2014 LARGE SCALE APPLIED RESEARCH PROJECT COMPETITION:
GENOMICS AND FEEDING THE FUTURE
APPLICATION FORM

Application Number:
Project Title: Genomics of Abiotic Stress Resistance in Wild and Cultivated Sunflowers
Term of Funding (years): 4
Total Budget Requested: $8,039,771 **Genome Canada Contribution: $3,215,247**

Project Leader		**Co-Project Leader**	
Name[1]	**Loren H. Rieseberg**	Name	**John M. Burke**
Affiliation	**University of British Columbia**	Affiliation	**University of Georgia**
Address	**Botany Department 3529-6270 University Blvd Vancouver, BC V6T 1Z4**	Address	**Department of Plant Biology Miller Plant Sciences Athens, GA 30602 USA**
Telephone	**604-827-4540**	Telephone	**706-583-5511**
E-mail	**lriesebe@mail.ubc.ca**	E-mail	**jmburke@uga.edu**
Date		Date	
Signature[2]		Signature[2]	

Certification Requirements

Applicants proposing to perform research that requires certification (such as research involving human subjects, human stem cells, animals, biohazards, radioactive materials or possible effects on the environment) must obtain the appropriate certification for the proposed project. Certificates are not required to be provided until after the project has been approved. Please check the box(es) below, if the proposed research involves any of the following:

Human subjects		Human stem cells		Animals		Biohazards	x	Environmental assessment	

Lead Organization (CEO, President or authorized representative)

Organization	University of British Columbia
Name	Martin Kirk
Title	Director of Research Services
Date	
Signature[2]	

Genome Centre CEO(s) or authorized representative(s)

Lead Centre [3]	Genome BC	Co-lead Centre (if applicable)	
Name	Alan E. Winter	Name	
Date		Date	
Signature[2]		Signature[2]	
Additional Centre (if applicable)			
Name			
Date			
Signature[2]			

[1] The Project Leader is responsible for the administrative and financial activities of the project.
[2] Signatures confirm acceptance of the terms as outlined in Meaning of Signatures.
[3] The Administrative Centre for projects which have been identified as being co-led by two or more Genome Centres.

TABLE OF CONTENTS

II RESEARCH AREAS AND KEYWORDS

Select the area(s) that relate(s) to the research proposed. If relevant to more than one area, use numbers to indicate the relative weighting (i.e., 1 = primary focus; 2 = secondary focus, etc.)

1	Agriculture/Agri-food		Energy		Environment
	Fisheries		Forestry	2	GE^3LS
	Human Health		Mining		Technology Development

Provide a **maximum** of five (5) words or phrases for each category (i.e., Research, and Methods and Technologies) that describes the proposed investigation.

CATEGORY	KEYWORDS
Research	Abiotic stress resistance, agricultural genomics, ecophysiology, evolutionary genomics, wild and cultivated sunflowers
Methods & Technologies	Association mapping, high-throughput genotyping and phenotyping, gene regulatory analyses, physiology of abiotic stress resistance, population genomic analyses

Provide a **maximum** of five (5) words or phrases for each category (i.e., Research questions, and Methods) that describes the proposed integrated GE^3LS investigation.

CATEGORY	KEYWORDS
Research Questions	Access and benefit sharing, crop yield models, genomic data sharing, international treaty framework, climate change impacts
Methods	Literature review, model development, model simulations, policy analysis, interviews

Sharing of application and reviews

Where applicable, Genome Canada seeks your consent to share the information included in your application and subsequent reviews on a confidential basis with the funding partners referenced in the Request for Applications (RFA), as well as project-specific co-funding partners.

I, the Project Leader **consent** to the sharing of the application and reviews with the organizations indicated above. [X]

I, the Project Leader **do not consent** to the sharing of the application and reviews with the organizations indicated above. [ ]

III RESEARCH TEAM

Please provide in the table below, a list of the research team members (excluding collaborators), their affiliation, role in the project, time commitment to the proposed research and their responsibilities in the context of the project:

Name, Title and Affiliation	Role[4]	Time Commitment (hrs/week)	Description of Responsibilities	Signature[5]
Dr. Loren Rieseberg, Professor & CRC Tier 1, Department of Botany, U. British Columbia	Project Leader	15	Rieseberg is leading overall project and Is lead on Activity 4 (functional analyses). Past experience in managing large genomics projects. Pioneer in plant ecological genomics.	
Dr. John Burke, Professor, Department of Plant Biology, U. of Georgia	Co-Project Leader	15	Burke is the co-project leader and is lead on Activity 1.2 (association mapping). Expert on sunflower genetics & genomics.	
Dr. Lisa Donovan, Head & Distinguished Professor, Department of Plant Biology, U. Georgia	Co-applicant	10	Will lead phenotyping of stress resistance traits (Activities 1.1 and 1.3). Expert on plant ecophysiology.	
Dr. Nicolas Langlade, Senior Researcher, INRA Toulouse, France	Co-applicant	10	Will lead transcriptomics component & Heliaphen phenotyping (Activities 1.4 and1.5). Expert on sunflower molecular physiology.	
Dr. Emily Marden, Research Associate & Lead of Intellectual Property and Policy Research Group, School of Law, U. British Columbia	Co-applicant	10	Will lead GE^{3}LS component on barriers to R&D caused by international treaties (Activity 6). Expert in CBD, Treaty & sharing of genomics data.	
Dr. Brent Hulke, Research	Co-applicant / End User	10	Will lead MAGIC population development	

[4] Role includes: Project Leader, Co-Project Leader, Co-applicant, and End User. Definitions of participant categories are provided in the *Guidelines for Funding Research Projects*.

