	29	85	20	11	9	1.15	4, 10, 13, 14, 20, 21, 22, 23, 24	0.94	0.02	1.68	0.48	0.43	3.82	1.80	HPS	0	6	2	2	2.5
<i>C. trinervius</i>	2	134	25	131	23	3	3	0.22	3, 7, 9, 12, 14, 19, 20, 26	0.96	0.05	4.71	0.46	0.67	2.40	1.10	HPS	0	3	1	1	1.25

GenePool 2 is secondary; 3 is tertiary. Distinct records denote occurrence records with unique coordinates. H records denote herbarium and other locality data used to inform species distribution models. G records denote germplasm accessions accessible to the global community. The final priority score (FPS) for the species is the mean of the sampling representativeness (SRS), ecogeographic variables (ERS), geographic representativeness score (GRS), and ecological representativeness score (EPS). Ecogeographic variables lists those species-specific bioclimatic and edaphic variables used in Maxent modeling per taxon. Maxent validation techniques include the 5-fold average area under the ROC curve of test data (ATAUC), the standard deviation of the test AUC of the 5 different folds (ST AUC), the proportion of the potential distribution coverage with standard deviation above 0.15 (ASD15), and the calibrated AUC (cAUC). Final priority categories (FPCAT) for the CWR of pigeonpea included high (HPS), medium (MPS), and low priority species (LPS) for further collecting for ex situ conservation. EPS denotes comparable expert priority scores for 4 experts, along with their mean value.

Supplementary Table 4. Countries identified for potential further collecting per crop wild relative species.

Taxon	Priority	Countries identified for further collecting
<i>C. acutifolius</i>	HPS	Australia
<i>C. albicans</i>	MPS	India, Sri Lanka
<i>C. cajanifolius</i>	HPS	India
<i>C. cinereus</i>	HPS	Australia
<i>C. confertiflorus</i>	HPS	Australia
<i>C. crassus</i>	HPS	China, India, Indonesia, Lao PDR, Malaysia, Papua New Guinea, Philippines, Thailand, Timor-Leste, Vietnam
<i>C. lanceolatus</i>	HPS	Australia
<i>C. latisepalus</i>	HPS	Australia
<i>C. lineatus</i>	HPS	India, Sri Lanka
<i>C. mollis</i>	LPS	Nepal
<i>C. platycarpus</i>	LPS	None modeled, but expected in India
<i>C. reticulatus</i>	HPS	Australia
<i>C. scarabaeoides</i>	HPS	Australia, Bangladesh, Bhutan, China, Fiji, India, Indonesia, Japan, Madagascar, Malaysia, Myanmar, Nepal, Papua New Guinea, Philippines, Sri Lanka, Taiwan, Thailand, Timor-Leste, Vietnam
<i>C. sericeus</i>	HPS	India
<i>C. trinervius</i>	HPS	India, Sri Lanka

REFERENCES

- Araújo, M.B., Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33 (10), 1677-1688.
- Ariyanayagam, R.P., Rao, A.N., Zavery, P.P. (1995). Cytoplasmic-genic male-sterility in interspecific matings of *Cajanus*. *Crop Sci.* 35, 981-985.
- Bivand, R., Keitt, T., Rowlingson, B. (2014). *Rgdal: Bindings for the geospatial data abstraction library*. R package version 0.8-16. <http://CRAN.R-project.org/package=rgdal>.
- Bivand, R., Lewin-Koh, N. (2014). *Maptools: Tools for reading and handling spatial objects*. R package version 0.8-30. <http://CRAN.R-project.org/package=maptools>.
- Bivand, R., Pebesma, E., Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition*. Springer, NY. <http://www.asdar-book.org/>.
- Bogdan, A.V. (1977). *Tropical Pasture and Fodder Plants* (New York: Longman).
- Brar, D.S., Khush, G.S. (1997). Alien introgression in rice. *Plant Mol. Biol.* 35 (1-2), 35-47.
- Choudhary, A.K., Kumar, S., Patil, B.S., Bhat, J.S., Sharma, M., Kemal, S., *et al.* (2013). Narrowing yield gaps through genetic improvement for *Fusarium* wilt resistance in three pulse crops of the semi-arid tropics. *SABRAO J. Breed. Genet.* 45 (3), 341-370.
- Choudhary, A.K., Sultana, R., Pratap, A., Nadarajan, N., Jha, U.C. (2011). Breeding for abiotic stresses in pigeonpea. *J. Food Legumes* 24 (3), 165-174.
- Cordell, D., Drangert, J.-O., White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Chang.* 19 (2), 292-305.
- Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodivers. Conserv.* 19, 883-899.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sust. Food Syst.* 38 (4), 369-77.
- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., *et al.* (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
- Esquinas-Alcázar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* 6 (12), 946-953.
- ESRI (2011). ArcGIS Desktop: Release 10.1.
- FAO (Food and Agriculture Organization of the United Nations). (2002). *International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2010a). *Crop Calendar*. Online Resource. <http://www.fao.org/agriculture/seed/cropcalendar/welcome.do>. (Rome: FAO). Accessed December 2013.
- FAO (Food and Agriculture Organization of the United Nations). (2010b). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- Flower, D.J., Ludlow, M.M. (1987). Variation among accessions of pigeonpea (*Cajanus cajan*) in osmotic adjustment and dehydration tolerance of leaves. *Field Crop. Res.* 17 (3-4), 229-243.
- Foley, J.A., Ramankutty, N., Brauman, K.A., Cassidy, E.S., Gerber, J.S., Johnston, M., *et al.* (2011). Solutions for a cultivated planet. *Nature* 478 (7369), 337-342.
- Google Maps Geocoder (2013). Online resource. <https://developers.google.com/maps/documentation/geocoding/?hl=pl> (accessed June 2013).
- GRIN (Germplasm Resources Information Network), USDA, ARS, National Genetic Resources Program (2012). National Germplasm Resources Laboratory, Beltsville, Maryland. Online resource.

- <http://www.ars-grin.gov/~sbmljw/cgi-bin/index.pl?language=en>. Accessed March 2013.
- Guarino, L., Lobell, D.B. (2011). A walk on the wild side. *Nat. Clim. Change* 1 (8), 374-375.
- Harlan, J.R., de Wet, J.M.J. (1971). Toward a rational classification of cultivated plants. *Taxon* 20, 509-517.
- Hengl, T., de Jesus, J.M., MacMillan, R.A., Batjes, N.H., Heuvelink, G.B.M., Ribeiro, E., *et al.* (2014). SoilGrids1km — Global soil information based on automated mapping. *PLoS One* 9 (8), e105992.
- Henry, R.J. (2014). Genomics strategies for germplasm characterization and the development of climate resilient crops. *Front. Plant Sci.* 5, 68.
- Hijmans, R.J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology* 93, 679-688.
- Hijmans, R.J. (2014). *Raster: geographic data analysis and modeling*. R package version 2.2-31. <http://CRAN.R-project.org/package=raster>.
- Hijmans, R.J., Cameron, S., Parra, J., Jones, P.G., Jarvis, A. (2005). *WorldClim*, Version 1.3 (Berkeley: University of California).
- Hijmans, R.J., Guarino, L., Cruz, M., Rojas, E. (2001). Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* 127, 15-19.
- Hijmans, R.J., Phillips, S., Leathwick, J., Elith, J. (2013). *Dismo: Species distribution modeling*. R package version 0.9-3. <http://CRAN.R-project.org/package=dismo>.
- Hijmans, R.J., Schreuder, M., De La Cruz, J., Guarino, L. (1999). Using GIS to check co-ordinates of genebank accessions. *Genet. Resour. Crop Ev.* 46, 291-296.
- Husson, F., Josse, J., Le, S., Mazet, J. (2009). *FactoMineR: multivariate exploratory data analysis and data mining with R*. v.1.26. <http://factominer.free.fr>.
- Jarvis, A., Lane, A., Hijmans, R.J. (2008). The effect of climate change on crop wild relatives. *Agr. Ecosyst. Environ.* 126 (1-2), 13-23.
- Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet. Resour. Crop Ev.* 52 (6), 671-682.
- Jones, R.B., Freeman, H.A., Lo Monaco, G. (2002). Improving the access of small farmers in Eastern and Southern Africa to global pigeonpea markets. *Agricultural Research and Extension Network* 120, 1-11.
- Kassa, M.T., Varma Penmetsa, R., Carrasquilla-Garcia, N., Sarma, B.K., Datta, S., Upadhyaya, H.D., *et al.* (2012). Genetic patterns of domestication in pigeonpea (*Cajanus cajan* (L) Millsp) and wild *Cajanus* relatives. *PloS One* 7 (6), e39563.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57 (4), 625-639.
- Kramer-Schadt, S., Niedballa J., Pilgrim J.D., Schroder B., Lindenborn J., Reinfelder, V., *et al.* (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19 (11), 1366-1379.
- Liu, C., Berry, P.M., Dawson, T.P., Pearson, R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28 (3), 385-393.
- Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* 319 (5863), 607-610.
- Loiselle, B., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., *et al.* (2008). Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35, 105-116.
- Lopez, F.B., Setter, T.L., McDavid, C.R. (1987). Carbon dioxide and light responses of photosynthesis

- in cowpea and pigeonpea during water deficit and recovery. *Plant Physiology* 85 (4), 990-995.
- Mallikarjuna, N., Jadhav, D., Reddy, M.V., Usharani, D.-T. (2005). Introgression of *Phytophthora* blight disease resistance from *Cajanus platycarpus* into short duration pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Indian J. Genet. Pl. Br.* 65 (4), 261-263.
- Mallikarjuna, N., Saxena, K.B. (2005). A new cytoplasmic nuclear male-sterility system derived from cultivated pigeonpea cytoplasm. *Euphytica* 142 (1-2), 143-48.
- Mallikarjuna, N., Saxena, K.B., Jadhav, D.R. (2011). "Cajanus," in: *Wild Crop Relatives: Genomic and Breeding Resources, Legume Crops and Forages*, ed. C. Kole (Berlin, Heidelberg: Springer-Verlag), 21-33.
- Mallikarjuna, N., Sharma, H.C., Upadhyaya, H.D. (2007). Exploitation of wild relatives of pigeonpea and chickpea for resistance to *Helicoverpa armigera*. *Journal of SAT Agricultural Research* 3 (1), 4.
- McCouch, S., Baute, G., Bradeen J., Bramel, P., Bretting, P.K., Buckler, E., *et al.* (2013). Agriculture: feeding the future. *Nature* 499, 23-24.
- Mlilo, J.K., Craufurd, P.Q. (2005). Adaptation and yield of pigeonpea in different environments in Tanzania. *Field Crop. Res.* 94 (1), 43-53.
- Monfreda, C., Ramankutty, N., Foley, J.A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochem. Cy.* 22, GB1022. Data: <http://www.geog.mcgill.ca/landuse/pub/Data/175crops2000/>.
- Mudaraddi, B., Saxena, K.B., Saxena, R.K., Varshney, R.K. (2013). Molecular diversity among wild relatives of *Cajanus cajan* (L.) Millsp. *Afr. J. Biotechnol.* 12 (24), 3797-3801.
- Odeny, D.A. (2007). The potential of pigeonpea (*Cajanus cajan* (L.) Millsp.) in Africa. *Nat. Resour. Forum* 31 (4), 297-305.
- Olson, D.M., Dinerstein, E., Wikramanayake, E.D., Burgess, N.D., Powell, G.V.N., Underwood, E.C., *et al.* (2001). Terrestrial ecoregions of the world: a new map of life on earth. *BioScience* 51 (11), 933-938.
- Pebesma, E.J., Bivand, R.S. (2005). Classes and methods for spatial data in R. *R News* 5 (2), <http://cran.r-project.org/doc/Rnews/>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190 (3-4), 231-259.
- Phillips, S.J., Dudik, M. (2008). Modeling of species distributions with maxent: new extensions and comprehensive evaluation. *Ecography* 31, 161-175.
- R Core Team (2013). *R: A language and environment for statistical computing*. <http://www.r-project.org/>.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop gene pools: a case study with *Phaseolus* beans. *PloS One* 5 (10), e13497.
- Ray, D.K., Mueller, N.D., West, P.C., Foley, J.A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8 (6), e66428.
- Reddy, L.J. (1990). "Pigeonpea: Morphology," in: *The Pigeonpea*, eds. Y.L. Nene, S.D. Hall, V.K. Sheila, (Wallingford: CAB International). 47-87.
- Reddy, L.J., Green, J.M., Singh, U., Bisen, S.S., Jambunathan, R. (1979). "Seed protein studies on *Cajanus cajan*, *Atylosia* spp. and some hybrid derivatives," in: *Seed protein improvement in cereals and grain legumes*. Proceedings of an international symposium jointly organized by IAEA and FAO in cooperation with GSF Neuherberg, 4-8 September 1978, 105-117.
- Rosenzweig, C., Elliot, J., Deryng, D., Ruane, A.C., Muller, C., Arneth, A., *et al.* (2013). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl Acad. Sci.* 111 (9), 3268-3273.

- Sahai, K., Rawat, K.K. (2014). A survey of the degrading population of *Cajanus lineatus* (Wight & Arn.) Maesen, in parts of the Western Ghats, India. *Genet. Resour. Crop Ev.*, 1-9. doi: 10.1007/s10722-014-0176-4.
- Sanchez, G. (2012). Package 'plsdepot'. *Partial Least Squares (PLS) Data Analysis Methods*, v. 0.1.17.
- Sardana, V., Sharma, P., Sheoran, P. (2010). "Growth and production of pulses," in: *Soils, Plant Growth and Crop Production- Vol III- Growth and Production of Pulses. Encyclopedia of Life Support Systems*. <http://www.eolss.net/sample-chapters/c10/e1-05a-44-00.pdf>.
- Saxena, K.B. (2005). Pigeonpea [*Cajanus cajan* (L.) Millsp.]. in: *Genetic Resources, Chromosome Engineering, and Crop Improvement: Volume I, Grain Legumes*, eds. R.J. Singh, P.P. Jauhar (Boca Raton: CRC Press), 86-110.
- Saxena, K.B., Kumar, R.V. (2003). Development of a cytoplasmic nuclear male-sterility system in pigeonpea using *C. scarabaeoides* (L.) Thouars. *Indian J. Genetics* 63 (3), 225-229.
- Saxena, K.B., Kumar, R.V., Gowda, C.L.L. (2010). Vegetable pigeonpea- a review. *J. Food Legumes* 23 (2), 91-98.
- Saxena, K.B., Kumar, R.V., Srivastava, N. Shiyang, B. (2005). A cytoplasmic-nuclear male-sterility system derived from a cross between *Cajanus cajanifolius* and *Cajanus cajan*. *Euphytica* 145 (3), 289-294.
- Saxena, K.B., Kumar, R.V., Tikle, A.N., Saxena, M.K., Gautam, V.S., Rao, S.K., et al. (2013). ICPH 2671 – the world's first commercial food legume hybrid. *Plant Breeding* 132, 479-485.
- Saxena, K.B., Sawargaonkar, S. L. (2014). First information on heterotic groups in pigeonpea [*Cajanus cajan* (L.) Millsp.]. *Euphytica* 200 (2), 187-196.
- Saxena, K.B., Srivastava, D.P., Tikla, S.B.S. (1998). "Breaking yield barrier in pigeonpea through hybrid breeding". in: *Proceedings of the National Symposium on Biotic and Abiotic Stress of Pulse Crops* (Kanpur: Indian Institute of Pulses Research), 55-64.
- Sharma, H.C., Pampapathy, G., Reddy, L.J. (2003). Wild relatives of pigeonpea as a source of resistance to the pod fly (*Melanagromyza obtusa* Malloch) and pod wasp (*Tanaostigmodes cajaninae* La Salle). *Genet. Resour. Crop Ev.* 50, 817-824.
- Sodhi, N.S., Koh, L.P., Brook, B.W., Ng, P.K.L. (2004). Southeast Asian biodiversity: an impending disaster. *Trends Ecol. Evol.* 19 (12), 654-660.
- Sodhi, N.S., Posa, M.R.C., Lee, T.M., Bickford, D., Koh, L.P., Brook, B.W. (2009). The state and conservation of Southeast Asian biodiversity. *Biodivers. Conserv.* 19 (2), 317-328.
- Subbarao, G.V., Chauhan, Y.S. Johansen, C. (2000). Patterns of osmotic adjustment in pigeonpea - its importance as a mechanism of drought resistance. *Eur. J. Agron.* 12 (3-4), 239-249.
- Subbarao, G.V., Johansen, C., Jana, M.K., Kumar Rao J.V.D.K. (1990). Comparative salinity responses among pigeonpea wild relatives. *Crop Sci.* 31, 415-418.
- Tapia, C., Torres, E., Parra-Quijano, M. (2014). Searching for adaptation to abiotic stress: ecogeographical analysis of highland Ecuadorian maize. *Crop Sci.* 55 (1), 262-274.
- Tester, M., Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327 (5967), 818-822.
- The Plant List. (2010). Version 1. <http://www.theplantlist.org/>.
- Tilman, D., Clark, M. (2014). Global diets link environmental sustainability and human health. *Nature* 515 (7528), 518-522.
- Upadhyaya, H.D. (2006). Improving pigeonpea with the wild. *SATrends* January 2006.
- Upadhyaya, H.D., Reddy, K.N., Pundir, R.P.S., Singh, S., Gowda, C.L.L., Irshad Ahmed, M. (2013a). Diversity and geographical gaps in *Cajanus scarabaeoides* (L.) Thou. germplasm conserved at the ICRISAT genebank. *Plant Genetic Resources* 11 (1), 3-14.
- Upadhyaya, H.D., Reddy, K.N., Sastry, D.V.S.S.R., Gowda, C.L.L. (2007). Identification of

- photoperiod insensitive sources in the world collection of pigeonpea at ICRISAT. *J. SAT. Agri. Res. J.* 3 (1), 46-49.
- Upadhyaya, H.D., Reddy, K.N., Singh, S., Gowda, C.L.L. (2013b). Phenotypic diversity in *Cajanus* species and identification of promising sources for agronomic traits and seed protein content. *Genet. Resour. Crop Ev.* 60 (2), 639-659.
- Valenzuela, H., Smith J. (2002). "Pigeonpea," in: *Sustainable Agriculture* (Honolulu: University of Hawaii).
- van der Maesen, L.J.G. (1986). *Cajanus* DC. and *Atylosia* W. & A. (Leguminosae). *Agricultural University Wageningen Papers* 85-4 (Wageningen: Wageningen University Press).
- van der Maesen, L.J.G. (1989). "*Cajanus cajan* (L.) Millsp.," in: *Plant Resources of South-East Asia No 1. Pulses*, eds. L.J.G. van der Maesen, S. Somaatmadja (Wageningen: Pudoc/Prosea), 39-42.
- van der Wal, J., Falconi, L., Januchowski, S., Shoo, L., Storlie, C. (2014). *SDMTools: species distribution modelling tools: tools for processing data associated with species distribution modelling exercises*. R package version 1.1-20. <http://CRAN.R-project.org/package=SDMTools>.
- Varshney, R.K., Chen, W., Li, Y., Bharti, A.K., Saxena, R.K., Schlueter, J.A., *et al.* (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30 (1), 83-89.
- Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., *et al.* (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265-275.
- Wanjari, K.B., Patil, A.N., Manapure, P., Manjappa, J.G., Manish, P. (1999). Cytoplasmic male sterility in pigeonpea with cytoplasm from *Cajanus volubilis*. *Annals of Plant Physiology* 13, 170-174.
- West, P.C., Gerber, J.S., Engstrom, P.M., Mueller, N.D. Brauman, K.A., Carlson, K.M., *et al.* (2014). Leverage points for improving global food security and the environment. *Science* 345 (6194), 325-328.
- Wickham, H. (2009). *Ggplot2: Elegant Graphics for Data Analysis* (New York: Springer).
- Williams, J.N., Seo, C., Thorne, J., Nelson, J.K., Erwin, S., O'Brian, J.M., *et al.* (2009). Using species distribution models to predict new occurrences for rare plants. *Divers. Distrib.* 15 (4), 565-576.
- Xiao, J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D., Li, J., *et al.* (1996). Genes from wild rice improve yield. *Nature* 384, 223-224.

CHAPTER 7

Interdependence among countries on crop genetic diversity

Colin K. Khoury^{1,2}, Harold A. Achicanoy¹, Anne D. Bjorkman³, Carlos Navarro-Racines^{1,4},
Luigi Guarino⁵, Ximena Flores-Palacios⁶, Johannes M. M. Engels⁷, John H. Wiersema⁸,
Hannes Dempewolf⁵, Julian Ramírez-Villegas^{1,4,9}, Nora P. Castañeda-Álvarez^{1,10},
Cary Fowler⁵, Andy Jarvis^{1,4}, Loren H. Rieseberg^{11,12,13}, and Paul C. Struik²

¹ International Center for Tropical Agriculture (CIAT), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, 763537 Cali, Colombia.

² Centre for Crop Systems Analysis, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands.

³ German Centre for Integrative Biodiversity Research, Leipzig, Germany.

⁴ CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Km 17, Recta Cali-Palmira, Apartado Aéreo 6713, 763537 Cali, Colombia.

⁵ Global Crop Diversity Trust, Platz der Vereinten Nationen 7, 53115 Bonn, Germany. ⁶Auckland University of Technology, 55 Wellesley St E, Auckland 1010, New Zealand. ⁷Bioversity International, Via dei Tre Denari, 472/a, 00057 Maccarese, Italy.

⁸ United States Department of Agriculture, Agricultural Research Service, National Germplasm Research Laboratory, Building 003, BARC-West, Beltsville, Maryland 20705-2350, USA.

⁹ School of Earth and Environment, University of Leeds, Leeds, UK.

¹⁰School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. ¹¹The Biodiversity Research Centre, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.

¹²Department of Botany, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada.

¹³Department of Biology, Indiana University, Bloomington, Indiana, 47405, USA.

Submitted for publication in an international scientific journal

Published as Research Study 8, International Treaty on Plant Genetic Resources for Food and Agriculture (Rome: FAO). A modified version of this chapter with the title “Estimation of countries’ interdependence in plant genetic resources provisioning national food supplies and production systems” was published by the Secretariat of the International Treaty on Plant Genetic Resources for Food and Agriculture following a call by the Secretariat and its Ad Hoc Open-Ended Working Group to Enhance the Functioning of the Multilateral System of Access and Benefit-Sharing, to encourage international organizations to make available information relevant to an analysis of the factors that influence the willingness of stakeholder groups to make contributions to the Benefit-sharing Fund and to access plant genetic resources for food and agriculture from the Multilateral System.

Abstract

The crops that feed the world originated in specific geographic regions across the planet. Genetic diversity within these crops and their wild relatives is considered to be historically particularly rich within these regions. These genetic resources are regularly employed in crop improvement: thus, preventing erosion of remnant genetic diversity occurring *in situ*, and ensuring the long-term access to this diversity conserved in genebanks *ex situ*, are critical to continued increases in agricultural productivity. The geopolitical significance of the geography of crop genetic diversity has not been quantified. Here we assess the degree to which the food supplies and production systems of countries worldwide are comprised of crops from each of these regions of diversity. We then examine dependence of countries upon crops from regions of diversity other than their own (“foreign crops”), and determine change in this dependence over the past 50 years. National food systems are thoroughly interconnected worldwide in regard to the geographic origins of crop diversity. Countries are highly dependent on foreign crops in their food supplies (68.7% as a global mean across food variables) and in their national production systems (69.3%). This reliance is evident even in countries located in regions of high indigenous crop diversity and has increased significantly over the past half century, stressing the need for effective national and international policies to promote genetic resource conservation and exchange.

Keywords: Food security, Crop diversity, Crop origins, Interdependence, Plant genetic resources

INTRODUCTION

As the source of traits employed for adaptation to biotic and abiotic stresses and for yield increases through breeding, as well as the palette from which food systems may be further diversified, crop genetic diversity represents a cornerstone of food security (Xiao *et al.* 1996, Hoisington *et al.* 1999, Zhu *et al.* 2000, Gepts 2006, Guarino & Lobell 2011). The need for utilization of this diversity to maintain or enhance crop productivity is likely only to grow given rising food demand and increasing constraints on the use of non-renewable agricultural inputs, limitations in further expansion of arable lands, soil degradation, and global climatic change (Lobell *et al.* 2008, Cordell *et al.* 2009, Kearney 2010, Kastner *et al.* 2012).

Crop genetic diversity is generated through genetic mutation and recombination, and further transformed through natural and artificial selection. Therefore, high levels of crop genetic diversity are associated with the persistence of crops and their wild relatives in specific geographic regions worldwide. A century ago, N. I. Vavilov described these “centers of origin”, characterized by notably high levels of inter- and intra-specific variability in food crops. These included Central America and Mexico; the Andes, Chile and Brazil-Paraguay; the Mediterranean; the Near East; Ethiopia; Central Asia; India; China; and Indo- Malaysia (Vavilov 1926, 1951, 1992). Since then, the number and boundaries of these centers have been investigated and refined (Harlan 1951, Zhukovsky 1968, Harlan 1971, Zeven & Zhukovsky 1975, Zeven & de Wet 1982). Here we use the term “primary regions of diversity” to describe these areas, which typically include the general geographic locations of the initial domestication of crops, encompass major geographic zones of crop genetic diversity generated since that time, and generally also include high species richness in related wild taxa.

New forms and combinations of crop genetic diversity may arise wherever farmers plant, harvest and re-sow their seed (Harlan 1975, Nuijten *et al.* 2009). The spread of crops outside their centers of origin and their increasing contribution to the diets of diverse cultures in different regions (Khoury *et al.* 2014) have therefore led to the development of “secondary centers of diversity”, as well as novel genotypes arising outside of any such defined centers. While all crop genetic diversity, regardless of geographic distribution, is of potential value to crop improvement, the diversity generated over time in the primary regions is considered to be the centerpiece of historical, current, and future crop improvement efforts due to its comparatively high values at allelic, genotypic, and species levels (Vavilov 1926, Harlan 1971,

Hoisington *et al.* 1999, Gepts 2006).

Severe erosion of crop genetic diversity is considered to have occurred over the past half century, particularly through the adoption of improved crop varieties or substitute crop species and subsequent neglect of traditional varieties and crops, economic development and associated shifts in consumer demand, land use change and habitat destruction, and urbanization and the displacement of cultures associated with particular crops and varieties (Hoisington *et al.* 1999, Gepts 2006, Wilkes 2007, van de Wouw *et al.* 2009, 2010). In some crops, a small fraction of the diversity once present is thought to still be found today in farmers' fields, e.g., in wheat varieties in parts of the Fertile Crescent (Harlan 1971, Gepts 2006). Due to the loss of variation in regions of diversity, the world's genebanks originally established to make plant genetic resources readily available to breeders for crop improvement, have become essential repositories for crop diversity conservation. A substantial portion of the world's remaining heritage of food crop genetic resources is likely now conserved exclusively in genebanks (Hoisington *et al.* 1999, Gepts 2006, FAO 2010, Thormann *et al.* 2015).

Thus, the long-term productivity and resilience of food systems depend upon the conservation and use of crop genetic diversity, accessed either directly from regions of diversity or via genebanks (Hoisington *et al.* 1999, McCouch *et al.* 2013). Unfortunately, significant gaps remain in the conservation of crop genetic resources globally (Hoisington *et al.* 1999, Gepts 2006, FAO 2010) and international access to this diversity requires improvement (Fowler & Hodgkin 2004, Bjørnstad *et al.* 2013). These deficiencies may be due, in part, to lack of information on the significance to modern food systems of the historical geographic distributions of crop genetic diversity. Here we determine the degree to which the national food supplies (measured in calories, protein, fat, and food weight) and national production systems (measured in production quantity, harvested area, and production value) of countries worldwide are comprised of crops from all the different primary regions of genetic diversity. In order to assess the level to which international collaboration is required in order to achieve access to crop genetic resources, we estimate the degree of dependence of countries upon crops from primary regions of diversity other than their own (i.e., "foreign" crops), and determine change in this dependence over the past 50 years.

MATERIALS AND METHODS

We used food supplies and production systems data provided by FAO (2015) [for food supplies: calories (kcal/capita/day), protein (g/capita/day), fat (g/capita/day), and food weight (g/capita/day); for production systems: production quantity (tonnes), harvested area (ha), and gross production value (million US\$)]. National food supply from plants represents national production plus imports plus or minus stock changes over the survey period; minus exports, quantities used for seed, animal feed, and in the manufacture of non-food products, and losses during storage and transport (Khoury *et al.* 2014). We utilized the full set of food crop commodities included in food supply and production data. While food supplies data accounts for direct human consumption, production data for crops such as maize and soybean is potentially inclusive of livestock and industrial uses as well as human food. In the production analysis we also included agricultural crops indirectly contributing to human food supplies via livestock production (i.e., alfalfa, clover, and vetch). Non-food (e.g., industrial and fiber) crops as well as animal product commodities were not included in the analysis. Plant commodities comprised of the same crop species were aggregated into single commodities representing the crop, e.g., sesame seed oil and sesame seed. After aggregation, 53 crop commodities remained in food supplies data, and 132 crop commodities in production data (Supplementary Table 1). See Table S1 of Khoury *et al.* (2014) for a comprehensive listing of the crop species included in the commodities treated in food supplies data.

For current food supplies and production systems, we analyzed data for each crop commodity per country per measurement over the most recent three years for which sufficient data were available (2009-2011). All (177) countries consistently reported during the time period were included for food supplies variables, as well as for production quantity and harvested area (Supplementary Table 2), covering 98.5% of the world's population. All (141) countries reported for (current million US\$) production value were included, covering 94.1% of the world's population (FAO 2015).

For the analysis of change in dependence over time, food supplies data were assessed for each year from 1961-2009, and production systems from 1961-2011, utilizing the full set of commodity and country listings, standardized across all years. In order to align all time periods and include as much of the world's population as possible, the current countries formerly comprising the USSR, Yugoslav SFR, Ethiopia PDR, and Czechoslovakia were aggregated into

their former countries, with national data summed per year for production measurements, and merged by weighted average based upon the population of the respective states during the respective reporting year for per capita food supplies measurements. Belgium and Luxembourg were reported together during 1961-1999 and therefore recent years listing the countries separately were merged as above. Countries that did not have estimates in every year between 1961 and 2009/2011 were removed from the analysis. The resulting 152 comparable countries treated in food supplies data comprised 98% of the world's population across the study period (Khoury *et al.* 2014). The 182 comparable countries covered in production quantity and harvested area data comprised 99.7% of the global population, and the 115 countries covered in (constant 2004-2006 million US\$) production value data covered 88.5% (Supplementary Table 2).

Primary regions of diversity were assigned based upon primary and secondary literature regarding centres of crop diversity, origins of crop domestication, and high species richness of closely related wild plants (Harlan 1951, Zhukovsky 1968, Harlan 1971, Zeven & Zhukovsky 1975, Zeven & de Wet 1982, Flores-Palacios 1998, Engels *et al.* 2001, Vincent *et al.* 2013, GRIN 2014, Prota 2014). Regional classifications followed those listed in Annex 2 of the FAO State of the World's Plant Genetic Resources for Food and Agriculture (2010), modified to more accurately represent eco-geographic parameters driving plant species distributions. Specifically, both western and eastern Europe were split into north and south regions to account for temperate versus Mediterranean ecologies; Australia and New Zealand were segregated from remaining (tropical) islands of the Pacific region; and South America was split into Andean, temperate, and tropical regions. A total of 23 eco-geographic regions were delineated (Supplementary Figure 1). Countries whose boundaries included more than one eco-geographic region were included in all appropriate regions (e.g., Colombia was assigned both to Andean and to tropical South American regions) (Supplementary Table 2).

Crops whose primary areas of diversity encompassed more than one eco-geographic region were listed in all appropriate regions (e.g., wheat was listed in Central Asia, West Asia, and the South and East Mediterranean). Forty-two of the 53 crop commodities treated in food supplies data, and 116 of the 132 crops in production data, were capable of being attributed to primary regions of diversity, with the remaining general commodities which were not clearly attributable to specific crop species listed as "not specified" (Supplementary Table 1).

We constructed circular plots displaying the relative importance of primary regions of diversity as sources of crops comprising current (2009-2011 average) national food supplies and production systems, using methods and code adapted from Abel and Sander (2014). For recipient data, regional food supply values (kcal or g, /capita/day) were formed per variable by deriving a weighted average across countries comprising each region, with national values weighted by population. Regional production values were calculated by summing values across countries for each variable.

We estimated the degree to which a country's food supplies and production systems are dependent upon crops of "foreign" primary regions of diversity by determining the extent to which such supplies/systems are composed of crops whose primary regions of diversity do not coincide with the regions within which that country is located. The method was initiated with the assumption that the primary diversity of crops within a given country's food supplies/production systems was completely foreign (100% dependence). The percent contribution of all crops whose primary diversity was identified as in the same region as the country was then subtracted to estimate a "maximum dependence" metric per country. In this metric, those general crop commodities whose regions could not be specified were assumed to be of foreign primary regions of crop diversity. The sum of the percent contribution of these non-specified general crop commodities was then subtracted, resulting in a "minimum dependence" metric which assumes that all non-specified crop commodities possess primary diversity within the same region as the country (modified from Flores-Palacios 1998).

Mean dependence in food supplies and production systems per country was estimated using an interval censoring method, where the response variable (the calculated dependence value in each country in each year) was bounded between the minimum and maximum dependence estimates for each observation. A model of this type allows the uncertainty around an observation to be incorporated into the parameter estimates for the parameter of interest. For estimates of current dependence, we modeled the mean of the most recent three years (2009-2011). For estimates of change in dependence from 1961-2009/2011, intercepts and slopes per country were modeled as random effects, where the mean hyper-parameter for the random slopes represented the estimated slope (change in dependence over time) across all countries. We allowed a correlation between country-level intercepts and slopes to account for the fact that countries with high dependence have weaker dependence-time relationships than countries

with low dependence (Gelman & Hill 2007). The interval-censored models were implemented using a Bayesian framework in JAGS (v. 3.4.0) called from R (v.3.1.1), using the packages *rjags* and *R2jags*. Non-informative (“flat”) priors were used for all coefficients. Convergence was assessed using the Gelman-Rubin diagnostic (Gelman & Rubin 1992) and by visual inspection of trace plots. Dependency values reported in the text represent the model-estimated coefficient, \pm the standard deviation. Credible intervals for each parameter are reported in Supplementary Table 3-5.

We used Simpson’s diversity index (2009-2011 mean) to correlate the degree of contributing crop diversity in current national food supplies/production systems with dependence on foreign primary regions of diversity. The diversity-dependence relationship was modeled using a simple linear model with both linear and quadratic terms. Diversity analyses were performed using the *vegan* package in R (v. 3.1.1). We also correlated dependence with national Gross Domestic Product (GDP) per capita purchasing power parity, using a mean GDP value across 2009-2011 for 169 available countries (World Bank 2014).

Crops were assigned importance individually for each food supplies and production systems variable into 10% quantiles, from 1 (low importance) to 10 (high importance), based upon their global aggregate (food supplies) and total global production values. A combined assessment was performed on (136) unique crop commodities covered in food supplies and production systems data (Supplementary Table 6). Thirty-seven of these commodities possessed both food supplies and production systems data and were directly compared. An additional 92 crop commodities with production systems values were embedded within 12 general commodities in food supplies data (i.e., cereals, other; fruits, other; oilcrops, other; oranges & mandarines; pulses, other; rape & mustard; roots, other; spices, other; sugar; tea; treenuts; and vegetables, other). Food supplies values for most of the individual commodities were estimated by dividing their total general commodity values equally across listed crops. For the sugar commodity, sugarcane was assigned 70% and sugar beet 30% of the total value; for the tea commodity, tea [*Camellia sinensis* (L.) Kuntze] was assigned 80%, mate 10%, and “not elsewhere specified” (nes) tea 10%. Three additional production systems crop commodities (alfalfa, clover, and vetches), which are livestock feed/forage crops and therefore are not recorded in food supplies data, were assessed through quantile values derived solely from production systems variables. Four general food supplies commodities (beverages, alcoholic; beverages, fermented;

miscellaneous; and sweeteners, other) were not recorded in production systems variables, thus these commodities were assessed through quantile values derived solely from food supplies variables. Coverage of each crop in the Multilateral System (MLS; i.e., Annex 1) of the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO 2002) was assessed, listing crops as covered, partially covered (often in the case of general crop commodities, in which some portion of the crops within the commodity are covered in the MLS and others not), or not covered (Supplementary Table 6). The extent of geographic importance of crops was additionally documented by counting the number of countries listing each commodity (>0) for each variable, as well as listing the plant commodities by decreasing importance until the total contribution equaled 90% of each country's food supply/ production for each variable, a threshold which is inclusive of major contributors to supply/ production systems and exclusive of commodities contributing very small quantities (Prescott-Allen & Prescott-Allen 1990, Khoury *et al.* 2014). The total count of countries including each crop commodity as important was then derived per crop commodity.

RESULTS

Primary regions of genetic diversity of agricultural crops were identified across the tropics and subtropics, extending into temperate regions in both hemispheres (Figure 1, Supplementary Figure 2). The food supplies and production systems of countries worldwide were found to be comprised of a wide range of crops from diverse geographic backgrounds, indicating a thoroughly interconnected global food system in regard to the geographic origins of crop genetic diversity (Figure 2, Supplementary Figure 3). The evident widespread importance in global food supplies particularly of major crops such as wheat, rice, sugarcane, maize, soybean, potatoes, barley, oil palm, beans, tomatoes, bananas & plantains, and sugarbeet, among others, lead to particular significance of key primary regions of diversity, including West, Central, South, Southeast, and East Asia, the South and East Mediterranean, West and Central Africa, Central America and Mexico, Andean and tropical South America, and southern Europe (Figure 2, Supplementary Figure 3-4). Cassava, rape & mustard, groundnut, grapes, apples, alfalfa, sorghum, and millets were among other crops of particular international importance for one or more food supply and/or production variables.

Dependence upon crops from foreign primary regions of genetic diversity in national food supplies and production systems was highest (i.e., up to 100%) in those countries

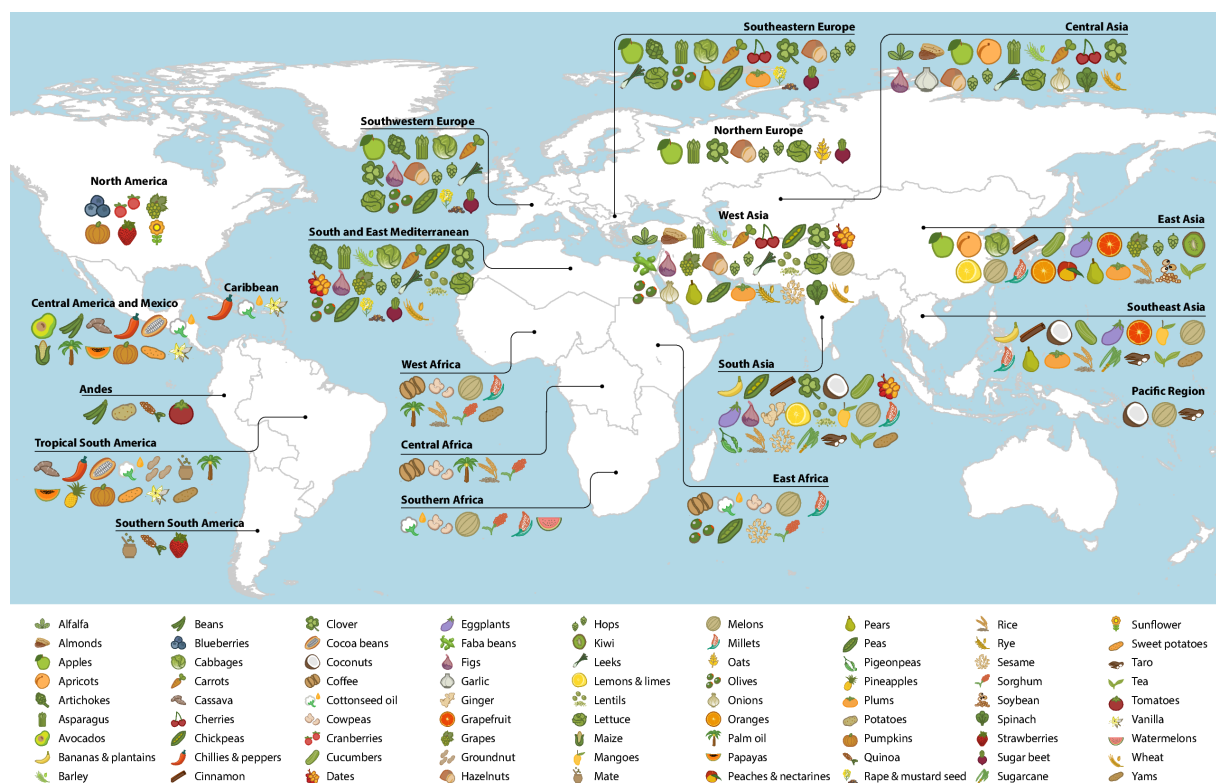


Figure 1. Primary regions of diversity of major agricultural crops worldwide. See Supplementary Table 1 for a list of primary regions for all assessed crop commodities.

geographically isolated from and/or located at great distance from the primary regions of diversity of major staple crops (Figure 3, Supplementary Figure 5, Supplementary Table 3). This includes Australia and New Zealand, the Indian Ocean Islands, the Caribbean, southern South America, North America, southern Africa, and northern Europe. These countries are generally in temperate climates, although tropical islands and some continental tropical regions, such as Central Africa, also demonstrated very high levels of dependence for most variables.

Conversely, dependence upon foreign crops was lowest in countries located within the primary regions of diversity of major crops, and where traditional staples are still cultivated and consumed, such as Southeast Asia, the South and East Mediterranean, South Asia, Central Asia, West Asia, and West Africa. The lowest levels of dependence were found in countries with food systems dominated by a limited number of traditional staples such as rice, wheat, yams, sorghum, and millets (see Supplementary Figure 6). Island nations predominantly dependent upon native crops for fat, such as coconut in the tropical Pacific Region, and countries with extreme agroecological conditions limiting national production to the cultivation of a select number of native crops (e.g., dates in the United Arab Emirates and other arid nations of West Asia) also exhibited very low levels of dependence for relevant food supply

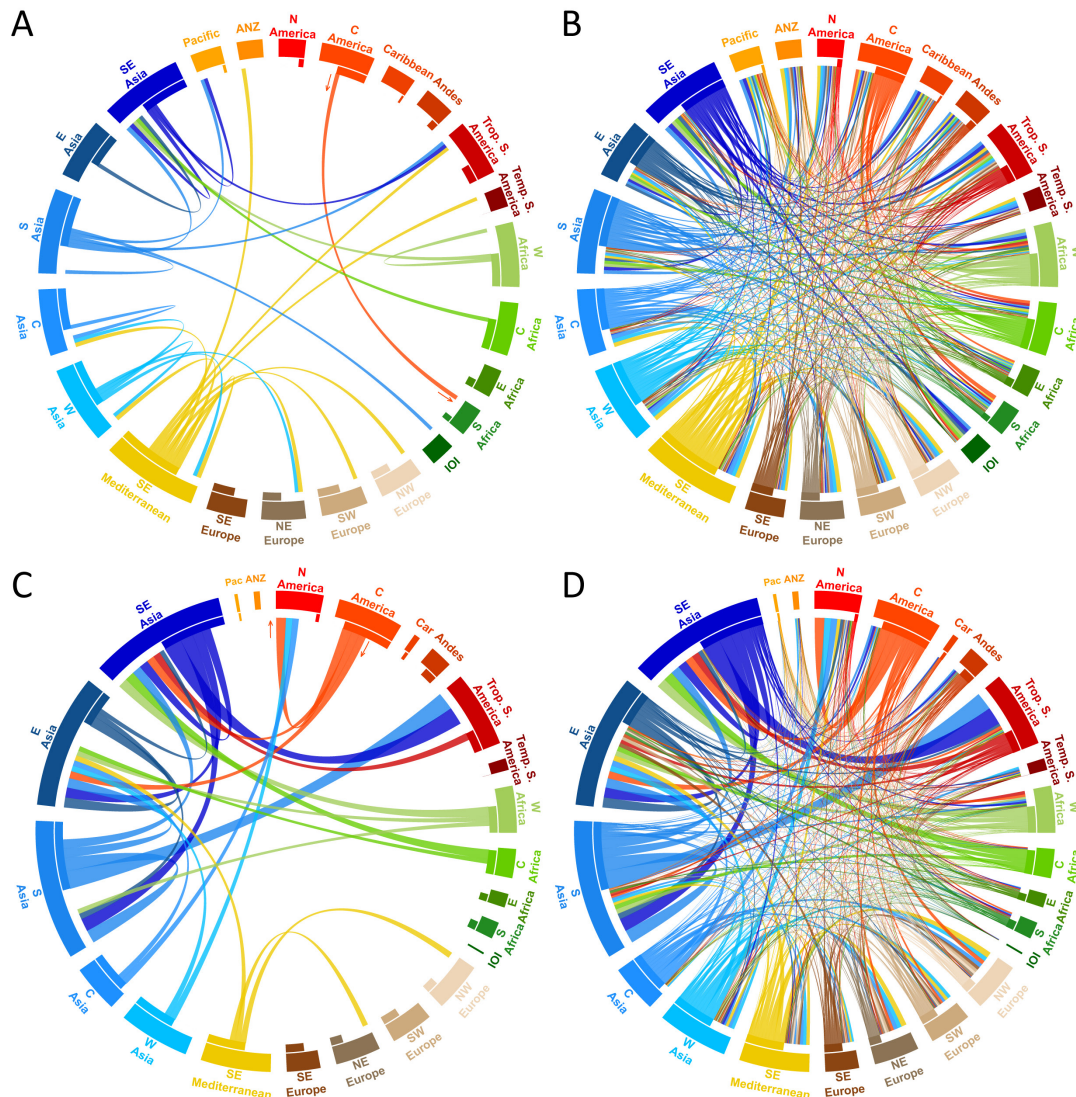


Figure 2. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (A-B) calories in national food supplies, and (C-D) production quantity in national production systems, averaged over 2009-2011. For recipient data, regional caloric food supply values (kcal/capita/day) were formed by deriving a weighted average across countries comprising each region, with national values weighted by population. Regional production quantity values were formed by summing values across countries. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, and C. America = Central America and Mexico; and in production quantity only, Car = Caribbean, and Pac = Tropical Pacific Region. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. Arrows indicating direction of contribution are also included as examples. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (A, C) display only the most significant contributions (i.e., 95th percentile) for visibility, which correspond to an approximate contribution of at least 1180 kcal/cap/day (A), and 212 million tonnes (C). In these figures, the importance of rice, wheat, maize, sugarcane, and oil palm are evident. Figures on the right (B, D) display all contributions. See Supplementary Figure 3 for circular plots for all measured food supply and production variables.

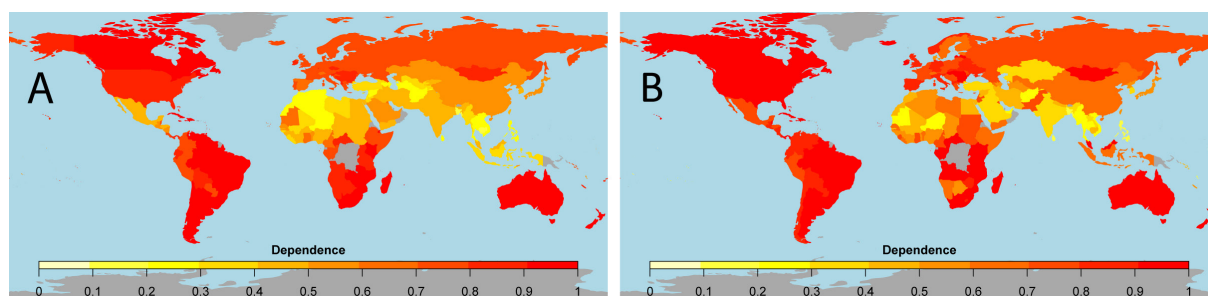


Figure 3. Dependence on foreign primary regions of diversity of crops per country in regard to (A) calories in national food supplies, and (B) production quantity in national production systems, as a modeled mean between minimum and maximum dependence per country, 2009-2011. Dependence scale is degree of dependence (1 = completely dependent). See Supplementary Figure 5 for world maps displaying dependence per country for all measured food supply and production variables.

and/or production metrics. In such extreme cases, though, low dependence was generally evident in only one or a few food supplies or production metrics, while other variables exhibited much higher dependence.

Although food supplies and production systems variables were highly correlated in degree of dependence (Supplementary Figure 7-8), variation was also visible across variables, with highest overall dependence evident in fat, production value, production quantity, and food weight (Supplementary Table 3). Considerable variation in dependence was found within geographic areas, e.g., ranging from (mean \pm SD) 48.2% \pm 1.6 for calories in Mexico, to 86.5% \pm 1.4 in Panama, within the Central America and Mexico region. Large variation in dependence in production systems was also found within regions, e.g., ranging from 25.7% \pm 3.5 for production value in the Philippines, to 94.1% \pm 1.1 for Malaysia, within the Southeast Asia region. Countries with very high dependence for such production variables were exemplified by the presence of extensive production systems dedicated to a limited number of high value, foreign commodity crops, such as oil palm in Malaysia.

Dependence upon crops from foreign primary regions of genetic diversity was positively correlated with diverse food supplies/production systems, although high dependence also occurred in numerous countries with exhibiting low diversity (Supplementary Figure 6). Very few countries, on the other hand, showed high diversity in their food supplies and/or production systems and at the same time low dependence on crops of foreign primary regions of diversity. National Gross Domestic Product was also associated with dependence, although with considerable variation worldwide (Supplementary Figure 9).

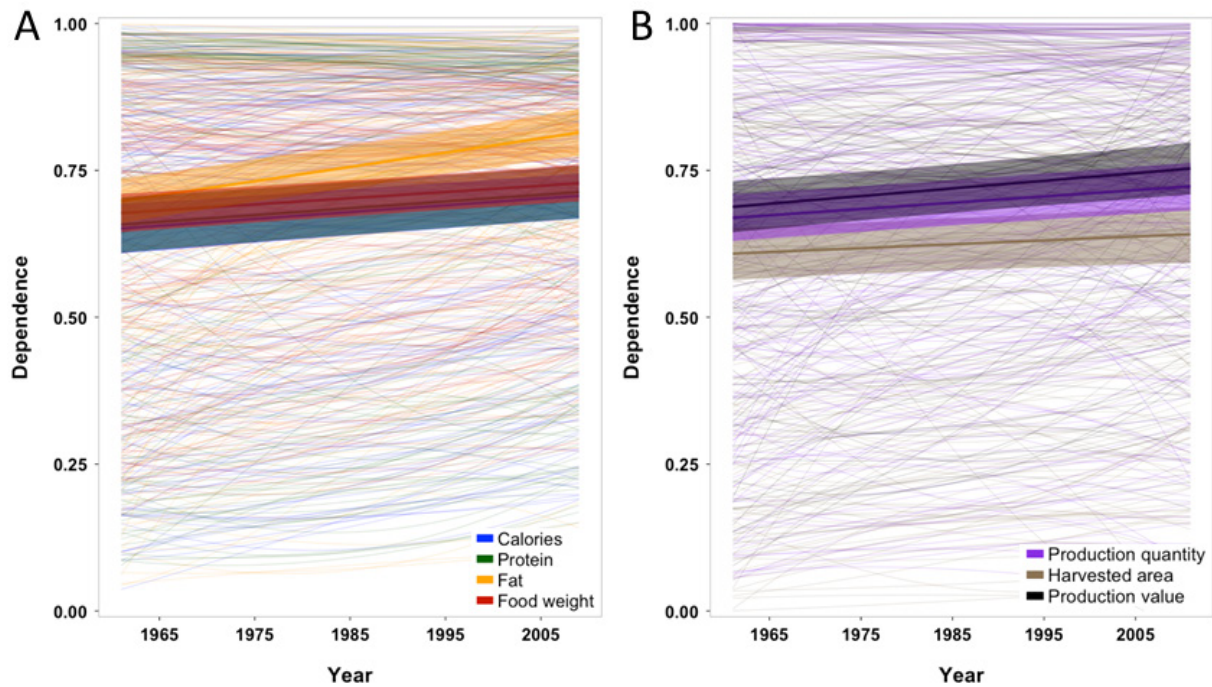


Figure 4. Change in dependence on foreign primary regions of diversity of crops in (A) national food supplies (1961-2009) and (B) production systems (1961-2011). Lines represent change over time in the mean between minimum and maximum dependence for each country in each year for each variable as predicted by a quadratic regression. Transparent ribbons represent modeled mean change across all countries (\pm 95% credible interval) in dependence for each variable, estimated using a Bayesian model with an interval censored response variable bounded between minimum and maximum dependence.

Mean dependence across all countries on crops of foreign primary regions of diversity in food supplies was $65.8\% \pm 1.8$ for calories, $66.6\% \pm 2.1$ for protein, $73.7\% \pm 1.6$ for fat, and $68.6\% \pm 1.4$ for food weight. Mean dependence in production systems was $71.0\% \pm 1.8$ for production quantity, $64.0\% \pm 2.2$ for harvested area, and $72.9\% \pm 1.9$ for production value. The combined mean dependence across food supply variables was estimated at 68.7%, across production systems at 69.3%, and across food systems worldwide (i.e., both food supplies and production systems, across all countries and all variables) at 68.9% (Supplementary Table 3).

National dependence upon crops of foreign primary regions of diversity increased significantly as a global mean for all food supplies and production systems variables over the past half century (Figure 4, Supplementary Table 4-5). Dependence in regard to calories increased from 62.7 to 67.4%, protein from 63.1 to 68.1%, fat from 63.4 to 73.2%, and food weight from 65.2 to 69.7% as measured in change in dependence from the mean of the first three years (1961-1963) to the last three years (2007-2009) per country, averaged across countries worldwide. Likewise, dependence in regard to production quantity increased from 64.1 to 69.0%, harvested

area from 59.0 to 62.1%, and production value from 64.4 to 70.5% between 1961 and 2011.

Countries with the greatest increases in dependence over the period were located in Africa, West, South, Southeast, and East Asia, Central America and Mexico, and Andean and tropical South America (Supplementary Figure 10). A number of countries with the largest changes in dependence upon foreign crops in contribution to their food supplies were also those with major transitions in their production systems during the past 50 years (e.g., the growth of oil palm cultivation in Malaysia and Indonesia, a crop whose primary regions of diversity are located in West and Central Africa and the Neotropics; and soybean in Brazil, a crop of East Asian origin). Most regions also contained countries with decreases in dependence over the period. Growing consumption of major staples within the native regions of these crops, such as soybean in China, or wheat in West Asia, may be a factor in this decrease. Dependence in regard to fat in food supplies increased the greatest degree over the past 50 years among all variables, a trend that is concordant with significant changes in the contributing crop species composition of national food supplies globally over this period (Khoury *et al.* 2014).

DISCUSSION

The food supplies and production systems of countries worldwide are primarily composed of crops that were initially domesticated and largely diversified elsewhere around the world. While geography and climate constrained for some period following the agricultural revolutions ca. 10,000 BP the availability of crops to their regions, growth in migration, colonialism, and trade, among other historical forces (Diamond 2004), increased the availability of crops beyond their primary regions, and current economic and agricultural development, urbanization, and globalization trends are making important food crops comprehensively available worldwide (Khoury *et al.* 2014). Even countries located within the most ancient and richest primary regions of diversity, e.g., West Asia, now exhibit considerable dependence on foreign crops in their food supplies and production systems.

The range of crops covered in this analysis is not fully inclusive of all foodstuffs produced and consumed in national food systems. Therefore, an underestimation and/or overgeneralization of diversity is assumed, particularly in regard to plants primarily encountered in home gardens and local markets, seasonally important foods, and culinary herbs, spices and other crops consumed in relatively small quantities (Prescott-Allen & Prescott-Allen 1990). Some of these

crops may be important to diets, particularly in their contribution of micronutrients (Doughty 1979). The aggregation of some crop commodities, the generality of the defined eco-geographic regions, uncertainty for some crops as to their primary regions of diversity, and the subjective nature of the boundaries of such regions, also lead to a degree of uncertainty in dependence metrics. Acknowledging these limitations, the results are a very strong indication of the extent of globalization of food systems and the resulting interdependence among nations on crop genetic resources.

Complementary approaches to the current study have been taken to assess the degree of interdependence among countries in provisioning the global food system. Analyses of production versus consumption in nations globally have revealed increasing interdependence in regard to trade in food (Fader *et al.* 2013, Porkka *et al.* 2013, D’Odorico *et al.* 2014, MacDonald *et al.* 2015). Investigations of exchange of crop genetic resources in the form of accession requests from international genebanks have also shown expanding transfers of germplasm (Kloppenburger & Kleinman 1987, Dudnik *et al.* 2001, Smale & Day Rubenstein 2002, Day Rubenstein & Smale 2004, Fowler & Hodgkin 2004). The increasingly global contribution of diverse breeding materials to the development of modern crop cultivars has also been documented for major crops (Smale 1996, Gollin 1998, Brennan *et al.* 1999, Zhou *et al.* 2000, Cassaday *et al.* 2001, Fowler *et al.* 2001, Smale *et al.* 2002, Johnston *et al.* 2003). Such studies bolster evidence of increased interdependence among nations in concert with greater globalization of the crops and crop varieties providing our global food supplies.

The importance of continued access to diverse crop genetic resources through international exchange in support of national production, and its corollary impact on national economies, is unequivocal. Yet access to genetic diversity in important crops by major producers, wherever their location, is equally critical for the reliable provisioning of global food supplies via international trade, especially as countries have transitioned from food insecurity to trade dependence (Fader *et al.* 2013, Porkka *et al.* 2013, D’Odorico *et al.* 2014, MacDonald *et al.* 2015). Production of the major crops is unevenly distributed across countries and for many crops now generally occurs outside of the primary regions of diversity of those crops, e.g., China, India, the USA, the Russian Federation, France and Canada for wheat; the USA, China, Germany, France, Brazil, and Argentina for maize; the USA, Brazil, Argentina and India for soybean; and China, India, the Russian Federation, Ukraine, and the USA for potatoes.

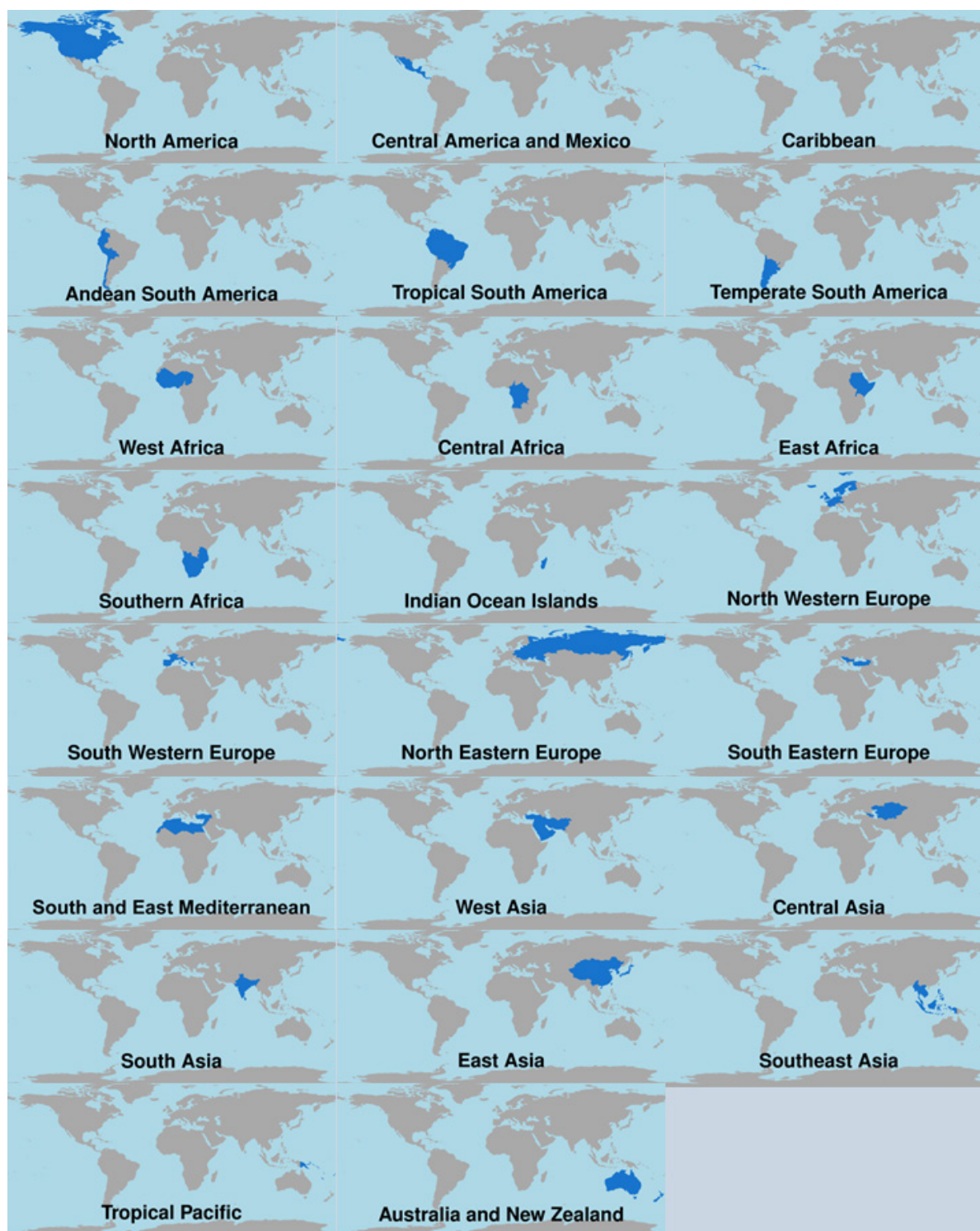
The evidently very high levels of interdependence among countries in regard to crop genetic diversity supports the rationale for internationally coordinated mechanisms to facilitate access to these resources worldwide, such as the Multilateral System (MLS) created within the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty) (FAO 2002). This interdependence also reinforces the importance of the international genebank collections safeguarded as global public goods by the CGIAR international agricultural research centres (Hoisington *et al.* 1999), which are covered under Article 15 of the Plant Treaty. While the long-term sustainability of funding for these collections has partially been achieved (Global Crop Diversity Trust 2013), an increased level of international support will be needed to secure their role in conserving and distributing the genetic resources of their mandate crops. Moreover, large gaps remain in the conservation of crop diversity not covered by CGIAR collections (FAO 2010, Khoury *et al.* 2010). The window of opportunity for securing the world's agricultural diversity threatened *in situ* and in under-funded genebanks will not remain open indefinitely (Wilkes *et al.* 2007, FAO 2010).

A comprehensive MLS should engender facilitated access to the genetic resources of all crops of present and future international importance, but current access is suboptimal (Bjørnstad *et al.* 2013). As food systems undergo further transition due both to dietary change (Kearney 2010, Khoury *et al.* 2014) and to novel production challenges (Lobell *et al.* 2008, Cordell *et al.* 2009, Jarvis *et al.* 2009), a broadly inclusive and adaptable effort to conserve and provide access to agricultural diversity internationally is prudent.

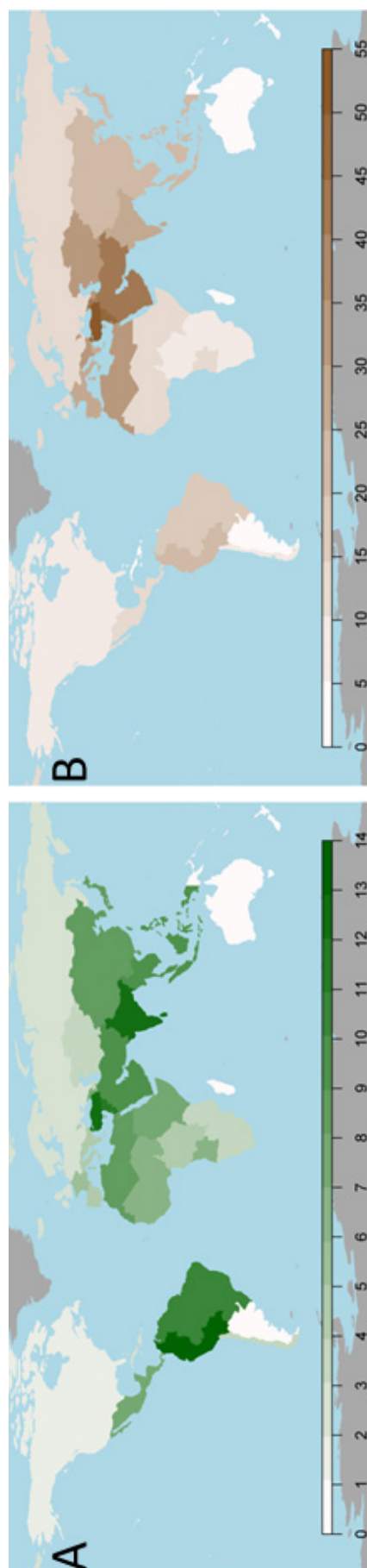
The Plant Treaty MLS has focused mostly on cereal, pulse, starchy root, and forage crops (listed in Annex 1 of the Treaty) (FAO 2002), thus oil crops, vegetables and fruits are not well covered. As much as 28.7% of global aggregate calories in food supplies, 19.0% of protein, 61.0% of fat, 43.4% of food weight; and 41.0% of total global production quantity, 27.0% of harvested area, and 41.2% of global production value are comprised of crops not covered by the MLS. These include soybean, oil palm, sugarcane, groundnut, tomatoes, onions, grapes, coffee, cocoa beans, and a variety of other vegetable, nut and fruit crops (Supplementary Figure 11, Supplementary Table 6). As limited genetic resources research on these crops has been carried out by CGIAR (Khoury & Jarvis 2014), their diversity is likewise underrepresented in the international collections. These crops therefore constitute major gaps in coordinated efforts among nations to conserve and make accessible valuable genetic resources. Furthermore, as

greater variation in genetic materials is increasingly needed for future breeding of major crops (Hoisington *et al.* 1999, Gepts 2006, Burke *et al.* 2009, Jarvis *et al.* 2009, McCouch *et al.* 2013), the expansion of MLS coverage to include wild relatives of staples such as maize and cassava will be important to the strengthening of a global system for the conservation and exchange of crop genetic diversity.

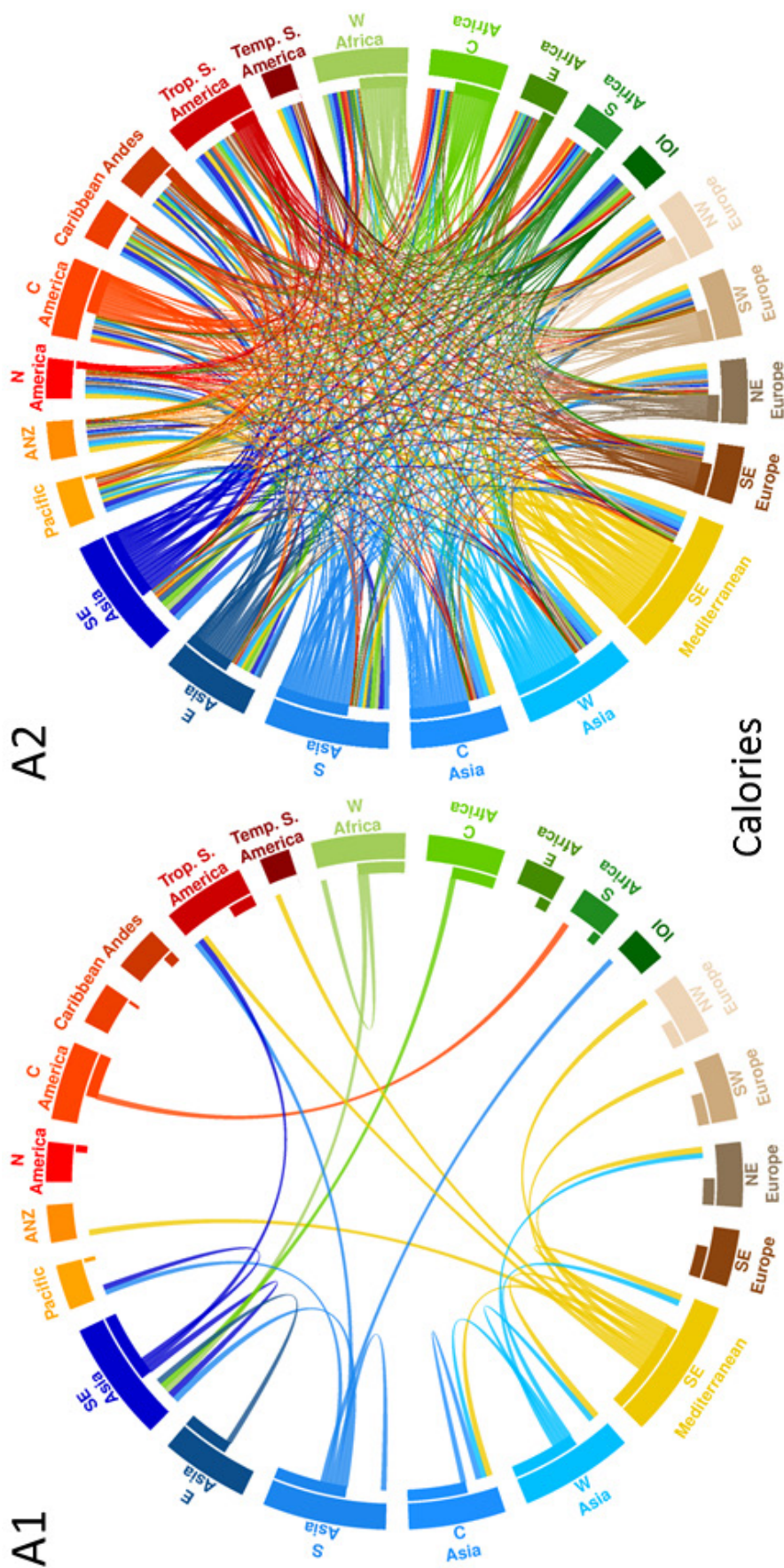
SUPPLEMENTARY INFORMATION



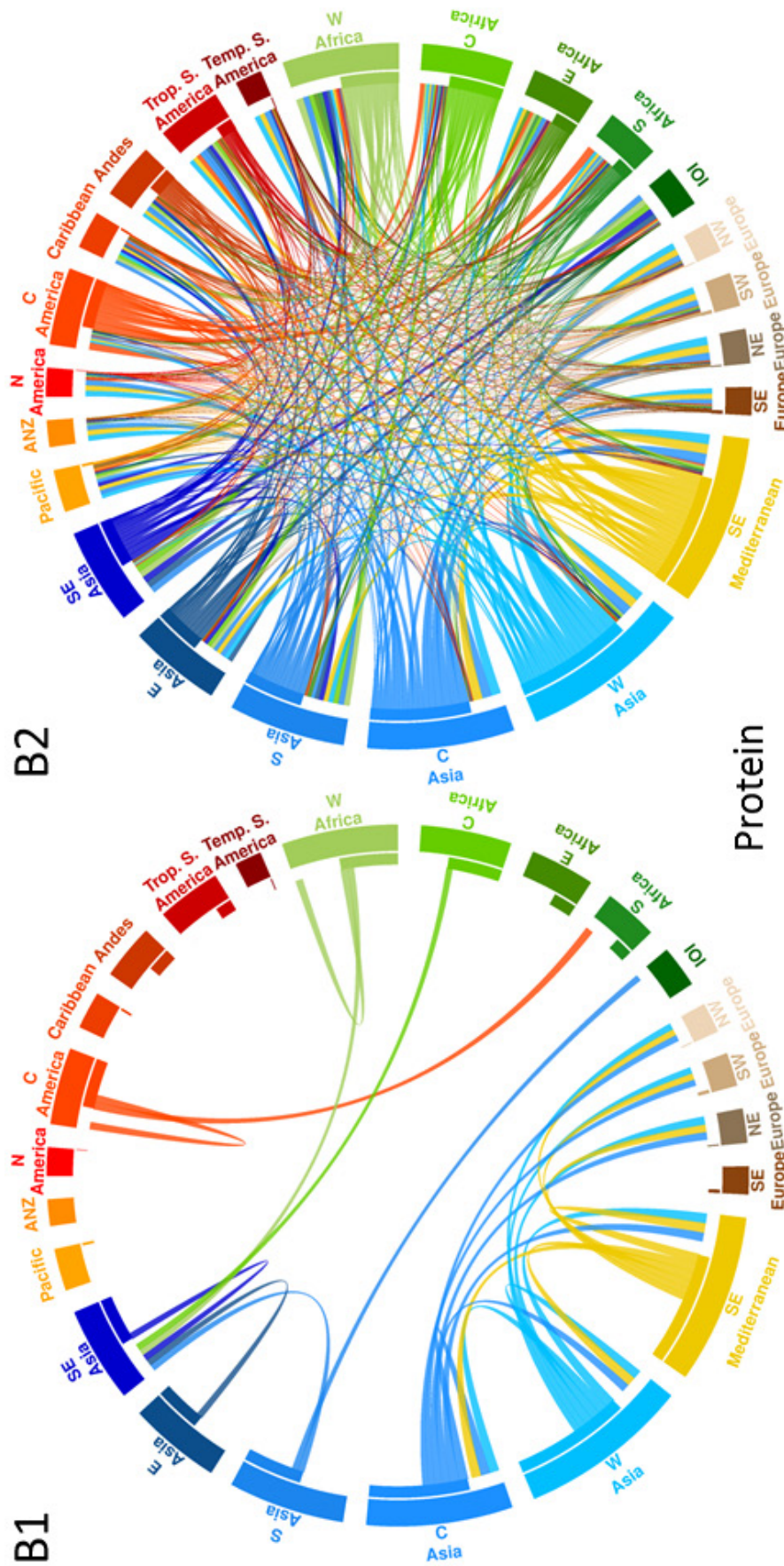
Supplementary Figure 1. Eco-geographic regions utilized in the analysis. See Supplementary Table 2 for a list of countries per region.



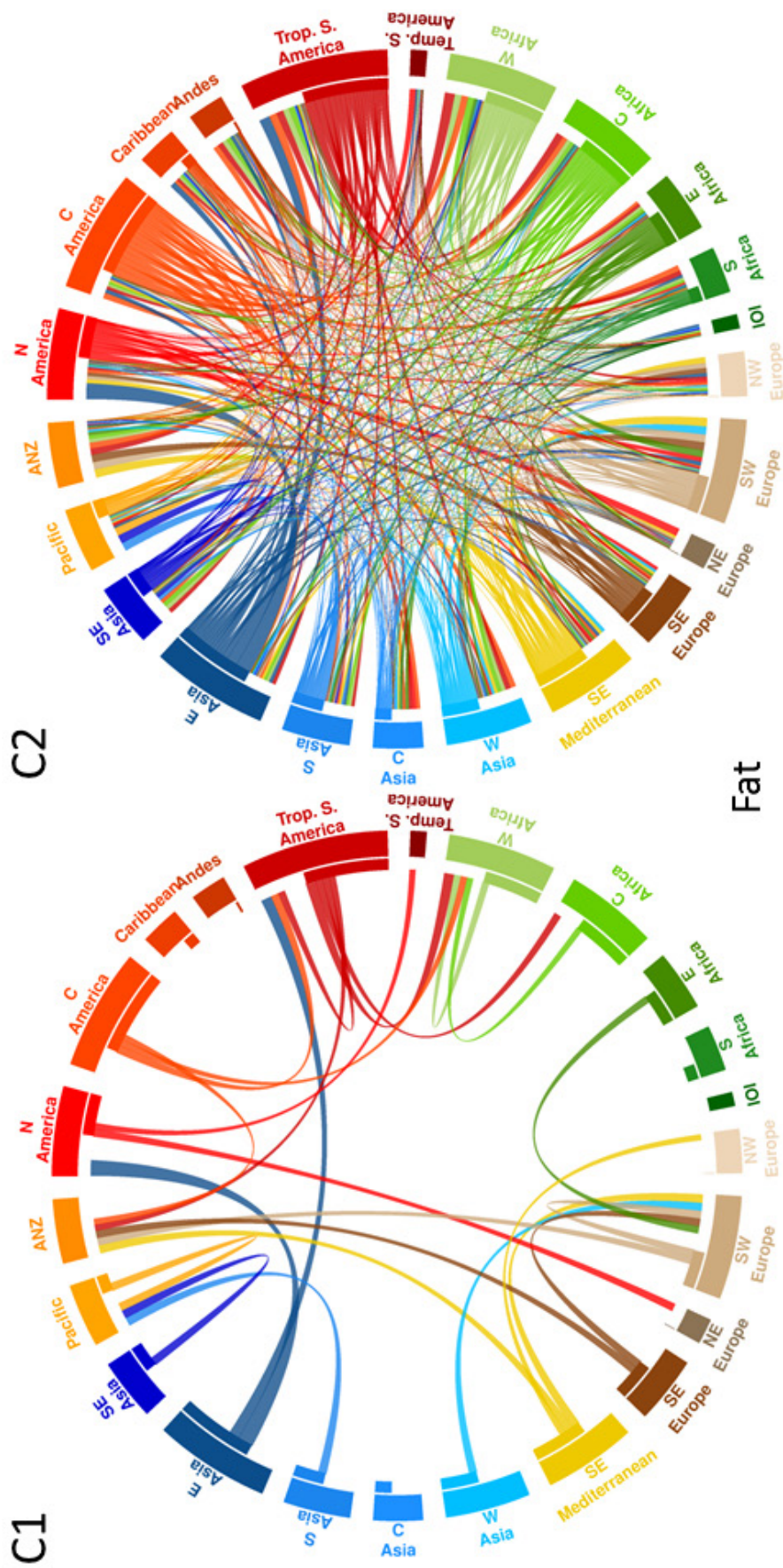
Supplementary Figure 2. Richness maps of primary regions of diversity of crop commodities. (A) Primary regions of diversity of 16 crop commodities assessed in food supplies; and (B) Primary regions of diversity of 42 crop commodities assessed in production systems. Scale is number of crops; darker colors represent regions where numerous primary regions of diversity of crops overlap.



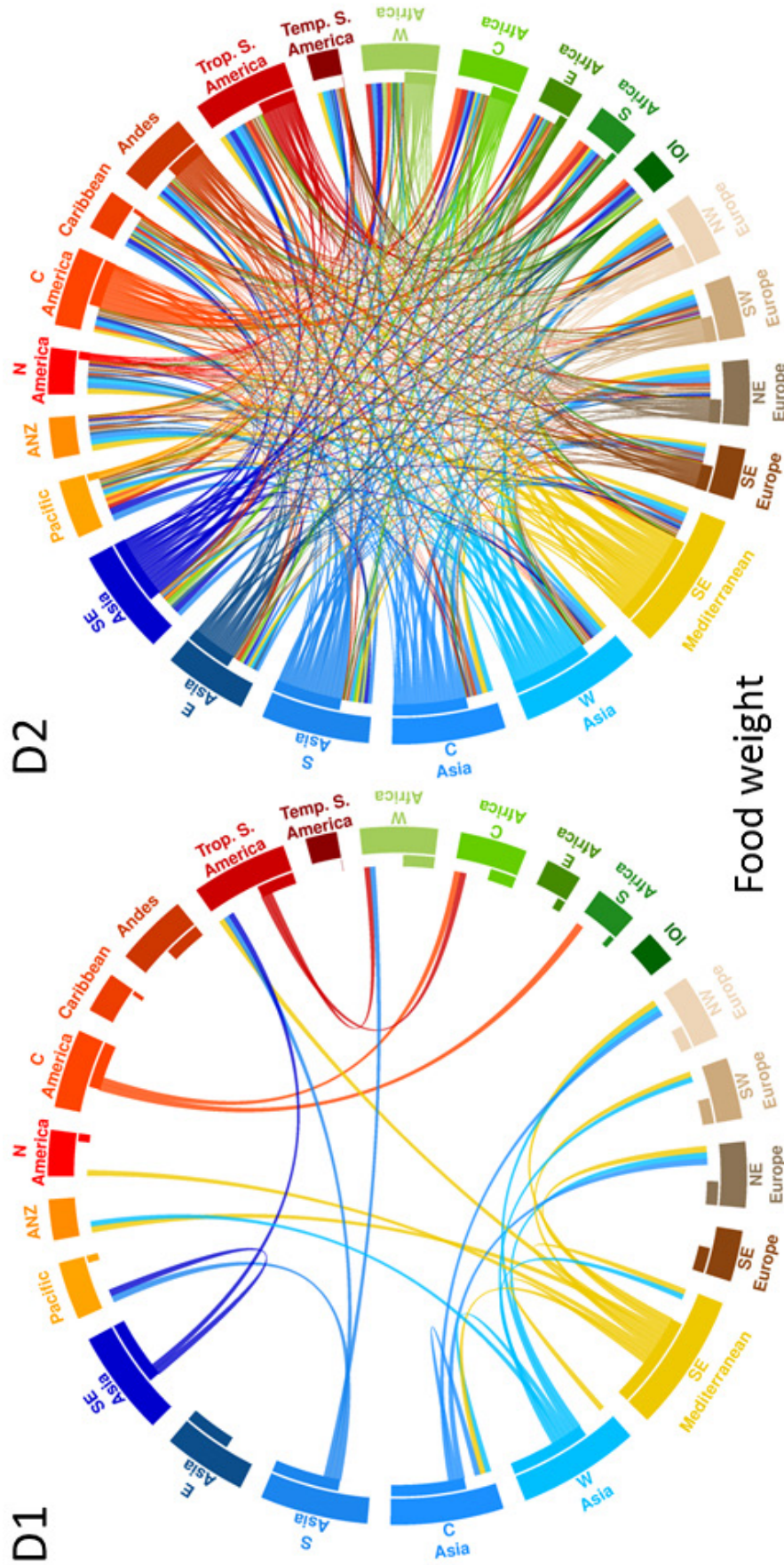
Supplementary Figure 3A. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (A) calories in national food supplies, averaged over 2009–2011. For recipient data, regional food supply values (kcal/capita/day) were formed per variable by deriving a weighted average across countries comprising each region, with national values weighted by population. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, and C. America = Central America and Mexico. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



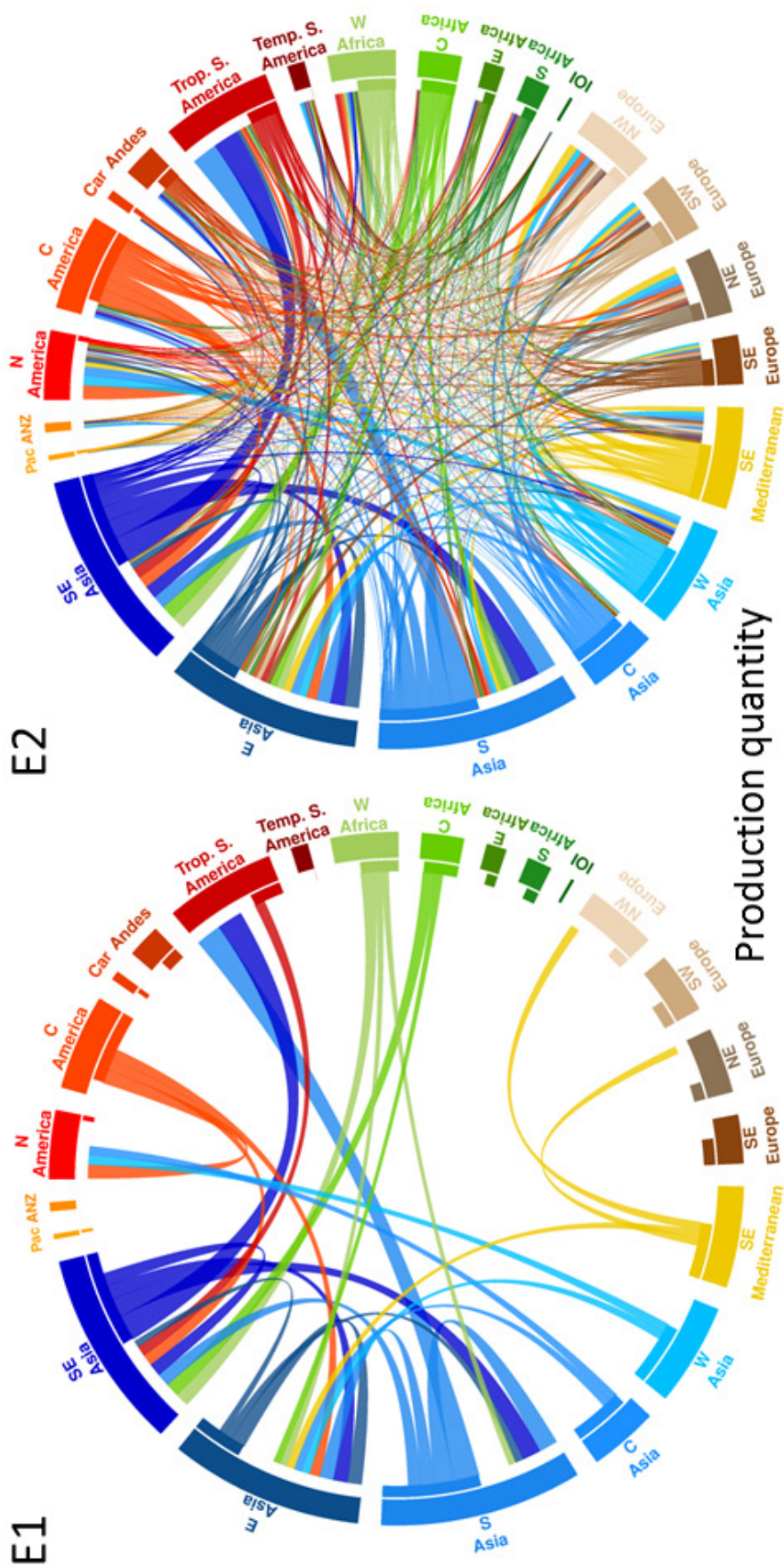
Supplementary Figure 3B. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (B) protein in national food supplies, averaged over 2009-2011. For recipient data, regional food supply values (g/capita/day) were formed per variable by deriving a weighted average across countries comprising each region, with national values weighted by population. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, and C. America = Central America and Mexico. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display all contributions contributions (i.e., 95th percentile) for visibility; figures on the right (2) display only the most significant



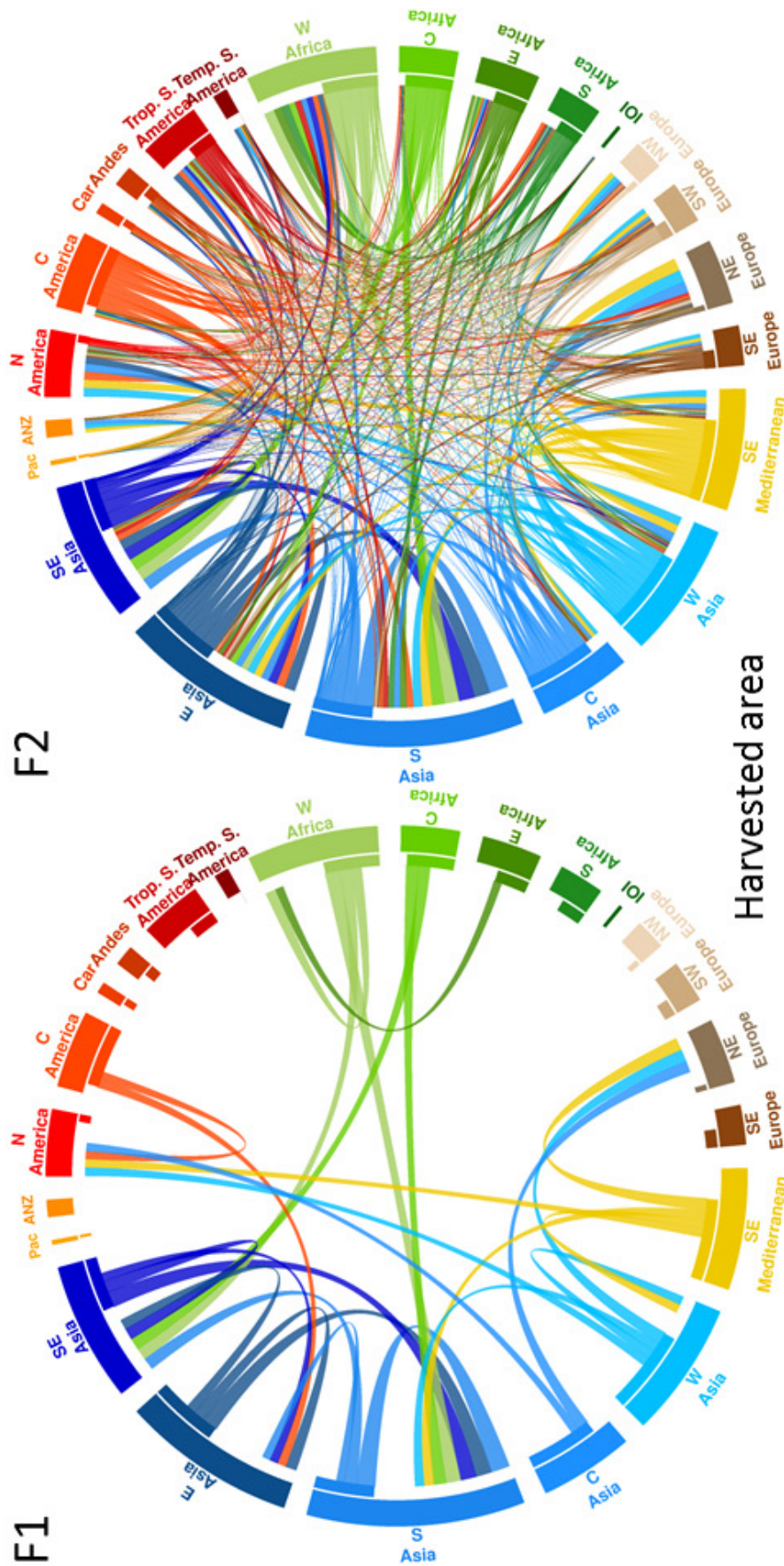
Supplementary Figure 3C. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (C) fat in national food supplies, averaged over 2009-2011. For recipient data, regional food supply values (g/capita/day) were formed per variable by deriving a weighted average across countries comprising each region, with national values weighted by population. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, and C. America = Central America and Mexico. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



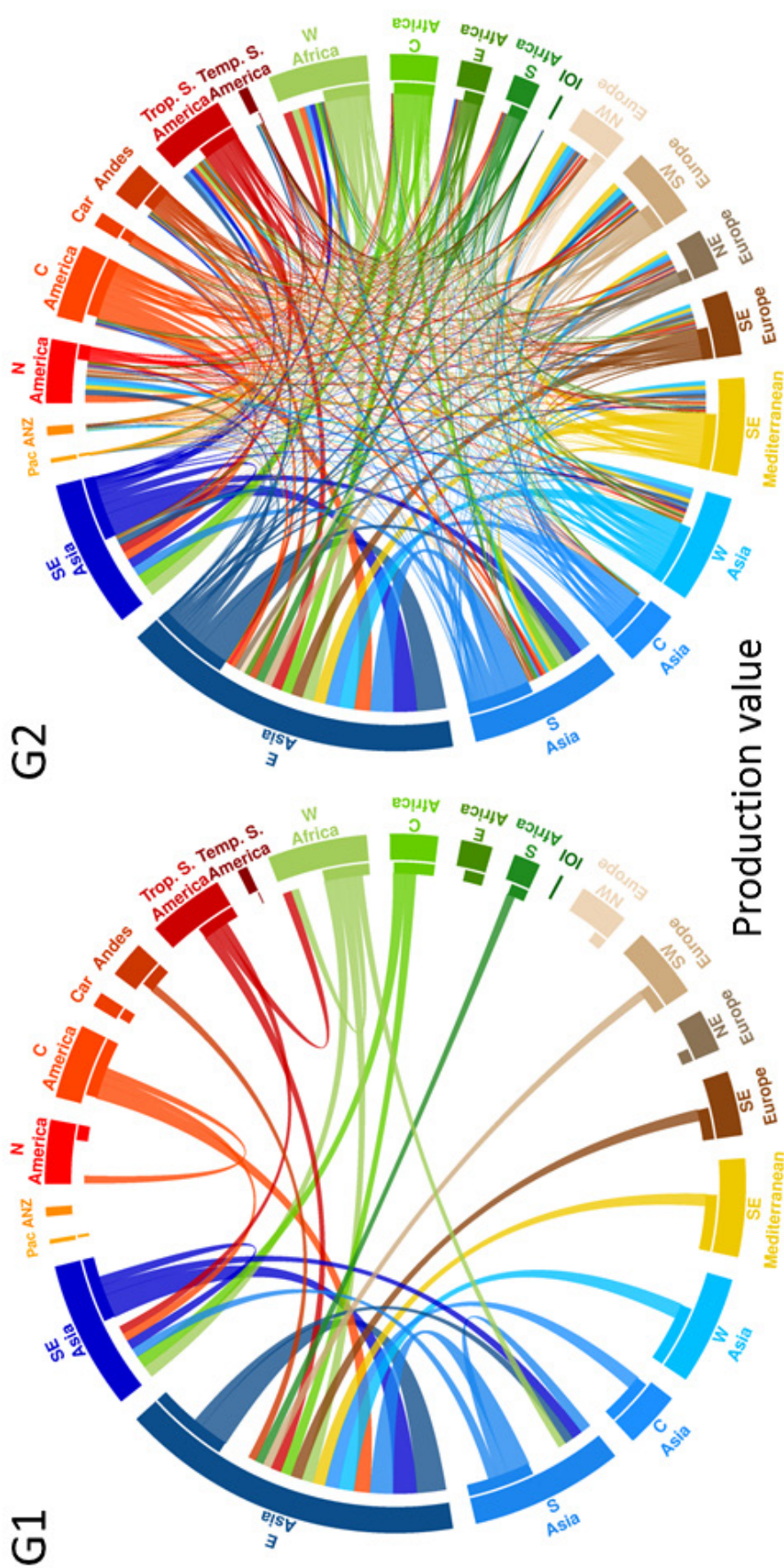
Supplementary Figure 3D. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (D) food weight in national food supplies, averaged over 2009-2011. For recipient data, regional food supply values (g/capita/day) were formed per variable by deriving a weighted average across countries comprising each region, with national values weighted by population. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, and C. America = Central America and Mexico. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



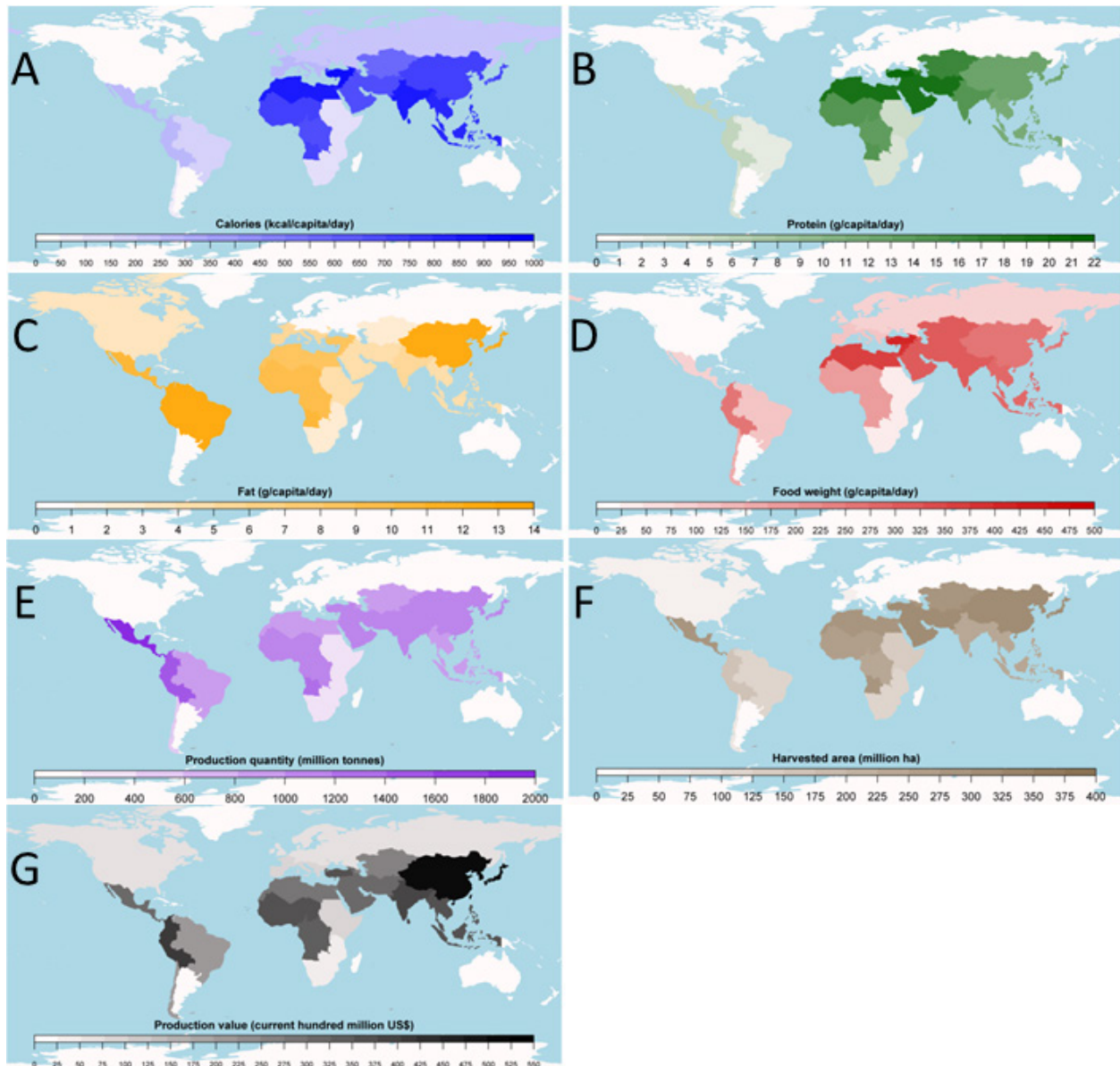
Supplementary Figure 3E. Circular plots indicating the importance of diversity as sources of crops comprising (E) production quantity in national production systems, averaged over 2009-2011. For recipient data, regional production values were formed by summing values across countries for each variable. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, C. America = Central America and Mexico, Car = Caribbean, and Pac = Tropical Pacific Region. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



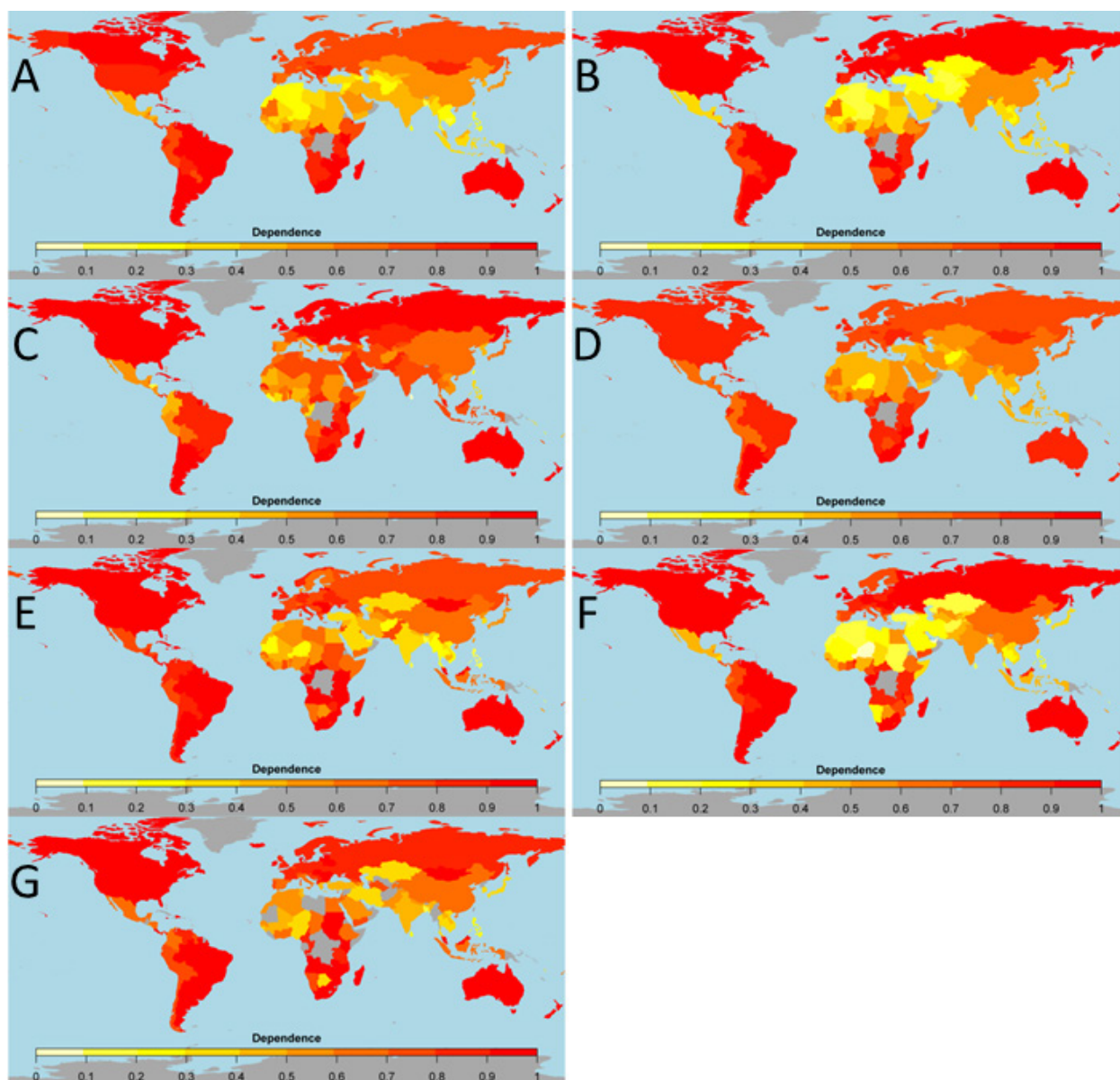
Supplementary Figure 3F. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (F) harvested area in national production systems, averaged over 2009-2011. For recipient data, regional production values were formed by summing values across countries for each variable. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, C. America = Central America and Mexico, Car = Caribbean, and Pac = Tropical Pacific Region. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



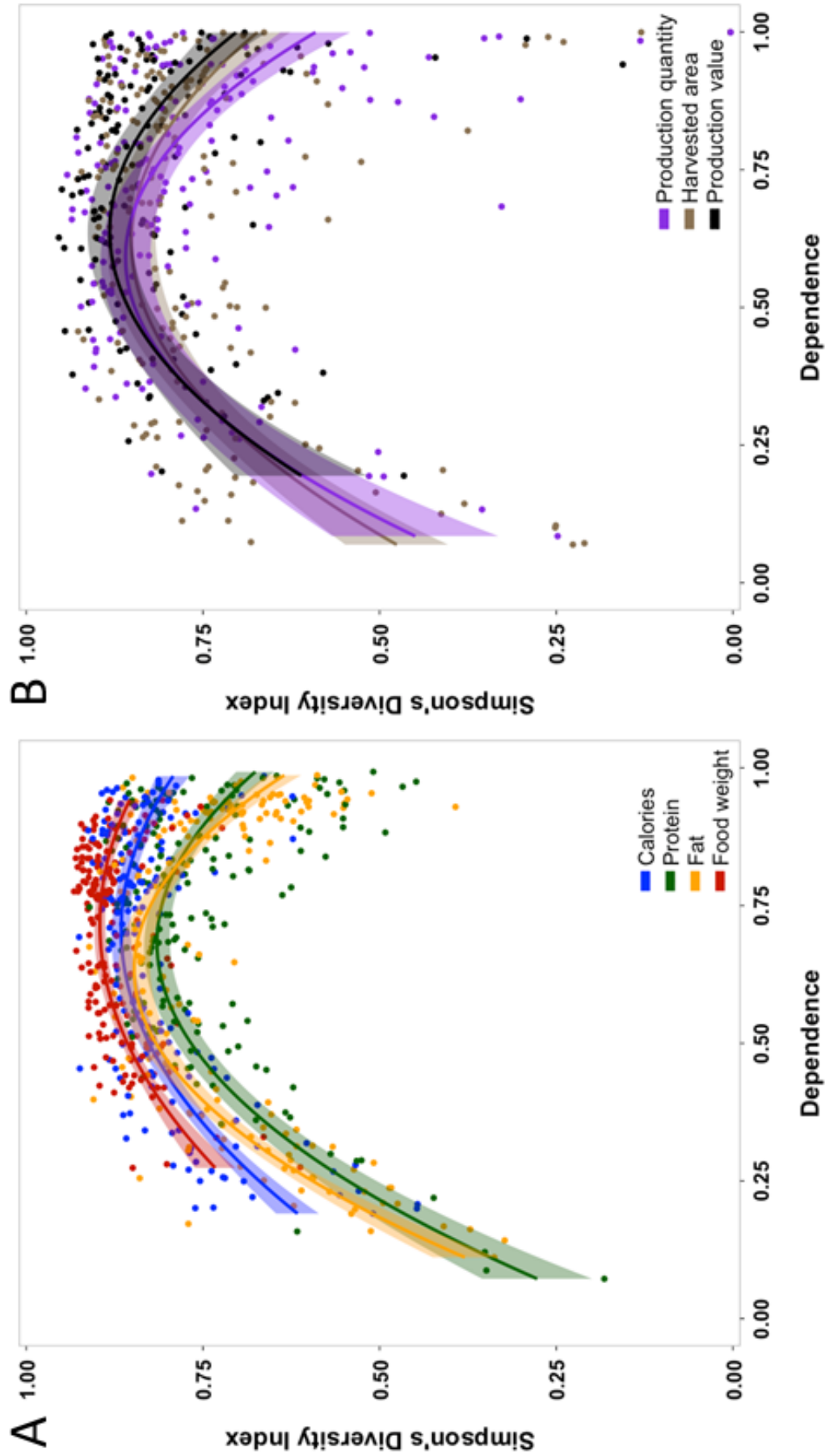
Supplementary Figure 3G. Circular plots indicating the importance of primary regions of diversity as sources of crops comprising (G) production value in national production systems, averaged over 2009–2011. For recipient data, regional production values were formed by summing values across countries for each variable. For countries within regions, see Supplementary Table 2. Region names are shortened in the figures; IOI = Indian Ocean Islands, ANZ = Australia and New Zealand, C. America = Central America and Mexico, Car = Caribbean, and Pac = Tropical Pacific Region. The direction of the importance contribution is indicated by both the origin region's color and a gap between the connecting line and the destination region's segment. The magnitude of contribution is indicated by the width of the connecting line. Because the line width is nonlinearly adapted to the curvature, it corresponds to the contribution size only at the start and end points. Figures on the left (1) display only the most significant contributions (i.e., 95th percentile) for visibility; figures on the right (2) display all contributions.



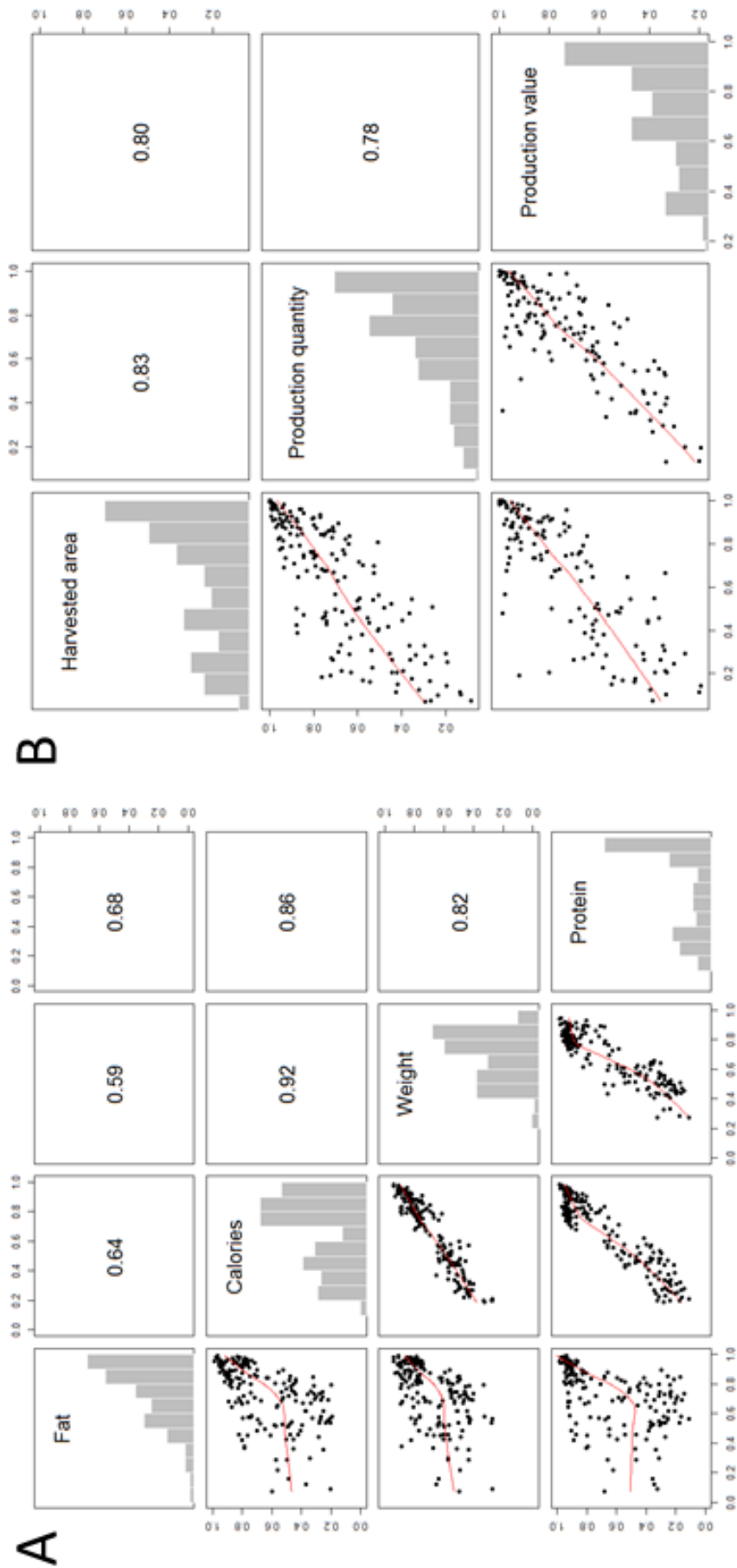
Supplementary Figure 4. Importance of primary regions of diversity of crops in contribution to global aggregate food supplies [(A) calories, (B) protein, (C) fat, and (D) food weight] and total global production systems [(E) production quantity, (F) harvested area, and (G) production value], averaged over 2009-2011.



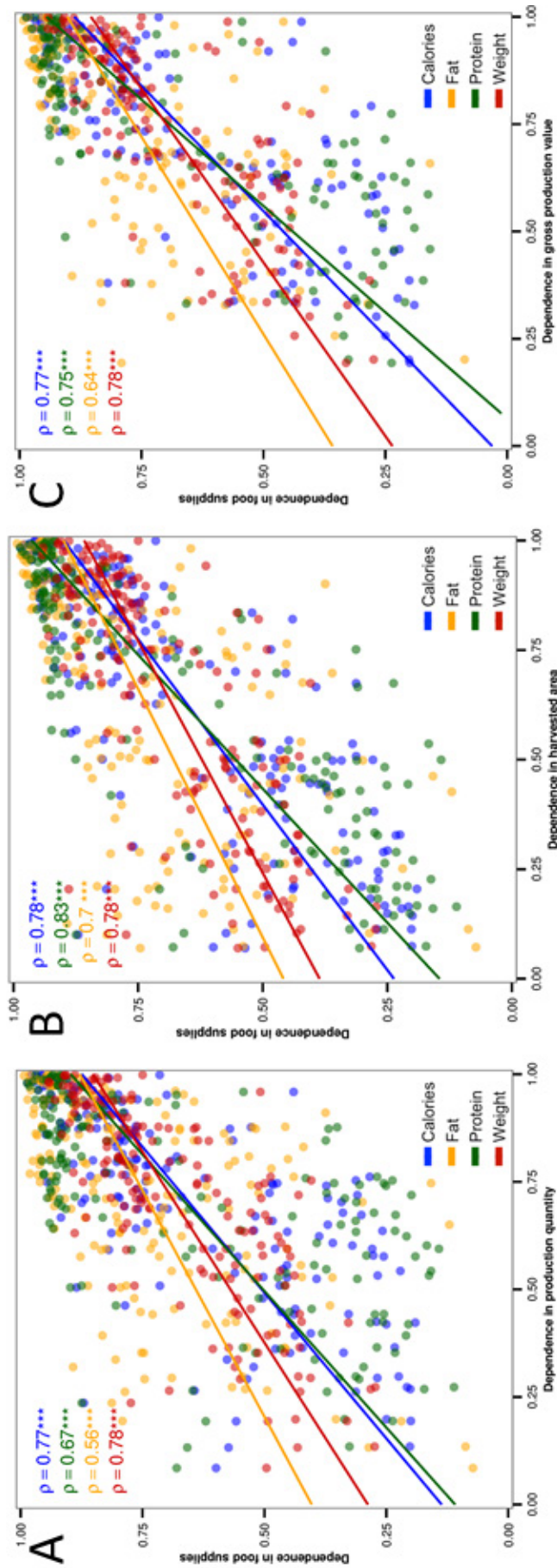
Supplementary Figure 5. Dependence on foreign primary regions of diversity of crops per country in national food supplies [(A) calories, (B) protein, (C) fat, and (D) food weight] and production systems [(E) production quantity, (F) harvested area, and (G) production value] as a modeled mean between minimum and maximum dependence per country, 2009-2011. Dependence scale is degree of dependence (1 = completely dependent).



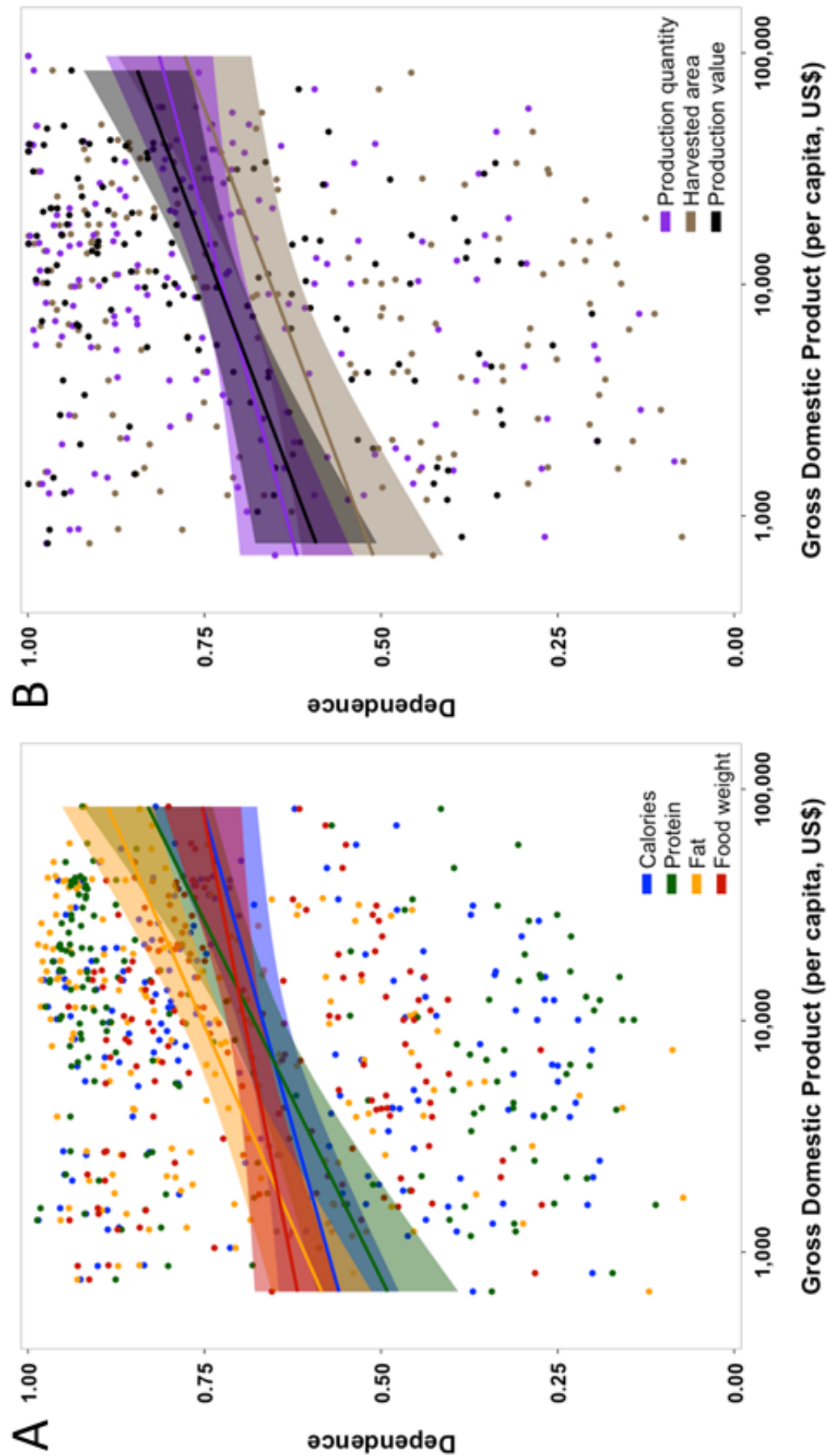
Supplementary Figure 6. Correlation within countries between national (A) food supply or (B) production system diversity, and dependence on foreign primary regions of crop diversity as a modeled mean between minimum and maximum dependence per country, 2009-2011. Dependence scale is degree of dependence (1 = completely dependent). Simpson's Index scale is degree of diversity in food supplies/production systems (1 = highly diverse). Shaded areas around each line represent the 95% confidence interval of the predicted relationship between mean diversity and mean dependence.



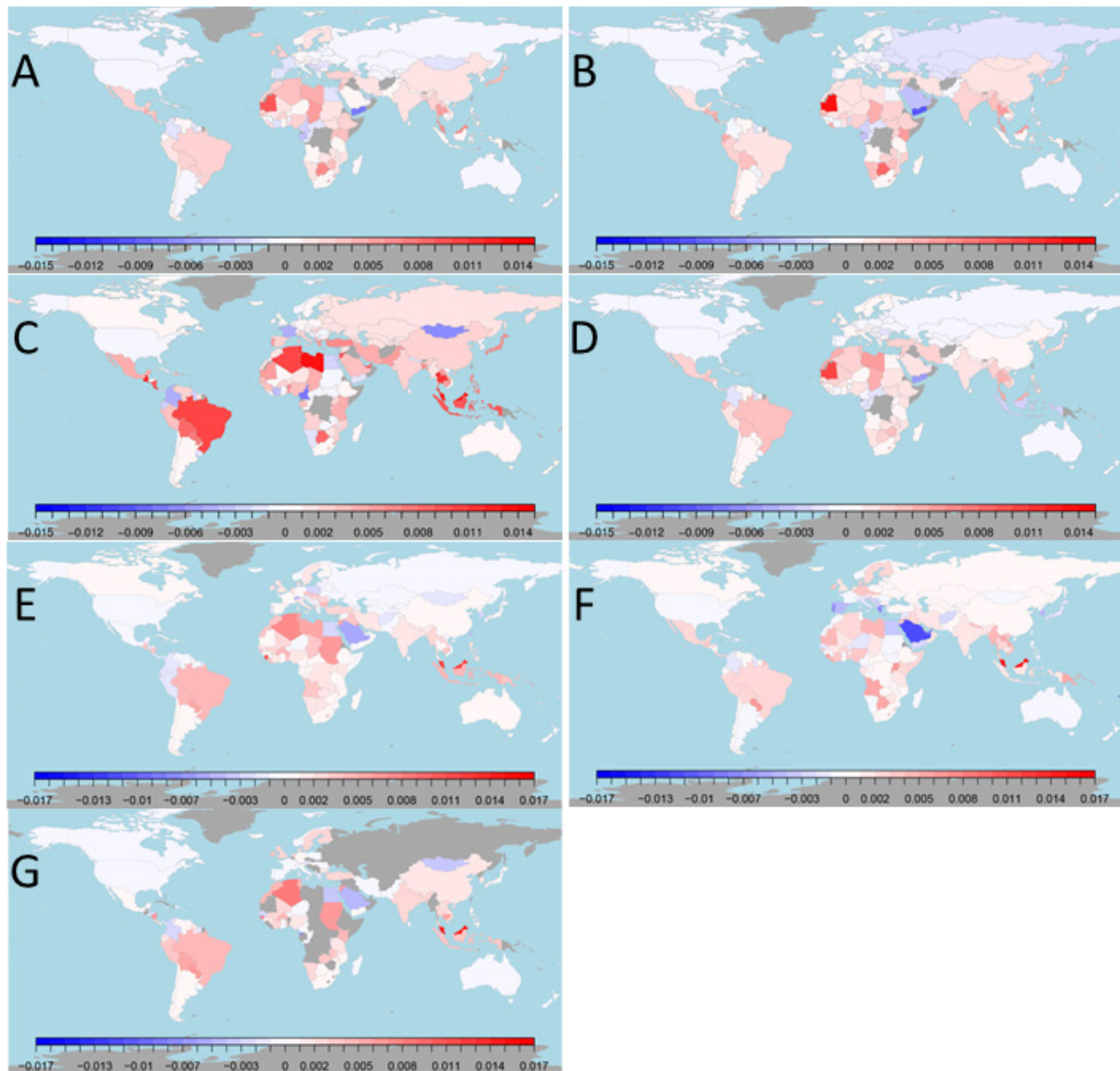
Supplementary Figure 7. Correlation within countries in regard to dependence on foreign primary regions of diversity in (A) food supplies variables, and (B) production systems variables as a modeled mean between minimum and maximum dependence per country, 2009-2011. Dependence scale is degree of dependence (1 = completely dependent).



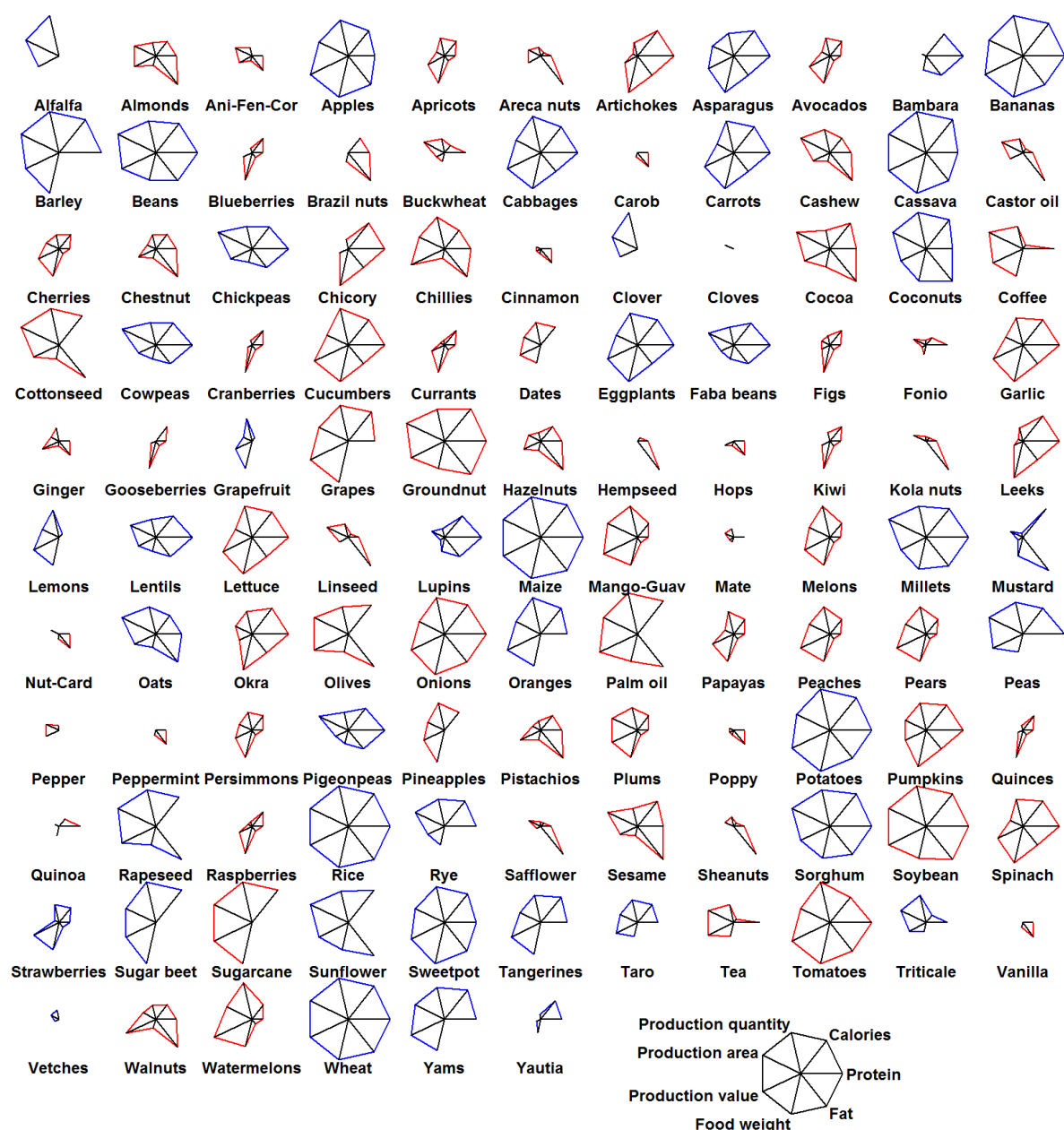
Supplementary Figure 8. Correlation within countries in regard to dependence on foreign primary regions of diversity between food supplies and production systems [(A) production quantity, (B) harvested area, and (C) production value] variables, as a modeled mean between minimum and maximum dependence per country, 2009–2011. Dependence scale is degree of dependence (1 = completely dependent).



Supplementary Figure 9. Correlation within countries between per capita gross domestic product (GDP) and national (A) food supply or (B) production systems dependence on foreign primary regions of crop diversity as a modeled mean between minimum and maximum dependence per country, 2009-2011. Dependence scale is degree of dependence (1 = completely dependent). Shaded areas around each line represent the 95% confidence interval of the predicted relationship between mean GDP and mean dependence.



Supplementary Figure 10. Slope of change in dependence from 1961-2009 on foreign primary regions of diversity in regard to national food supplies [(A) calories, (B) protein, (C) fat, and (D) food weight] and from 1961-2011 for national production systems [(E) production quantity, (F) harvested area, and (G) production value], measured as change in the modeled mean value between minimum and maximum dependence for each country in each year for each variable.



Supplementary Figure 11. Relevance of crop commodities in global food supplies and production systems worldwide, and their coverage in Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO 2002). Crops were assigned importance individually for each variable into 10% quantiles, from 1 (low importance) to 10 (high), based upon their global aggregate (food supplies) and total global production values. The center of the web is equivalent to 1; the outside of the web to 10 in importance. All specific crops are displayed; 20 general commodities are not shown here (see Supplementary Table 6 for all crops). Ani-Fen-Cor denotes Anise, Fennel and Coriander treated together; Mango Guav denotes Mango and Guava; and Nut Card denotes Nutmeg and Cardamoms. Blue outlines identify crop commodities covered in Annex 1, while red are not covered.

The following supplementary tables contain extensive information on dependence metrics per country and importance metrics per crop. These multi-sheet tables are too large to be reproduced in the supplementary information of this chapter. They are available in the “Interdependence among countries on crop genetic resources” project on Figshare (permanent links below).

Supplementary Table 1. Crop commodities and eco-geographic regions assessed in food supplies and production systems analyses.

permanent link: figshare.com/s/667b2840f33111e4a89b06ec4bbcf141

Supplementary Table 2. Countries and eco-geographic regions assessed in food supplies and production systems analyses.

permanent link: figshare.com/s/87d4197cf4d311e4971b06ec4bbcf141

Supplementary Table 3. Estimated percent dependence on foreign primary regions of diversity in national food supplies and production systems. Data includes the raw mean minimum and maximum dependence values across 2009-2011 per country, and the mean value between minimum and maximum per country across these years, as well as modeled mean values and variation metrics as estimated in a Bayesian framework using an interval-censored response variable bounded between minimum and maximum dependence estimates.

permanent link: figshare.com/s/f78dad72f33011e4b9f606ec4b8d1f61

Supplementary Table 4. Change in dependence on foreign primary regions of diversity in national food supplies, 1961-2009. Data includes minimum and maximum dependence values for each variable for each country in each year, as well as slopes of change and variation metrics over the time period as estimated in a Bayesian framework using an interval-censored response variable bounded between minimum and maximum dependence estimates. Year was centered at 1985 for modeling purposes, thus model intercepts represent mean dependence in this year.

permanent link: figshare.com/s/2e0455fef33111e4898306ec4bbcf141

Supplementary Table 5. Change in dependence on foreign primary regions of diversity in national production systems, 1961-2011. Data includes minimum and maximum dependence values for each variable for each country in each year, as well as slopes of change and variation metrics over the time period as estimated in a Bayesian framework using an interval-censored response variable bounded between minimum and maximum dependence estimates. Year was centered at 1985 for modeling purposes, thus model intercepts represent mean dependence in this year.

permanent link: figshare.com/s/4134f534f33111e4a89b06ec4bbcf141

Supplementary Table 6. Relevance of crop commodities in national food supplies and production systems, 2009-2011. Importance by quantiles: crops were assigned importance individually for each variable into 10% quantiles, from 1 (low importance) to 10 (high), based upon their global aggregate (food supplies) and total global production values. Data also includes the number of countries counting the crop within food supplies/production systems (>0) for each variable, as well as the number of these countries in which the crop is important (i.e. within the top 90% of food supply/production system variable). Within these importance counts, the number of foreign (i.e., countries whose location do not overlap with the primary regions of diversity of the crop) countries and native countries are listed. In addition, details on coverage in Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (FAO 2002) are provided for each crop.

permanent link: figshare.com/s/5029166af33111e48d2906ec4bbcf141

REFERENCES

- Abel, G.J., Sander, N. (2014). Quantifying global international migration flows. *Science* 343, 1520-1522.
- Björnstad, Å., Tekle, S., Göransson, M. (2013). 'Facilitated access' to plant genetic resources: does it work? *Gen. Res. Crop Evol.* 60, 1959-1965.
- Brennan, J.P., Godden, D., Smale, M., Meng, E. (1999). Breeder demand for and utilization of wheat genetic resources in Australia. *Plant Varieties and Seeds* 12, 113-127.
- Burke, M.B., Lobell, D.B., Guarino, L. (2009). Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. *Global Environ. Chang.* 19 (3), 317-325.
- Cassaday, K., Smale, M., Fowler, C., Heisey, P.W. (2001). Benefits from giving and receiving genetic resources: the case of wheat. *Plant Gen. Res. News.* 127, 1-10.
- Cordell, D., Drangert, J-O., White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Chang.* 19 (2), 292-305.
- D'Odorico, P., Carr, J.A., Laio, F., Ridolfi, L., Vandoni, S. (2014). Feeding humanity through global food trade. *Earth's Future* 2, 458-469.
- Day Rubenstein, K., Smale, M. (2004). *International Exchange of Genetic Resources, The Role of Information and Implications for Ownership: The Case of the U.S. National Plant Germplasm System*. EPTD Discussion Paper no. 119 [Washington D.C.: Environment and Production Technology Division, International Food Policy Research Institute (IFPRI)].
- Diamond, J. (2004). The wealth of nations. *Nature* 429, 616-617.
- Doughty, J. (1979). Dangers of reducing the range of food choice in developing countries. *Ecol. Food Nutr.* 8, 275-283.
- Dudnik, N.S., Thormann, I., Hodgkin, T. (2001). The extent of use of plant genetic resources in research - a literature survey. *Crop Sci.* 41, 6-10.
- Engels, J., Hodgkin, T., Thormann, I., Robinson, J., Fowler, C. (2001). *Crops Proposed for the Multilateral System: Centres of Diversity, Locations of Ex Situ Collections, and Major Producing Countries*. Background Study Paper No. 12 (CGRFA/CG-4/00/Inf.4) of the Fourth Inter-Sessional Meeting of the Contact Group of the CGRFA Secretariat (Rome: FAO).
- Fader, M., Gerten, D., Krause, M., Lucht, W., Cramer, W. (2013). Spatial decoupling of agricultural production and consumption: quantifying dependences of countries on food imports due to domestic land and water constraints. *Environ. Res. Lett.* 8 (1), 014046.
- FAO (Food and Agriculture Organization of the United Nations). (2015). FAOSTAT (Rome: FAO). <http://faostat3.fao.org/>.
- FAO (Food and Agriculture Organization of the United Nations). (2002). *The International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2010). *Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- Flores-Palacios, X. (1998). *Contribution to the Estimation of Countries' Interdependence in the Area of Plant Genetic Resources*. Commission on Genetic Resources for Food and Agriculture, Background Study Paper No. 7, Rev. 1 (Rome: FAO).
- Fowler, C., Hodgkin, T. (2004). Plant genetic resources for food and agriculture: assessing global availability. *Annu. Rev. Env. Resour.* 29 (1), 143-179.
- Fowler, C., Smale, M., Gaiji, S. (2001). Unequal exchange? Recent transfers of agricultural resources and their implications for developing countries. *Devel. Pol. Rev.* 19, 181-204.
- Gelman, A., Hill, J. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models* (New

- York: Cambridge University Press).
- Gelman, A., Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Sci.* 7, 457-472.
- Gepts, P. (2006). Plant genetic resources conservation and utilization: the accomplishments and future of a societal insurance policy. *Crop Sci.* 46 (5), 2278-2292.
- Global Crop Diversity Trust. (2013). *Fundraising Strategy 2014-2018* (Bonn: Global Crop Diversity Trust).
- Gollin, D. (1998). "Valuing farmers' rights", in *Agricultural Values of Plant Genetic Resources*, eds. R.E. Evenson, D. Gollin, V. Santaniello (CAB International, Wallingford, UK), pp. 233-245.
- GRIN (Germplasm Resources Information Network). (2014). USDA, ARS, National Genetic Resources Program, Beltsville. http://www.ars-grin.gov/cgi-bin/npgs/html/tax_search.pl.
- Guarino, L., Lobell, D.B.A. (2011). A walk on the wild side. *Nat. Climate Chang.* 1. 374-375.
- Harlan, J.R. (1951). Anatomy of Gene Centers. *American Naturalist* 8.
- Harlan, J.R. (1971). Agricultural origins: centres and noncentres. *Science* 174, 468-474.
- Harlan, J.R. (1975). *Crops and Man* (Madison: American Society of Agronomy).
- Hoisington, D., Khairallah, M., Reeves, T., Ribout, J.-M., Skovmand, B., Taba, S., *et al.* (1999). Plant genetic resources: what can they contribute toward increased crop productivity? *Proc. Natl. Acad. Sci.* 96, 5937-5943.
- Jarvis, A., Ramirez-Villegas, J., Hanson, J., Leibing, C. (2009). "Crop and forage genetic resources: international interdependence in the face of climate change", in: *The Impact of Climate Change on Countries' Interdependence on Genetic Resources for Food and Agriculture*. Background Study Paper No.48, eds. S. Fujisaka, D. Williams, M. Halewood (Rome: FAO, Commission on Genetic Resources for Food and Agriculture).
- Johnson, N.L., Pachico, D., Voysest, O. (2003). The distribution of benefits from public international germplasm banks: the case of beans in Latin America. *Agric. Econ.* 29, 277-286.
- Kastner, T., Rivas, M.J.I., Koch, W., Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. *Proc. Natl. Acad. Sci.* 109, 6868-6872.
- Kearney, J. (2010). Food consumption trends and drivers. *Philos. T. R. Soc. B.* 365 (1554), 2793-2807.
- Khoury, C.K., Bjorkman, A.D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., *et al.* (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci.* 111 (11), 4001-4006.
- Khoury, C.K., Jarvis, A. (2014). *The Changing Composition of the Global Diet: Implications for CGIAR Research*. CIAT Policy Brief No. 18. (Cali: Centro Internacional de Agricultura Tropical), 6 pp.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Gen. Res. Crop Evol.*, 57, 625-639.
- Kloppenborg, J.Jr., Kleinman, D.L. (1987). Analyzing empirically the distribution of the world's plant genetic resources: the plant germplasm controversy. *Bioscience* 37, 190-198.
- Lobell, D.B., Burke, M.B., Tebaldi, C., Mastrandrea, M.D., Falcon, W.P., Naylor, R.L. (2008). Prioritizing climate change adaptation needs for food security in 2030. *Science* 319 (5863), 607-610.
- MacDonald, G.K., Brauman, K.A., Sun, S., Carlson, K.M., Cassidy, E.S., Gerber, J.S., *et al.* (2015). Rethinking agricultural trade relationships in an era of globalization. *Bioscience* 65, 275-289.
- McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., *et al.* (2013). Agriculture: feeding the future. *Nature* 499 (7456), 23-24.
- Nuijten, E., van Treuren, R., Struik, P.C., Mokuwa, A., Okry, F., Teeken, B., *et al.* (2009). Evidence for the emergence of new rice types of interspecific hybrid origin in West African farmers' fields. *PLoS One* 4, e7335.
- Porkka, M., Kumm, M., Siebert, S., Varis, O. (2013). From food insufficiency towards trade

- dependency: a historical analysis of global food availability. *PLoS One* 8, e82714.
- Prescott-Allen, R., Prescott-Allen, C. (1990). How many plants feed the world? *Conserv. Biol.* 4 (4), 365-374.
- Prota (Plant Resources of Tropical Africa). (2014). Prota4u online database. <http://www.prota4u.info/>.
- Smale, M. (1996). *Understanding Global Trends in the Use of Wheat Diversity and International Flows of Wheat Genetic Resources*. Economics Working Paper, 96 (2). [International Maize and Wheat Improvement Center (CIMMYT)].
- Smale, M., Day Rubenstein, K. (2002). The demand for crop genetic resources: international use of the U.S. National Plant Germplasm System. *World Dev.* 30, 1639-1655.
- Smale, M., Reynolds, M.P., Warburton, M., Skovmand, B., Trethowan, R.M., Singh, R.P., *et al.* (2002). Dimensions of diversity in modern spring bread wheat in developing countries from 1965. *Crop Sci.* 42, 1766–1779.
- Thormann, I., Fiorino, E., Halewood, M., Engels, J.M.M. (2015). Plant genetic resources collections and associated information as a baseline resource for genetic diversity studies: an assessment of the IBPGR-supported collections. *Genet. Resour. Crop Evol.* doi 10.1007/s10722-015-0231-9.
- van de Wouw, M., Kik, C., van Hintum, T., van Treuren, R., Visser, B. (2009). Genetic erosion in crops: concept, research results and challenges. *Plant Genetic Resources* 8, 1-15.
- van de Wouw, M., van Hintum, T., Kik, C., van Treuren, R., Visser, B. (2010). Genetic diversity trends in twentieth century crop cultivars: a meta analysis. *Theor. Appl. Genet.* 120 (6), 1241-1252.
- Vavilov, N.I. (1926). Tzentry proiskhozhdeniya kulturnykh rastenii (The centers of origin of cultivated plants). *Works of Applied Botany and Plant Breeding* 16, 248pp.
- Vavilov, N.I. (1951). The origin, variation, immunity and breeding of cultivated plants (Translated by K. Start). *Cron. Bot.* 13, 1-366.
- Vavilov, N.I. (1992). *Origin and Geography of Cultivated Plants* (Translated by Doris Löve) (Cambridge: Cambridge University Press).
- Vincent, H., Wiersema, J., Kell, S., Fielder, H., Dobbie, S., Castañeda-Álvarez, N.P., *et al.* (2013). A prioritized crop wild relative inventory to help underpin global food security. *Biol. Conserv.* 167, 265-275.
- Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. A modest proposal for teosinte evolution and conservation *in situ*: the Balsas, Guerrero, Mexico. *Maydica* 52, 49-58.
- World Bank. (2014). International Comparison Program database. <http://data.worldbank.org/>.
- Xiao, J., Grandillo, S., Ahn, S.N., McCouch, S.R., Tanksley, S.D. (1996). Genes from wild rice improve yield. *Nature* 384, 223-224.
- Zeven, A.C., De Wet, J.M.J. (1982). *Dictionary of Cultivated Plants and Their Regions of Diversity: Excluding Most Ornamentals, Forest Trees and Lower Plants* (Wageningen: CAPD).
- Zeven, A.C., Zhukovsky, P. (1975). *Dictionary of Cultivated Plants and Their Centres of Diversity: Excluding Most Ornamentals, Forest Trees and Lower Plants* (Wageningen: CAPD).
- Zhou, X., Carter, T.E.Jr., Cui, Z., Miyazaki, S., Burton, J.W. (2000). Genetic base of Japanese soybean cultivars released during 1950 to 1988. *Crop Sci.* 40, 1794-1802.
- Zhu, Y., Chen, H., Fan, J., Wang, Y., Li, Y., Chen, J., *et al.* (2000). Genetic diversity and disease control in rice. *Nature* 406 (6797), 718-722.
- Zhukovsky, P.M. (1965). Main gene centres of cultivated plants and their wild relatives within the territory of the U.S.S.R. *Euphytica* 14, 177-188.
- Zhukovsky, P.M. (1968). New centres of origin and new gene centres of cultivated plants including specifically endemic microcentres of species closely allied to cultivated species. *Bot. J. (Russian Bot Z.)* 53, 430-460.

CHAPTER 8

General discussion

DISCUSSION OF RESULTS AND IMPLICATIONS

This thesis aimed to contribute to the knowledge required to answer a set of key questions regarding the need for, potential of, challenges and constraints regarding, and necessary steps to enhance the conservation and use of crop genetic diversity. A discussion of the results and their implications is presented in the following subsections.

The status of diversity and associated vulnerabilities in the global food system

In the past half century, very substantial changes have occurred in human diets worldwide and in the production systems that sustain them. National food supplies around the world have become increasingly similar (Figure 1), gaining in calories, protein, and fat, as animal-derived foods and high-calorie plant foods (oils and sugars) have risen in importance. The proportion of diets consisting of major cereals, sugar crops and oil crops has increased, while regionally and locally important cereals, root crops, and oil crops have generally become further marginalized [Khoury *et al.* 2014 (Chapter 2)].

These changes have been driven by globalization, urbanization, and economic development, including agricultural research (Khoury & Jarvis 2014). While this nutrition transition has enhanced food security by making macronutrients more readily available worldwide, it has had mixed effects on micronutrient sufficiency and the over-consumption of macronutrients has contributed to a global surge in diet-related non-communicable diseases (Popkin 2006, Pingali 2007, Kearney 2010). Dietary change is also linked with greater homogeneity in farmers' fields and the associated commodity trading systems, thus heightening concerns about genetic vulnerability to biotic and abiotic stresses as well as food system vulnerability to climatic and political instability (Chapter 2).

What can be done to mitigate the vulnerabilities created by placing our proverbial eggs in one basket - an increasingly interconnected global food system, highly contingent upon trade in a handful of selected food crop commodities, highly dependent on the extensive application of

Each country's food supply composition in contribution to calories in:

● 1961 ● 1985 ● 2009

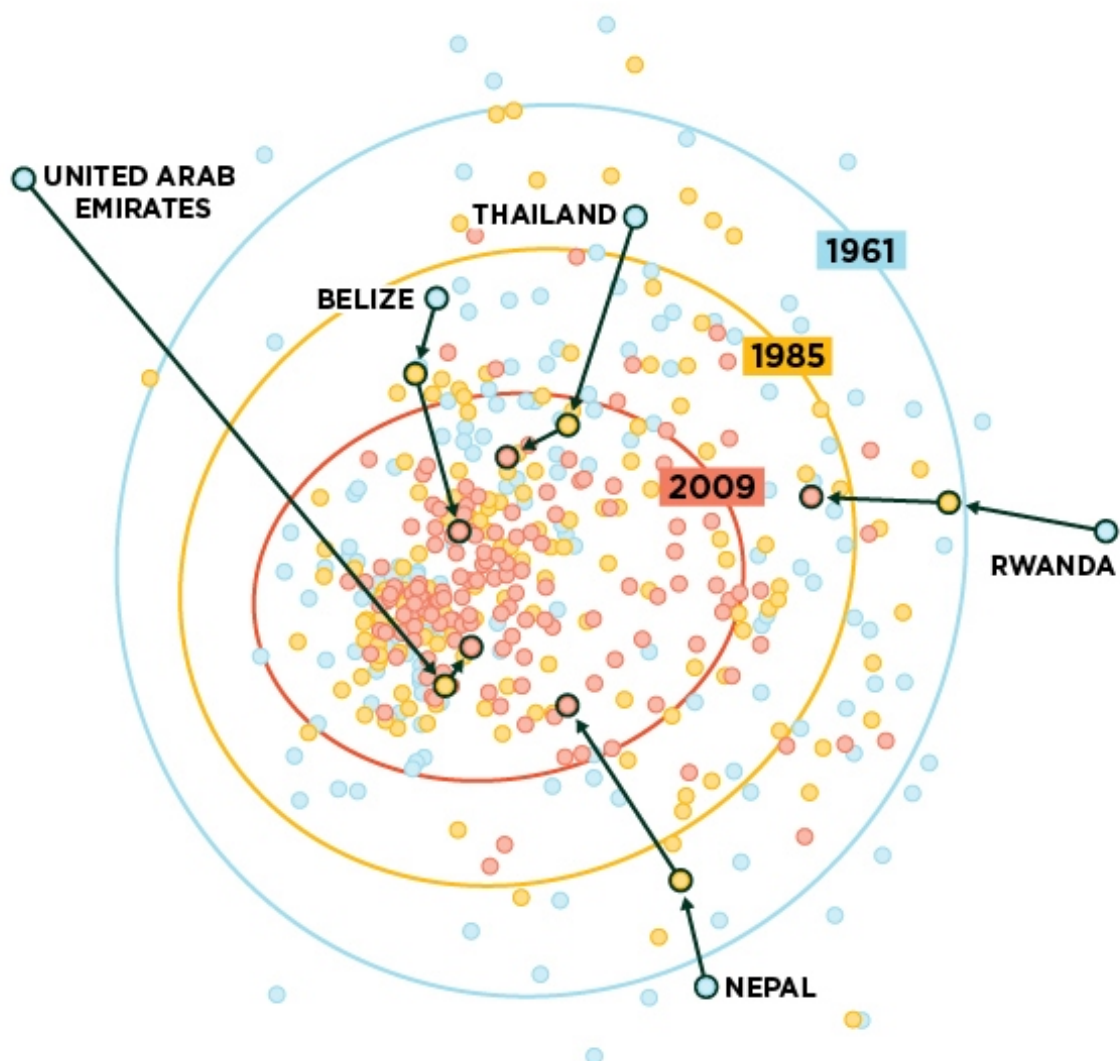


Figure 1. Increase in similarity (homogeneity) in national food supplies, 1961-2009. The figure depicts a multivariate ordination of crop commodity composition in contribution to calories in national food supplies in 1961, 1985 and 2009. Blue points represent the multivariate commodity composition of each country in 1961, yellow points in 1985, and red points in 2009. Circles represent 95% confidence intervals around the centroid in each year. Both x- and y-axis have been deliberately removed for visualization purposes. See Khoury *et al.* (2014) (Chapter 2) for further details.

renewable and non-renewable resources, and creating major environmental impacts?

A comprehensive view of food and nutrition security encompasses consistent availability and access to adequate, culturally acceptable and nutritious food as well as empowerment of consumers to use this food for improved health. Over the long term, food security also requires actions to mitigate the negative ecological effects of food systems and adapt agriculture to

climatic change as well as to natural resource limitations. Achieving long-term food security thus requires a holistic approach to agricultural development. From a crop diversity perspective, such an approach aims to build food systems in which diverse varieties of a broad range of crops can flourish in terms of production, markets, and consumption.

Key steps that can help to mitigate global food system vulnerabilities, increase agricultural productivity and sustainability, and enhance human nutrition are proposed below. These actions are divided into sections devoted to major staple crops, increasingly dominant oil crops, and neglected or marginalized crops, with particular focus on the role of crop genetic resources in enhancing food and nutrition security.

As wheat, rice, maize, and other critical staples gain importance in the global food system (Figure 2), their production stability, nutritional quality, and environmental impacts become ever more critical issues. Improving these crops in the face of land, water, and other resource limitations, climate change, and agriculture's increasing pressure on ecosystems, will necessitate a range of research and political actions, including (a) conserving and describing the genetic variation within these crops and their wild relatives and making diverse genetic material and associated information available to researchers and breeders, (b) facilitating sharing of diversity through the International Treaty on Plant Genetic Resources for Food and Agriculture (Plant Treaty) and aligned political efforts, (c) breeding more productive and resilient varieties, with emphasis on enhancing nutritional quality, stress tolerance, and resource use efficiency, (d) promoting widespread adoption of genetically diverse, locally adapted varieties to minimize vulnerability associated with genetic uniformity, and (e) developing more ecologically efficient agronomic, storage, processing and distribution practices in order to reduce their negative environmental impacts (Khoury & Jarvis 2014).

Challenges related to the production of oil crops such as soybean and oil palm that have become an increasingly important part of global food supplies over the past 50 years (Figures 2 and 3) include major impacts on natural ecosystems through deforestation, greenhouse gas emissions, and vulnerabilities due to genetic uniformity in crop varieties. Key objectives for production systems focused on oil production should be on reducing these ecological impacts, improving species diversity at the system level (i.e., diversify crops) and enhancing genetic diversity at the varietal level. Steps that can be taken to address these challenges include:

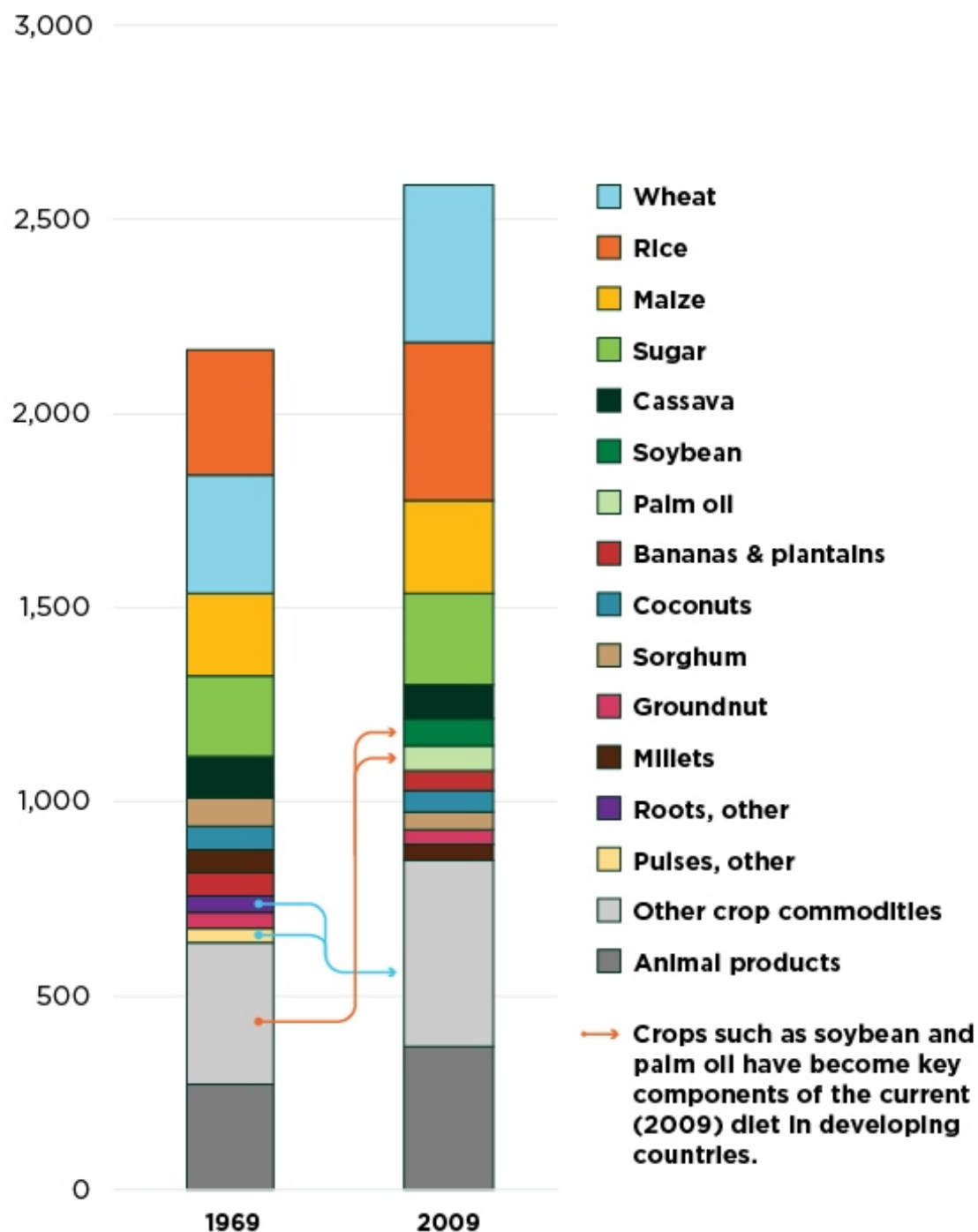


Figure 2. Contribution of crop commodities to mean food supplies in developing countries for calories (kcal/capita/day), 1969 and 2009 [Khoury & Jarvis 2014, based on Khoury *et al.* 2014 (Chapter 2)].

(a) developing and promoting ecologically efficient production and processing methods, (b) broadening the genetic diversity within major varieties, and (c) diversifying oil crop production, processing, and markets by promoting less globally dominant oil crops such as coconut, cottonseed, groundnut, olive, rape and mustardseed, sesame, shea nut, and sunflower.

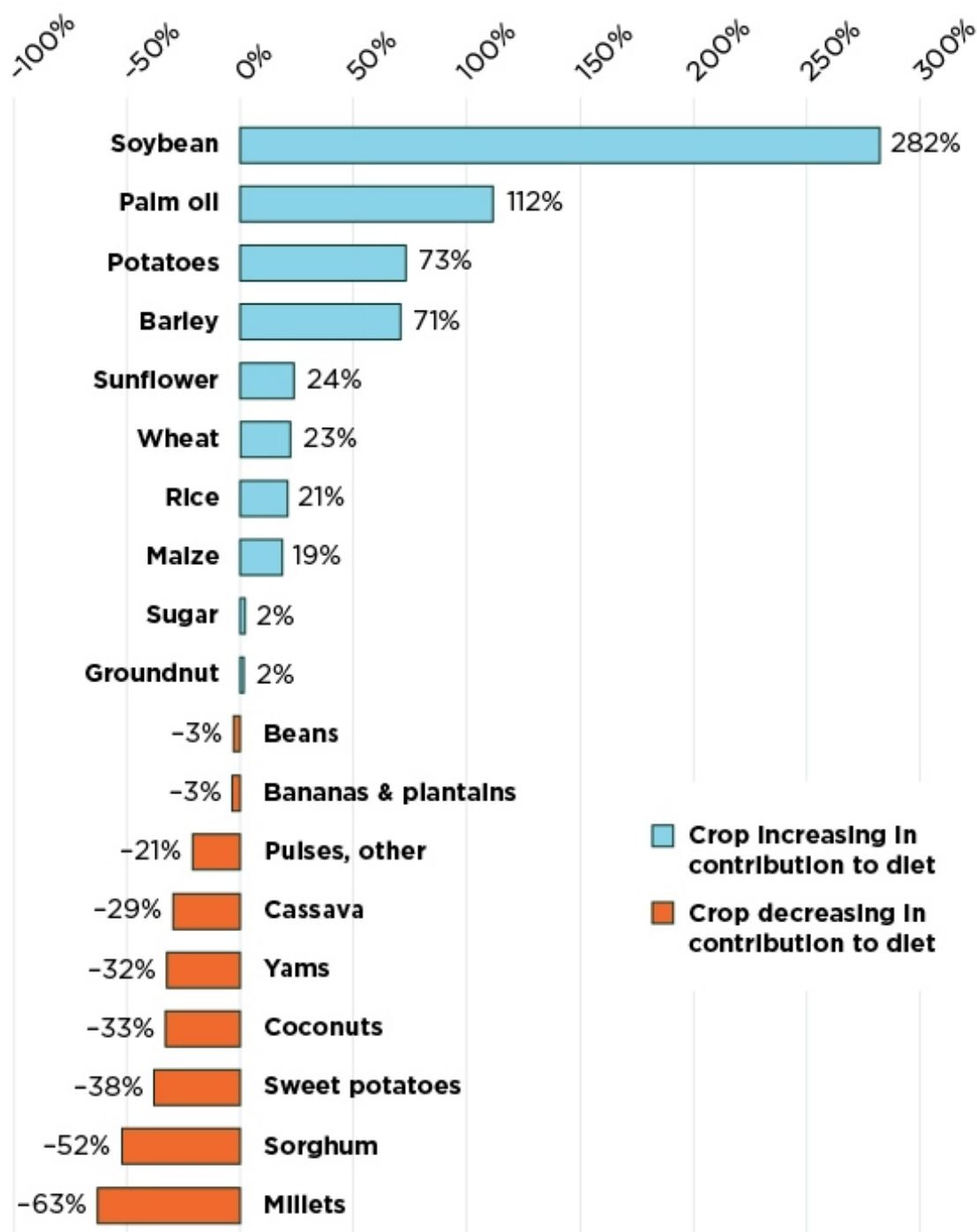


Figure 3. Median percent change in the relative contribution to calories (kcal/capita/day), from assorted crops in national food supplies in developing countries, 1969-2009 [Khoury & Jarvis 2014, based on Khoury *et al.* 2014 (Chapter 2)].

A number of cereal, root, and oil crop species have generally declined in terms of their relative contribution to national and global food supplies in the past half century (Figure 3), although they may remain significant particularly for rural communities in some developing regions. These include sorghum, cassava, millets, sweetpotato, coconut, yams, and grain legumes, among others. Agricultural research funding for these crops is minimal compared with funding

for the major cereals (Khoury & Jarvis 2014). Since many of these plants are both stress tolerant and nutritionally rich, investment in their conservation, improvement, and promotion offers a wise long-term option for diversifying global food supplies, particularly as the environmental challenges that agriculture faces intensify and as more people suffer from the negative health effects of the nutrition transition (Popkin 2006, Pingali 2007, Kearney 2010).

The recent rise of quinoa (Zurita-Silva *et al.* 2014) demonstrates how global food supplies can be diversified, particularly against a background of growing consumer interest in diverse and healthy food alternatives. A potential benefit of globalization and urbanization trends may be an increased tolerance for healthy non-traditional foods. Increasing the contribution of marginalized crops in global food supplies requires focus both on establishment or improvement of incentives for farmers and seed systems at the production level as well as promotion at the consumer level. Specific actions may include: (a) identifying regionally and locally important crop species that show potential for improved productivity, enhanced nutritional quality, and greater competitiveness under challenging conditions, (b) conserving as well as fostering the use of their genetic diversity, (c) breeding productive and resilient varieties of these crops, with particular emphasis on traits such as increased micronutrient levels, storage life, and versatility in processing and consumer use, (d) making these materials widely available to breeders and other researchers, facilitating their uptake through agricultural extension and training, and developing robust seed systems for their multiplication and distribution, and (e) stimulating policy measures that strengthen market demand for these crops.

Potential, geographic distributions, and conservation priorities for crop wild relatives

A number of key impressions emerge from the present research devoted to identifying, investigating the potential of, and determining the conservation status of the genetic diversity represented in crop wild relatives. First, a great deal of wild diversity exists and is therefore potentially available for use in breeding. The national assessment for the USA, a country generally considered to be depauperate in regard to crop diversity (Vavilov 1926), identified a great range of native and introduced species associated with a long list of important crops (Chapter 3). Case studies on the wild relatives of beans, sweetpotato and pigeonpea revealed diverse crop genetic resources widely distributed over regions and continents (Chapters 4-6).

Second, wild genetic resources have considerable potential to contribute traits of value to crop

improvement. Populations of wild species related to sweetpotato (Chapter 5) and pigeonpea (Chapter 6) are distributed in areas modeled to be well outside of the temperature, precipitation, and soil characteristic comfort zone of these crops (see Figure 4 of Chapter 5, and Figures 5 and 6 of Chapter 6). Similar results have been reported for wild relatives and/or landraces of other crops (e.g., Endresen *et al.* 2011, Tapia *et al.* 2014). Given the continual advances in information on wild relatives, as well as improving breeding technologies facilitating the utilization even of distantly related species (Zamir 2001, Cao *et al.* 2009, McCouch *et al.* 2013), there is good reason to consider crop wild relatives as genetic resources of increasing potential for major contributions to crop improvement.

Third, these wild genetic resources are drastically under-represented as a whole in *ex situ* conservation systems. With an estimated 57% of wild relatives of bean (Chapter 4), 79% of sweetpotato (Chapter 5), and 80% of pigeonpea (Chapter 6) assessed as high priority for further collecting, and virtually all crop wild relative species determined to be insufficiently represented in genebanks in regard to the full range of geographic and ecological variation in their distributions, it is evident that much remains to be done to conserve their genetic diversity. Such results are concordant with conservation assessments for other crop gene pools (Khoury *et al.* 2010).

Given such large gaps remaining in conservation needs for crop wild relatives, significant resources needed for their maintenance *ex situ*, and seemingly chronic funding deficits for genetic resources conservation (Khoury *et al.* 2010), a further assessment ideally informed by genotypic diversity analyses of what constitutes sufficient germplasm collections for wild relatives is warranted. As extremely valuable traits sourced from crop wild relatives have been found in only a few populations despite screening of a large number of accessions (Brar & Khush 1997), and given modeling of large variation within species in regard to traits such as resistance to insect pests (Sharma *et al.* 2003), seed protein content and days to maturity (Upadhyaya *et al.* 2013), and tolerance to high precipitation (Martin & Jones 1973, Nimmakayala *et al.* 2011), the case can be made for increasing funding for collecting and conservation so as to comprehensively conserve wild genetic resources. The window of opportunity to fully collect and conserve this genetic diversity will not remain open indefinitely (Wilkes 2007, FAO 2010).

Along with ensuring their maintenance *ex situ*, further investment in genebank information systems, *ex situ* conservation technologies (i.e., storage, testing, and regeneration), safety duplication of unique germplasm, characterization and evaluation for traits of interest, taxonomic and systematic studies, and breeding research are needed to mobilize wild genetic resources for use in crop improvement (FAO 2010, Khoury *et al.* 2010, Tester and Langridge 2010, Guarino & Lobell 2011, Henry 2014).

Access priorities for crop genetic diversity

Given the global geography of crop genetic diversity, it is evident that international collaboration is critical to achieving access to these resources. The vast majority of countries, even those located within the most ancient and richest primary regions of crop genetic diversity, e.g., West Asia, are significant consumers and producers of crops whose genetic diversity largely sources from outside their borders (Chapter 7). The importance of ‘foreign’ crops in national food supplies and production systems is growing over time as national food systems become more diverse and at the same time more homogeneous worldwide. These trends bolster the rationale for considering genetic resources of internationally important crops as global public goods, the common heritage of humankind, the conservation and facilitated access of which should be proactively supported by all countries worldwide (FAO 2002, Esquinas-Alcázar 2005). They also reaffirm the critical importance of the germplasm collections safeguarded by CGIAR and key national genebanks, which have borne most of the burden of germplasm distribution to the global agricultural research community over the past few decades.

LIMITATIONS AND UNCERTAINTIES

The main limitations and uncertainties inherent in the thesis research relate to the reliability and sufficiency of input information, and the robustness of modeling methods. Food supply and production data managed by the Food and Agriculture Organization of the United Nations (FAO) lack the specificity needed for an exhaustive analysis of changes in crop diversity [Khoury *et al.* 2014 (Chapter 2)], and although significant efforts are being taken to improve data variably reported by national institutions (e.g., FAO 2015), it is disappointing that annual data was probably of greater scope and quality 35 years ago than it is now (R. Prescott-Allen, personal communication 2014). Broad support for and increased investment in agricultural statistics by all countries are critically needed in order to resolve these deficiencies.

Significant further research is still needed to identify wild plant taxa, assess species boundaries within taxonomic groups, and determine degrees of relatedness with associated crops, as evidenced by the study on sweetpotato (Chapter 5). Incorrect or outdated species identifications, synonymy, and erroneous or low resolution occurrence information are recognized challenges in the utilization of the current openly available biodiversity resources (Gaiji *et al.* 2013).

The rapidly evolving field of species distribution modeling is continuously improving in its methodologies, but deficiencies and uncertainty are acknowledged (Dormann 2006, Hijmans & Graham 2006, Graham *et al.* 2008, Jimenez-Valverde *et al.* 2008, Lobo *et al.* 2008, Loiselle *et al.* 2008, Wisz *et al.* 2008, Fitzpatrick & Hargrove 2009), particularly in regard to sampling bias due to the lack of absence information and imperfect sampling techniques (Araújo & Guisan 2006, Costa *et al.* 2010, Kramer-Schadt *et al.* 2013). The methods utilized in the thesis drew upon the scientific literature to refine data inputs, capitalized on a large set of climatic and edaphic variables to determine distributions, performed modeling with different variations of this data in order to test robustness, and employed a set of statistical analyses to determine confidence in the results. Further studies that validate modeling results (Jarvis *et al.* 2005, Cobben *et al.* 2014) will be useful in the application of species distribution modeling to genetic resources conservation planning. As the potential distributions maps created in this thesis and elsewhere (e.g., Castañeda-Alvarez *et al.*, 2015) are informing ongoing collecting missions (Dempewolf *et al.* 2014), feedback from the fieldwork can result in improvements in the methods used here.

While the analyses presented in Chapters 5 and 6 of this thesis revealed large eco-geographic variation in wild populations compared with the cultivated species, further studies are needed to correlate such variation with useful adaptive traits that can be bred into commercial genotypes. Such analyses would improve the knowledge base on potential uses of crop wild relatives.

An important tool employed in the gap analysis methodology to mitigate deficiencies and uncertainty in input data and modeling techniques is the expert evaluation process, which can involve a number of researchers and utilizes both quantitative and qualitative metrics (Chapters 4-6). Agricultural research in parallel with technology appears to be transitioning from high quality restricted scope studies performed by a limited number of experts, to ‘big data’

approaches drawing overall trends from large amounts of information. The expert evaluation and the data processing and modeling steps bring together these classical and emerging methods in the gap analysis, attempting to draw upon the strengths of both worlds. While the expert assessments were considered to be extremely valuable in verifying the modeled results for the crop genepools assessed here, the unfortunate reality is that there are fewer and fewer experts left that have a full career of experience in taxonomy, genetic resources, and/or breeding. The loss of these human encyclopedias on crop diversity has some regrettable parallels to the folk tale of John Henry (Bradford 1931). Henry's ability as a (human) steel-driver was measured in a race against the emerging technology of a steam powered hammer. Henry won, only to die in victory with his hammer in his hand as his heart gave out from stress. As research transitions from dependence on knowledge stored in human brains to data stored in online platforms, improvements in data quality, modeling methods, and statistical analyses of significance of trends will become ever more critical to the validity of results.

An additional uncertainty in regard to conceptualizing the future potential of crop genetic diversity is that advancing molecular and breeding tools are expanding the capacity to successfully introgress traits from distant relatives (Zamir 2001, McCouch *et al.* 2013, Henry 2014), making the boundaries of what might be considered potentially valuable related species less obvious. Developments in cisgenic, transgenic, and gene editing techniques (Sander & Joung 2014) will only widen the scope of possibility, with implications for genetic resource conservation, exploration, and use.

This thesis represents concrete steps taken to conceptualize global change in crop diversity at the species level, advance information on crop wild relatives for one country and for three crop genepools, and contribute to policy arguments in support of facilitated access to crop genetic resources of global importance. A comprehensive understanding of the diversity, conservation, and utilization needs for genetic resources associated with the world's crops requires parallel and extended research efforts across other levels of diversity, regions and crop genepools. As the present study proposes methods that rely upon openly available data, tools and software, it is hoped that the effort will facilitate this further needed research.

IMPACT

The results presented in this thesis have been received by the scientific community and the public with remarkable interest.

The research on homogeneity in global food supplies (Khoury *et al.* 2014) was covered extensively by the media, with over 300 articles, blogs, and podcasts published in the first year following its release, including in top tier newspapers and journals such as *Scientific American*, *Time Magazine*, and *National Geographic* (see the Publications of author section of this thesis for references for a selection of these articles). This attention led to an altmetric score considering the research among the top 100 articles ever published in *PNAS*, and within the 99th percentile of all scientific articles ever tracked. The study was remarked upon by leaders of the World Bank and CGIAR, inspired significant discussion on research and funding priorities in CGIAR, generated a series of invited articles, blogs, and presentations given internationally, and was further reproduced in academic textbooks, radio and television shows.

Research on crop wild relatives has also received considerable attention. The highly cited gap analysis methodology (Ramírez-Villegas *et al.* 2010) has become a central method paper in the field of exploration of diversity of wild relatives, and subsequent articles (e.g., Khoury *et al.* 2015) have a high number of viewings by scientists and the public. The national inventory of the United States (Khoury *et al.* 2013) was also covered in American media and received the 2014 Crop Science Society of America C8 Division “Outstanding Papers in Plant Genetic Resources” award.

A number of concrete actions associated with the present research are ongoing. Inspired by the assessment of increasing homogeneity in global food supplies and subsequent research focused on its implications in regard to agricultural development (Khoury & Jarvis 2014), CGIAR international research centers are engaging in wider discussions regarding funding and research priorities. Numerous crop genepool based gap analyses, bringing together a variety of associated researcher experts, are in planning or have been initiated. A cumulative global analysis communicating conservation concerns for over 1100 wild species related to 80 crops is also nearing completion at the time of writing. Building upon the national inventory of the United States, agencies are collaborating on refining knowledge on conservation concerns, performing targeted collecting, and establishing management plans for the conservation of key

iconic wild relatives *in situ*. Finally, the research on interdependence among countries on crop genetic resources is being used as a formal submission to the Plant Treaty in contribution to current negotiations for the potential expansion of scope and membership.

CLOSING REMARKS

How humanity will feed itself into the future remains an open question. Whether our ingenuity will prove successful in finding new ways to innovate with natural resources to overcome nutrition and production challenges as we surpass what appear to be safe planetary boundaries (Rockström *et al.* 2009) remains to be seen. More than six decades ago, Aldo Leopold, a preeminent American ecologist and conservationist, warned of the importance of natural resource conservation to human wellbeing:

“To keep every cog and wheel is the first precaution of intelligent tinkering”
(Leopold 1949).

Sixty years later, it is clear that much work remains to be done to accomplish a comprehensive global system for the conservation, exploration, access to, and use of crop genetic resources. This work is more pressing than ever, given increasing homogeneity in the global food system, the myriad of challenges facing agriculture, and threats to the survival of this diversity in the wild, in farmers’ fields, and in under-funded genebanks. N. I. Vavilov’s oft cited statement on the urgency of these efforts is still poignant:

“Time is short, time is short, there is so much to do” [N. I. Vavilov, as recorded in Cohen (1991)].

REFERENCES

- Araújo, M.B., Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *J. Biogeogr.* 33 (10), 1677-1688.
- Bradford, R. (1931). *John Henry* (Harper & Brothers).
- Brar, D.S., Khush, G.S. (1997). Alien introgression in rice. *Plant Mol. Biol.* 35 (1-2), 35-47.
- Cao, Q., Zhang, A., Ma, D., Li, H., Li, Q., Li, P. (2009). Novel interspecific hybridization between sweetpotato (*Ipomoea batatas* (L.) Lam.) and its two diploid wild relatives. *Euphytica* 169, 345-352.
- Castañeda-Álvarez, N.P., de Haan, S., Juárez, H., Khoury, C.K., Achicanoy, H., Sosa, C.C., *et al.* (2015). *Ex situ* conservation priorities for the wild relatives of potato (*Solanum* L. section *Petota*). *PLoS One* 10 (4), e0122599.
- Cobben, M.M.P., van Treuren, R., Castañeda-Álvarez, N.P., Khoury, C.K., Kik, C., van Hintum, T.J.L. (2014). Robustness and accuracy of Maxent niche modelling for *Lactuca* species distributions in light of collecting expeditions. *Plant Genetic Resources: Characterization and Utilization*, 1-9. doi: 10.1017/S1479262114000847.
- Cohen, B.M. (1991). Nikolai Ivanovich Vavilov: the explorer and plant collector. *Econ. Bot.* 45 (1), 38-46.
- Costa, G.C., Nogueira, C., Machado, R.B., Colli, G.R. (2010). Sampling bias and the use of ecological niche modeling in conservation planning: a field evaluation in a biodiversity hotspot. *Biodiv. Conserv.* 19, 883-899.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: A global initiative to collect, conserve, and use crop wild relatives. *Agroecol. Sust. Food Syst.* 38 (4), 369-377.
- Dormann, C.F. (2006). Promising the future? Global change projections of species distributions. *Basic Appl. Ecol.* 8, 387-397.
- Endresen, F., Terje, D., Street, K., Mackay, M., Bari, A., De Pauw, E. (2011). Predictive association between biotic stress traits and eco-geographic data for wheat and barley landraces. *Crop Sci.* 51 (5), 2036-2055.
- Esquinas-Alcázar, J. (2005). Protecting crop genetic diversity for food security: political, ethical and technical challenges. *Nat. Rev. Genet.* 6 (12), 946-953.
- FAO (Food and Agriculture Organization of the United Nations). (2010). *The Second Report on the State of the World's Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2002). *International Treaty on Plant Genetic Resources for Food and Agriculture* (Rome: FAO).
- FAO (Food and Agriculture Organization of the United Nations). (2015). CountrySTAT. <http://www.fao.org/economic/ess/ess-capacity/countrystathome/en/>.
- Fitzpatrick, M.C., Hargrove, W.W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodiv. Conserv.* 18, 2255-2261.
- Gaiji, S., Chavan, V., Ariño, A.H., Otegui, J., Hobern, D., Sood, R., *et al.* (2013). Content assessment of the primary biodiversity data published through GBIF network: status, challenges and potentials. *Biodiversity Informatics* 8, 94-172.
- Graham, C., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., *et al.* (2008). The influence of spatial errors in species occurrence data used in distribution models. *J. Appl. Ecol.* 45, 239-247.
- Guarino, L., Lobell, D.B. (2011). A walk on the wild side. *Nat. Clim. Change* 1 (8), 374-375.
- Henry, R.J. (2014). Genomics strategies for germplasm characterization and the development of climate resilient crops. *Front. Plant Sci.* 5, 68.

- Hijmans, R.J., Graham, C. (2006). The ability of climate envelope models to predict the effect of climate change on species distributions. *Glob. Change Biol.* 12, 2272-2281.
- Jarvis, A., Williams, K., Williams, D.E., Guarino, L., Caballero, P.J., Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. *Genet. Resour. Crop Ev.* 52, 671-682.
- Jimenez-Valverde, A., Lobo, J.M., Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Divers. Distrib.* 14, 885-890.
- Kearney, J. (2010). Food consumption trends and drivers. *Philos. T. R. Soc. B.* 365 (1554), 2793-2807.
- Khoury, C., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57 (4), 625-639.
- Khoury, C.K., Bjorkman, A.D., Dempewolf, H., Ramírez-Villegas, J., Guarino, L., Jarvis, A., *et al.* (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci.* 111 (11), 4001-4006.
- Khoury, C.K., Greene, S., Wiersema, J., Maxted, N., Jarvis, A., Struik, P.C. (2013). An inventory of crop wild relatives of the United States. *Crop Sci.* 53 (4), 1496.
- Khoury, C.K., Heider, B., Castañeda-Álvarez, N.P., Achicanoy, H.A., Sosa, C.C., Miller, R.E., *et al.* (2015). Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I. series Batatas*]. *Front. Plant Sci.* 6, 251.
- Khoury, C.K., Jarvis, A. (2014). *The Changing Composition of the Global Diet: Implications for CGIAR Research*. CIAT Policy Brief No. 18 (Cali: Centro Internacional de Agricultura Tropical).
- Krame-Schadt, S., Niedballa J., Pilgrim J.D., Schroder B., Lindenborn J., Reinfelder, V., *et al.* (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19 (11), 1366-1379.
- Leopold, A. 1949. *A Sand County Almanac and Sketches Here and There* (New York: Oxford University Press).
- Lobo, J.M., Jiménez-Valverde, A., Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17 (2), 145-151.
- Loiselle, B.A., Jorgensen, P.M., Consiglio, T., Jimenez, I., Blake, J.G., Lohmann, L.G., *et al.* (2008). Predicting species distributions from herbarium collections, does climate bias in collection sampling influence model outcomes? *J. Biogeogr.* 35(1), 105-116.
- Martin, F.W., Jones, A. (1973). The species of *Ipomoea* closely related to the sweetpotato. *Econ. Bot.* 26, 201-215.
- McCouch, S., Baute, G.J., Bradeen, J., Bramel, P., Bretting, P.K., Buckler, E., *et al.* (2013). Agriculture: feeding the future. *Nature* 499 (7456), 23-24.
- Nimmakayala, P., Vajja, G., Reddy, U.K. (2011). “*Ipomoea*,” in: *Wild Crop Relatives: Genomic and Breeding Resources*, ed. C. Kole (Berlin: Springer), 123-132.
- Pingali, P. (2007). Westernization of Asian diets and the transformation of food systems: implications for research and policy. *Food Pol.* 32 (3), 281-298.
- Popkin, B.M. (2006). Technology, transport, globalization and the nutrition transition food policy. *Food Pol.* 31 (6), 554-569.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. *PloS One* 5, e13497.
- Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F.S., Lambin, E.F., *et al.* (2009). A safe operating space for humanity. *Nature* 461 (7263), 472-475.
- Sander, J.D., Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347-355.

- Sharma, H.C., Pampapathy, G., Reddy, L.J. (2003). Wild relatives of pigeonpea as a source of resistance to the pod fly (*Melanagromyza obtusa* Malloch) and pod wasp (*Tanaostigmodes cajaninae* La Salle). *Genet. Resour. Crop Ev.* 50, 817-824.
- Tapia, C., Torres, E., Parra-Quijano, M. (2014). Searching for adaptation to abiotic stress: ecogeographical analysis of highland Ecuadorian maize. *Crop Sci.* 55 (1), 262-274.
- Tester, M, Langridge, P. (2010). Breeding technologies to increase crop production in a changing world. *Science* 327 (5967), 818-822.
- Upadhyaya, H.D., Reddy, K.N., Pundir, R.P.S., Singh, S., Gowda, C.L.L., Irshad Ahmed, M. (2013). Diversity and geographical gaps in *Cajanus scarabaeoides* (L.) Thou. germplasm conserved at the ICRISAT genebank. *Plant Genetic Resources* 11 (1), 3-14.
- Vavilov, N.I. (1926). Studies on the origin of cultivated plants. *Bulletin of Applied Botany, Genetics and Plant Breeding* 16, 1-248.
- Wilkes, G. (2007). Urgent notice to all maize researchers: disappearance and extinction of the last wild teosinte population is more than half completed. A modest proposal for teosinte evolution and conservation *in situ*: The Balsas, Guerrero, Mexico. *Maydica* 52, 49-58.
- Wisn, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A., and NCEAS Predicting Species Distributions Working Group. (2008). Effects of sample size on the performance of species distribution models. *Divers. Distrib.* 14, 763-777.
- Zamir, D. (2001). Improving plant breeding with exotic genetic libraries. *Nat. Genet.* 2, 983-989.
- Zurita-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S.-E., Schwember, A.R. (2014). Breeding quinoa (*Chenopodium quinoa* Willd.): potential and perspectives. *Mol. Breed.* 34 (1), 13-30.

Summary

The challenges to long-term global food and nutrition security are complex and compounding. The increasing availability of energy-dense foods worldwide is reducing stunting and other measures of undernutrition, but this nutrition transition demonstrates mixed success in resolving micronutrient deficiencies, and the over-consumption of macronutrients is contributing to the global surge in diet-related non-communicable diseases. Our growing population and dietary expectations are projected to increase demand on food systems for at least the next four decades, outpacing current yield trends. Limitations in land, water, and natural resource inputs, competition for arable soils with non-food crops and other land uses, the need to minimize harmful impacts on biodiversity and other ecosystem services, and greater climatic variability further constrain production potential. Although future gains in food availability may partially be obtained through dietary change and food waste reduction, an increase both in productivity and sustainability on current agricultural lands is necessary. This increase will be achieved through improved agronomic practices combined with the use of varieties of crops with reliable yields under more adverse conditions.

As the source of traits employed for adaptation to biotic and abiotic stresses and for yield increases through breeding, and the palette from which food systems may be further diversified, crop genetic diversity is critical to increased productivity and sustainability. This biological cornerstone of food security is generated through genetic mutation and recombination, and further transformed through natural and artificial selection, and is therefore the product of the evolution over time of crops and closely related wild plants in their agricultural and natural habitats.

Expectations of increases in utilization of crop genetic diversity in order to address production challenges are made under the assumption that adequate variation will be available for exploration. Unfortunately, considerable erosion of crop genetic diversity is occurring through economic development, demographic change, and habitat destruction. Due to the disappearance of this diversity *in situ*, i.e., in farmers' fields and natural habitats, the world's *ex situ* genebanks originally established to make plant genetic diversity readily available to breeders for crop improvement, have become essential repositories for crop diversity conservation, but these too are vulnerable due to insufficient support and resources.

Summary

Two points regarding the current state of the world's heritage of crop genetic diversity are thus clear - much remains to be done to secure the diversity threatened *in situ* and in under-funded genebanks, and the window of opportunity to accomplish a comprehensive global system for genetic resources conservation will not remain open indefinitely.

Among the factors hindering the conservation of crop genetic resources is a lack of essential information regarding this diversity. A number of primary questions have not been fully answered, including: (a) what is the status of diversity in our food systems, and where are the greatest vulnerabilities?, (b) where can genetic diversity be found that can be useful in increasing productivity and mitigating these vulnerabilities?, (c) is this genetic diversity available in the present and in the long term?, and (d) if not available, what steps are needed to improve the ability for researchers to access genetic resources critical for present and future crop improvement?

This thesis aims to contribute to the knowledge required to answer these questions through an exploration of the need for, potential of, challenges and constraints regarding, and necessary steps to enhance the conservation and use of crop genetic resources. The research starts with an investigation of the current state of diversity in global food supplies (Chapter 2). This also represents an exploration into a longstanding two-part assumption in the field of genetic resources – that humanity relies on relatively few crops for its survival, and that this list of important crops is growing smaller over time.

In order to understand the state of diversity in global food supplies, trends over the past 50 years in the richness, abundance, and composition of crop species in the national food supplies of countries worldwide were assessed. Over this period national per capita food supplies expanded in total quantities of food calories, protein, fat, and weight, with increased proportions of those quantities sourcing from energy-dense foods. At the same time, the number of measured crop commodities contributing to national food supplies increased, the relative contribution of these commodities within these supplies became more even, and the dominance of the most significant commodities decreased. As a consequence, national food supplies worldwide became more similar in composition, correlated particularly with an increased supply of a number of globally important cereal and oil crops, and a decline of other cereals, oil crops, and starchy root and tuber species. The increase in homogeneity worldwide portends the

establishment of a global standard food supply, which is relatively species rich in regard to measured crops at the national level, but species poor globally. This research contextualizes the need for the use of diverse genetic resources in crop breeding in order to mitigate vulnerability created by greater homogeneity in global food supplies.

The thesis then delves into the potential for utilization of a particular set of genetic resources of increasing interest globally – crop wild relatives. These weedy and wild cousins of cultivated species have been used for many decades in order to introduce traits of value to crops through breeding, particularly for pest and disease resistance. Their use in crop improvement is likely only to increase because (a) information is improving regarding the identity, potential, and conservation status of these plants, and digital data platforms are making this information more readily available, (b) advancements in geographic information systems, due to increasingly high resolution ecological data as well as evolving models and methods, are enabling a more comprehensive conceptualization of the geographic distribution and conservation status of wild species, (c) rapidly progressing classical and genomic tools, technologies, and methods are facilitating their use in crop breeding, and (d) there is a growing interest in the use of exotic genetic diversity in order to address compounding agronomic challenges.

Crop wild relatives are genetic resources at the nexus of a number of critical global challenges. They have the potential to contribute significantly to crop improvement, helping to address food security and development goals, while improving the adaptation of crops to climate change. At the same time, as wild plants they are subject to a myriad of human caused threats to natural ecosystems, including habitat modification, urbanization, mining, logging, changing fire regimes, pollution, invasive species, overharvesting, and climate change. A focus on wild genetic resources is thus timely both for conservation and food security objectives.

Research on crop wild relatives first concentrates on the identification of potentially important wild genetic resources at the national level (Chapter 3). Focusing on the United States of America, a large and eco-geographically diverse country with relatively advanced conservation policy and active national genetic resources conservation efforts, the chapter presents an inventory of crop wild relatives and other wild species of potential use in agricultural research. The chapter also develops a method for prioritizing these wild species based upon their potential to contribute to food security. The resulting National Inventory listed 4,600 taxa from 985

Summary

genera and 194 plant families, including wild relatives of potential value via breeding as well as wild species of direct use for food, forage, medicine, herb, industrial, ornamental, and environmental restoration purposes. Crop wild relatives were found to be related to a broad range of important crops. Some potentially valuable species are threatened in the wild, and few accessions of such taxa are currently conserved *ex situ*. The prioritization identified 821 taxa from 69 genera primarily related to major food crops, with emphasis on approximately 285 native taxa from 30 genera that are most closely related to such crops.

Once potentially valuable crop genetic resources are identified, subsequent information is needed regarding where they occur, what diversity they may possess, and how well conserved and therefore available to crop breeders they are. Chapter 4 offers a ‘gap analysis’ methodology to answer these questions at the crop genepool level, i.e., for the wild relatives associated with any particular crop, with a case study on the wild relatives of bean (*Phaseolus* L.). This chapter capitalizes on developments in the generation of and access to digital occurrence and eco-geographic data as well as improvements in modeling wild plant species distributions, intentionally utilizing freely available software and data. The method also includes a novel expert evaluation methodology, using researchers knowledgeable in the distributions and conservation concerns of crop wild relatives to assess the results. Of 85 assessed taxa, over half were found to be highly under-represented in genebanks. Priority areas for collecting were identified, particularly in central Mexico.

Chapters 5 and 6 represent advances on the basic methodology outlined in Chapter 4, taking advantage of improvements in species targeting, occurrence data, modeling, and expert feedback methods. The chapters also take an additional step by utilizing eco-geographic information to indicate the potential for species and specific populations to possess traits of value to crop improvement, particularly for abiotic stress tolerance. The chapters focus on crops differing substantially in the state of existing information regarding associated wild relatives, as well as historical use of the resources in crop improvement.

The potential for use of the crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I.* series *Batatas*] is constrained by uncertainty in regard to species boundaries and their phylogenetic relationships, the limited availability of germplasm with which to perform crosses, and the difficulty of introgressing genes from wild species. Chapter 5 modeled the distributions

of the fourteen species that are considered the closest wild relatives of sweetpotato, and found that currently designated species differed among themselves and in comparison with the crop in their adaptations to temperature, precipitation, and edaphic characteristics, and most species also showed considerable intraspecific variation. With 79% of species identified as high priority for further collecting, these crop genetic resources were found to be highly under-represented in *ex situ* conservation systems and thus inadequately available to breeders and researchers. Taxa and specific geographic locations were prioritized for further collecting, particularly in diversity hotspots in Mesoamerica and in the extreme southeastern United States, in order to improve the completeness of germplasm collections.

The use of crop wild relatives in pigeonpea [*Cajanus cajan* (L.) Millsp.] breeding has been successful in providing important resistance, quality, and breeding efficiency traits. Current breeding objectives for pigeonpea include increasing its tolerance to abiotic stresses, including heat, cold, drought, and waterlogging. Like the sweetpotato study, Chapter 6 found considerable variation among the fifteen wild relatives of the crop in regard to adaptations to climatic and soil conditions. Likewise, the research assessed that these wild genetic resources are broadly under-represented in *ex situ* conservation systems, with 80% of species identified as high priority for further collecting. Species and geographic locations particularly in southern India and northern Australia were highlighted for further collecting in order to improve the completeness of germplasm collections, with particular emphasis on collecting and conserving populations possessing tolerance to abiotic stresses.

While conservation of crop genetic resources is fundamental to the availability of this diversity for breeding, it is not the only major constraint to utilization. National and international policies on crop genetic resources determine the real capacity for researchers to acquire diversity of potential interest. The thesis culminates in an exploration of the implications of the global geographic distribution of crop genetic diversity for food security, in particular the level to which international collaboration is required in order to achieve access to genetic resources where they are needed.

The research presented in Chapter 7 demonstrated that rich historical areas of crop genetic diversity occur across the tropics and subtropics, extending into temperate regions in both hemispheres. National food systems are thoroughly interconnected worldwide in regard to the

Summary

geographic origins of crop diversity. Countries are highly dependent on crops whose genetic diversity largely sources from outside their borders in their food supplies (68.7% as a global mean across food variables) and in their national production systems (69.3%). This reliance is evident even in countries located in regions of high indigenous crop diversity and has increased significantly over the past half century, bolstering evidence for the need for effective national and international policies to promote genetic resource conservation and exchange.

Chapter 8 provides a general discussion of the research and its main results. The chapter offers a summary of the current status of conservation of the crop genetic resources targeted in the thesis, identifying limitations in the study and suggesting future research in order to make further progress on key questions in the field. The chapter also discusses the impact of the research to date, and highlights ongoing activities that are building upon the efforts documented here. Increasing awareness and information about the critical role of crop genetic diversity in overcoming nutrition and production challenges offer a renewed opportunity to establish a more comprehensive global system for its conservation and availability for use. Collecting, improving conservation, performing further taxonomic, breeding, and associated research, and resolving the politics of access to genetic resources are urgently needed in order to maximize the potential for crop genetic diversity to contribute to improving food security.

Publications of author

Scientific journals

- Castañeda-Álvarez, N.P., de Haan, S., Juárez, H., Khoury, C.K., Achicanoy, H., Sosa, C.C., *et al.* (2015). *Ex situ* conservation priorities for the wild relatives of potato (*Solanum* L. section *Petota*). *PLoS One* 10 (4), e0122599.
- Cobben, M.M.P., van Treuren, R., Castañeda-Álvarez, N.P., Khoury, C.K., Kik, C., van Hintum, T.J.L. (2014). Robustness and accuracy of Maxent niche modelling for *Lactuca* species distributions in light of collecting expeditions. *Plant Genetic Resources: Characterization and Utilization*, 1-9. doi: 10.1017/S1479262114000847.
- Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: a global initiative to collect, conserve and use crop wild relatives. *Agroecol. Sust. Food Syst.* 38 (4), 369-377.
- Khoury, C.K., Bjorkman, A.D., Dempewolf, H., Ramírez-Villegas, J., Guarino, L., Jarvis, A., *et al.* (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proc. Natl. Acad. Sci.* 111 (11), 4001-4006.
- Khoury, C.K., Castañeda-Álvarez, N.P., Achicanoy, H., Sosa, C.C., Bernau, V., Kassa, M.T., *et al.* (2015). Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: Distributions, *ex situ* conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* 184, 259-270.
- Khoury, C.K., Greene, S., Wiersema, J., Maxted, N., Jarvis, A., Struik, P.C. (2013). An inventory of crop wild relatives of the United States. *Crop Sci.* 53 (4), 1496.
- Khoury, C.K., Heider, B., Castañeda-Álvarez, N.P., Achicanoy, H.A., Sosa, C.C., Miller, R.E., *et al.* (2015). Distributions, *ex situ* conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [*Ipomoea batatas* (L.) Lam., *I. series Batatas*]. *Front. Plant Sci.* 6, 251.
- Khoury, C.K., Laliberté, B., Guarino, L. (2010). Trends in *ex situ* conservation of plant genetic resources: a review of global crop and regional conservation strategies. *Genet. Resour. Crop Ev.* 57 (4), 625-639.
- Ramírez-Villegas, J., Khoury, C.K. (2013). Reconciling approaches to climate change adaptation for Colombian agriculture. *Clim. Chang.* 119, 575-583.
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D.G., Guarino, L. (2010). A gap analysis methodology for collecting crop genepools: a case study with *Phaseolus* beans. *PLoS One* 5 (10), e13497.

Book chapters, policy briefs and other published articles

- Arnaud, E., Castañeda-Álvarez, N., Dulloo, E., Drucker, A., de Haan, S., Jarvis, D., *et al.* (2014). *Global Information System for In Situ Conservation and On-farm Management of Plant Genetic Resources for Food and Agriculture*. Input Paper for the ITPGRFA Consultation on the Global Information System on Plant Genetic Resources for Food and Agriculture (COGIS-PGRFA) (Rome: FAO).
- Guarino, L., Khoury, C., Fowler, C. (2011). “Adapting to climate change: the importance of *ex*

- situ* conservation of crop genetic diversity,” in: *Food security and Climate Change in Dry Areas: Proceedings of an International Conference*, eds. M. Solh, M.C. Saxena, 1-4 February 2010, Amman, Jordan [Aleppo: International Center for Agricultural Research in the Dry Areas (ICARDA)].
- Hunter, D., Guarino, L., Khoury, C., Dempewolf, H. (2011). “A community divided: lessons from the conservation of crop wild relatives around the world,” in: *Agrobiodiversity Conservation: Securing the diversity of Crop Wild Relatives and Landraces*, eds. N. Maxted, E. Dulloo, B. Ford-Lloyd, L. Frese, J Iriondo (Wallingford: CAB International).
- Khoury, C.K., Achicanoy, H.A., Bjorkman, A.D., Navarro-Racines, C., Guarino, L., Flores-Palacios, X., *et al.* (2015). *Estimation of Countries’ Interdependence in Plant Genetic Resources Provisioning National Food Supplies and Production Systems*. International Treaty on Plant Genetic Resources for Food and Agriculture, Research Study 8 (Rome: FAO).
- Khoury, C.K., Dempewolf, H., Castañeda-Álvarez, N.P. (2014). Wild Relatives. *The Scientist*. June 2014.
- Khoury, C., Guarino, L. (2010). Back to the roots: wild genes for food security. *Samara* 18/19, January-December 2010 (London: Kew Publications), 1-2.
- Khoury, C.K., Jarvis, A. (2014). *The Changing Composition of the Global Diet: Implications for CGIAR Research*. CIAT Policy Brief No. 18 (Cali: Centro Internacional de Agricultura Tropical), 1-6.
- Khoury, C., Laliberté, B., Guarino, L. (2009). *Trends and Constraints in Ex Situ Conservation of Plant Genetic Resources: A Review of Global Crop and Regional Conservation Strategies* (Rome: Global Crop Diversity Trust), 1-56.

Thesis

- Khoury, C. (2007.) *An Ecogeographic Survey and Conservation Analysis for Arid North American Phaseolus L. (Fabaceae)*, Sections *Acutifolii*, *Coriacei*, *Minkelersia* and *Rugosi Species* (Birmingham: University of Birmingham), UK MSc. thesis.

Publications coordinated

- Global Strategy for the Ex Situ Conservation of Chickpea (Cicer L.)* (2008). (Rome: Global Crop Diversity Trust).
- Global Strategy for the Ex Situ Conservation of Faba Bean (Vicia faba L.)* (2009). (Rome: Global Crop Diversity Trust).
- Global Strategy for the Ex Situ Conservation of Lentil (Lens Miller)* (2008). (Rome: Global Crop Diversity Trust).

Media articles featuring research (selected works)

- Media articles regarding Khoury et al. (2014)* Proc. Natl. Acad. Sci.
- Baragona, S. (2014). Global diets growing similar. *Voice of America*. 3 March 2014.
- Becker, M. (2014). Food security declines with global crop diversity. *Conservation Magazine*. 17 March 2014.
- Charles, D. (2014). In the new globalized diet, wheat, soy and palm oil rule. *National Public*

Radio. 4 March 2014.

Cherfas, J. (2014). The global standard diet. *Eat this podcast*. 31 March 2014.

Doyle, A. (2014). Global diets get more similar in threat to food security: study. *Reuters*. 3 March 2014.

Doyle, A. (2014). Global diets get more similar in threat to food security: study. *Chicago Tribune*. 3 March 2014.

Doyle, K. (2014). Decreasing diversity in global food supplies threatens crop production and nutrition. *Crop Science Society of America*. 5 March 2014.

Graber, C. (2014). Lower diet diversity threatens crops and us. *Scientific American*. 10 March 2014.

Kinver, M. (2014). Crop diversity decline 'threatens food security'. *BBC News*. 3 March 2014.

MacVean, M. (2014). Global food supply grows increasingly homogeneous, study says. *Los Angeles Times*. 3 March 2014.

Mintz, S. (2014). World food security at risk as variety of crops shrinks. *Thomson Reuters Foundation*. 4 March 2014.

Muller, G. (2014). 'Global average diet' threatens food security. *Australian Broadcasting Corporation, Bush Telegraph*. 14 March 2014.

Roach, J. (2014). Increasingly similar global food supply poses risks, Study Says. *NBC News*. 3 March 2014.

Stuart, H. (2014). A narrowing global food supply will make us all fat, sick and less safe, study says. *Huffington Post*. 6 March 2014.

Tran, M. (2014). Growing reliance on fewer crops will increase risk of drought and disease. *The Guardian*. 4 March 2014.

Treat, J., Scalomogna, A. (2014). Diet similarity- we'll have what they're having. *National Geographic Magazine*. September 2014, p. 41.

Vence, T. (2014). Crop coalescence. *The Scientist*. 5 March 2014.

Walsh, B. (2014). Our global diet is becoming increasingly homogenized - and that's risky. *Time Magazine*. 4 March 2014.

Media articles regarding Khoury et al. (2013) Crop Sci.

Fisher, M., Fisk, S. (2013). U.S. a surprisingly large reservoir of crop plant diversity. Press release. 29 April 2013.

Gewin, V. (2013). US takes stock of valuable "weeds". *Frontiers in Ecology and the Environment*. Dispatches. The Ecological Society of America.

Suszkiew, J. (2014). National inventory takes stock of crops' wild relatives. *USDA Agricultural Research*. January 2014.

Media articles regarding crop wild relative initiatives

Braw, E. (2013). World food day: the search for sustainable crops. *The Guardian*. 16 October 2013.

Fisher, M. (2012). Crop wild relatives and their potential for crop improvement. *CSA News*. May 2012, p. 4-10.

Fisher, M., Fisk, S. (2013). Protecting the weedy and wild kin of globally important crops. Press

Publications of author

release. American Society of Agronomy (ASA), Crop Science Society of America (CSSA), and Soil Science Society of America (SSSA) International Annual Meetings, Nov. 3-6 in Tampa, Florida.

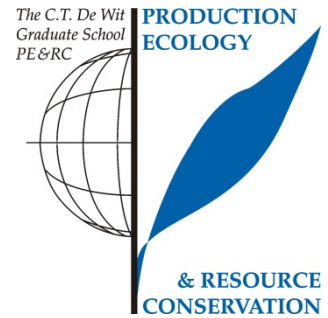
Gewin, V. (2013). Weeds warrant urgent conservation. *Nature News*. 22 July 2013.

Gilbert, N. (2010). Searching for wild relatives. *Nature News Blog*. 10 December 2010.

Kew. (2013). Major global analysis offers hope for saving the wild side of staple food crops. Press release. 22 July 2013.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review of literature (6 ECTS)

- Adapting agriculture to climate change: collecting, protecting and preparing crop wild relatives

Writing of project proposal (4.5 ECTS)

- Adapting agriculture to climate change: collecting, protecting and preparing crop wild relatives

Post-graduate courses (4.5 ECTS)

- Contemporary approaches in genetic resources conservation and use: integrated and participatory approaches in agro-biodiversity management; WUR (2010)
- Multivariate analysis; WUR (2011)

Invited review of (unpublished) journal manuscript (2 ECTS)

- *Evolutionary Applications*: The red queen and seed bank: pathogen resistance of ex situ and in situ conserved barley (2011)
- *Journal of Plant Breeding and Crop Science*: Selection of yam seeds from open pollination for adoption in yam (*Dioscorea rotundata* Poir) production (2012)
- *Plant Genetic Resources: Characterization and utilization*: Identification of gaps in the world collection of pigeonpea germplasm from East and Southern Africa conserved at the ICRISAT genebank (2013)
- *Economic Botany*: Factors influencing diversity of farmers' varieties of sweetpotato in Uganda: implications for conservation (2014)
- *Genetic Resources and Crop Evolution*: Evaluation and improvement of the ecogeographical representativeness of a collection of the genus *Trichloris* in Argentina (2014)
- *Journal for Nature Conservation*: Private reserves in Brazil: distribution patterns, logistical challenges and conservation contributions (2015)

Deficiency, refresh, brush-up courses (0.6 ECTS)

- Introduction to R (2011)

Competence strengthening / skills courses (1.5 ECTS)

- Ethics and Philosophy in Life Sciences (E&P LS); WUR (2012)

PE&RC Annual meetings, seminars and the PE&RC weekend (1 ECTS)

- PE&RC Weekend (2010)
- PE&RC Day (2011)

Discussion groups / local seminars / other scientific meetings (4.5 ECTS)

- Modelling and statistics network; WUR (2011-2013)
- GIS Modelling group – CIAT (DAPA SDM group) (2011-2015)

International symposia, workshops and conferences (9 ECTS)

- USDA National Plant Germplasm System (NPGS) Plant Germplasm Operations Committee (PGOC); oral presentation; Washington DC (2011)
- Botany conference (Society for Economic Botany); poster presentations (2011)
- USDA NPGS Curators and PGOC; oral presentation; Spokane WA (2012)
- USDA NPGS Curators and PGOC; oral presentation; Washington DC (2013)
- ASA, CSSA and SSSA International Water, Food, Energy and Innovation for Sustainable World; oral presentation (2013)
- Center for Development Research (ZEF); oral presentation (2014)
- University of Arizona WEES Food Systems Scholars & Native Seeds/SEARCH; oral

PE&RC Training and Education Statement

- presentation; University of Arizona (2014)
- USDA NPGS Curators and PGOC; oral presentation; Davis CA (2014)
- USDA National Program Leaders; oral presentation; Washington DC (2014)
- ASA, CSSA and SSSA International Grand Challenges, Great Solutions; oral and poster presentations (2014)
- International Food Policy Research Institute (IFPRI); oral presentation; Washington DC (2014)

Curriculum vitae

Colin Kahlil Khoury was born on the 14th of March, 1975 in La Jolla, California, USA. Raised at the interface of wild lands, subtropical fruit orchards, and urban areas, his mother regularly phoned the California Poison Control System to inquire as to the toxicity of the wild plants that he often sampled. Satsuma mandarins, pineapple guavas, and avocados remain his cultivated fruits of preference. His feeling of foreboding as natural habitats and agricultural soils are urbanized continues as well. Colin completed two years of undergraduate studies in Biology at Oberlin College before leaving university to pursue an agricultural career full time. He cultivated vegetables, fruits, and grains in Argentina, Mexico, and various parts of the USA before returning to complete a Bachelor of Arts in Environmental Studies with an emphasis in Agricultural Ecology from Prescott College, Arizona, USA in 2000. He then served as a crop curator at Native Seeds/SEARCH, a regional seedbank dedicated to the conservation and celebration of agricultural seeds and their associated traditions adapted to the arid southwestern USA and northwestern Mexico. Following a further year exploring seed diversity movements in Latin America, he returned to California, where he surveyed rare wild plants in the coastal scrub, mountains, and deserts of the region. In 2007, Colin completed a Master of Science in Conservation and Utilisation of Plant Genetic Resources at the University of Birmingham, UK. For his MSc. thesis, he analyzed the distributions and conservation status of crop wild relatives of bean (*Phaseolus* spp.) native to Mexico and the USA. He then served as a scientific associate at the Global Crop Diversity Trust, located in Rome, Italy, where he informed strategies for more comprehensive conservation of crop diversity worldwide. In 2011, he transferred to the International Center for Tropical Agriculture (CIAT), Colombia, to work as a visiting research scientist on a Crop Trust managed project while pursuing his doctoral studies at the Centre for Crop Systems Analysis, C.T. de Wit Graduate School of Production Ecology and Resource Conservation, Wageningen University, The Netherlands. He is a member of the Species Survival Commission, Crop Wild Relative Specialist Group of the International Union for Conservation of Nature (IUCN); the Standing Committee on Science Communications of the Crop Science Society of America; the Crop Science Society of America, Soil Science Society of America, and American Society of Agronomy; the Botanical Society of America; and the California Native Plant Society.

Funding

Funding for this thesis came from the initiative "Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives" which is supported by the Government of Norway. The project is managed by the Global Crop Diversity Trust with the Millennium Seed Bank of the Royal Botanic Gardens, Kew, UK, and implemented in partnership with national and international genebanks and plant breeding institutes around the world (Dempewolf *et al.* 2014). Support also came from the International Center for Tropical Agriculture (CIAT), the CGIAR Research Program on Climate Change, Agriculture, and Food Security (CCAFS), and the CGIAR Research Program on Roots, Tubers, and Bananas (RTB). The Grains Research and Development Corporation (GRDC); the Global Public Goods Phase 2 project of the System-wide Genetic Resources Program (SGRP), Bioversity International; and the World Bank supported the early stages of development of the gap analysis methodology and occurrence data compilation.

Dempewolf, H., Eastwood, R.J., Guarino, L., Khoury, C.K., Müller, J.V., Toll, J. (2014). Adapting agriculture to climate change: a global initiative to collect, conserve and use crop wild relatives. *Agroecol. Sust. Food Syst.* 38 (4), 369-377.