[5] **Signatures of the co-applicants and End Users are required** and confirm that the application has been reviewed and approved for submission to the Genome Centre and Genome Canada by all investigators. **Signatures of Collaborators are not required.**

Geneticist, USDA ARS Fargo, ND			and phenotyping (Activity 3) & deploy resistance alleles in breeding program. Strong track record in molecular breeding.	
Dr. Sam Yeaman, AIHS/CAIP Chair in computational evolutionary biology (start date: 8/15), U. Calgary	Co-applicant	10	Will lead population genomics component of project (Activity 2). Expert in population genomics & bioinformatics.	
Dr. Navin Ramankutty, Professor & CRC Tier 1, Liu Institute, U. British Columbia	Co-applicant	10	Expert on global land use, especially in relation to agriculture. Will lead GE^{3}LS component on crop yield modeling (Activity 5).	
Dr. Louisa Staton, Research Associate, U. British Columbia	Project Manager	40	Will provide project management & coordination. Expert in plant genetics.	
Dr. Jinhua Xiao, Trait Genomics Leader for sunflowers, Dow Agrosciences	Collaborator / End User /	5	Will contribute to MAGIC population development & will deploy resistance alleles in breeding program. Strong track record in molecular breeding.	
Dr. Marie Coque, Sunflower Project Lead, Biogemma	Collaborator / End User / Co-funder	5	Will contribute to MAGIC population development & will deploy resistance alleles in breeding program. Strong track record in molecular breeding.	
Dr. Silke Wieckhorst, Project Lead for molecular breeding in sunflower, KWS Seeds	Collaborator / End User / Co-funder	5	Will contribute to MAGIC population development & will deploy resistance alleles in breeding program. Strong track record in molecular breeding.	
Dr. Andrés Zambelli, Biotechnology Manager, Advanta Semillas SAIC	Collaborator / End User / Co-funder	5	Will contribute to MAGIC population development & will deploy resistance alleles in breeding program. Strong track record in molecular	

			breeding.	
Dr. Bill May, Crop Management Agronomist, Agriculture and Agri-Food Canada	Collaborator / End User	5	Will contribute to field-based salt tolerance phenotyping, economic modeling & will deploy resistance alleles in breeding program. Strong track record in agronomy.	
Dr. Walter Anyanga, Research Officer, National Semi-arid Resources Research Institute, Uganda	Collaborator / End User	10	Will contribute to field-based low nutrient stress phenotyping & deploy resistance alleles in breeding program. Strong track record in sunflower breeding.	
Dr. Jim Gerdes, Global R&D Director for Sunflower Breeding, NuSeed Americas	Collaborator / End User / Co-funder	5	Will contribute to MAGIC population development & evaluation & will deploy resistance alleles in breeding program. Strong track record in molecular breeding.	
Dr. Laura Marek, Oil Seeds Crops Curator, USDA-ARS	Collaborator	5	Will contribute to collection and distribution of sunflower germplasm	
Dr. Pedro Andrade-Sanchez, Assistant Professor, Agricultural and Biosystems Engineering, U. Arizona	Collaborator	5	Will adapt field-based HTP approaches to sunflower. Expert in agricultural engineering.	
Dr. Khaled Bali, Irrigation Advisor & Interim Director, DREC	Collaborator	5	Will contribute to drought stress phenotyping at DREC. Expert in water use management.	
Dr. Wayne Parrott, Professor, Crop and Soil Sciences, U. Georgia	Collaborator	5	Will conduct genome editing in soybean. Expert in crop biotechnology.	
Dr. Timo Kubach, Vice President, Portfolio & New Business | Product & Innovation, SAP AG, Germany	Collaborator / Co-funder	5	Will coordinate SAP effort to develop data analytic tools for plant genomics, using sunflower as model.	

IV PARTICIPATING ORGANIZATIONS' SIGNATURES

To be completed by the organizations in which the research will be undertaken.

The following organizations have reviewed and approved this application and agree to respect the general principles guiding the use of Genome Canada funds, specific guidelines on eligible costs and co-funding, and the specific conditions associated with the Release of Genome Canada funds, as outlined in the *Request for Applications: 2014 Large-Scale Applied Research Project Competition Genomics: Feeding the Future*, and *Guidelines for Funding Research Projects* including adherence to commonly accepted guidelines with respect to ethical, environmental and safety requirements.

In addition, the following organizations agree to respect applicable policy and program guidelines of other funding agencies, which are identified as sources of co-funding in this application.

Organization	Name & Title of Authorized Representative	Signature	Date dd/mm/yy
University of British Columbia	Martin Kirk, Director of Research Services		
University of Georgia			
INRA Toulouse, France			
USDA-ARS, Fargo, ND			
University of Calgary			
Biogemma			
KWS Seeds			
Advanta Semillas SAIC			
National Semi-arid Resources Research Institute, Uganda			
NuSeed Americas			

SAP AG, Germany			

V LAY SUMMARY

In a maximum of **one-half (1/2) page**, describe the proposed research in a way that can be understood by a lay audience. Be sure to indicate how the proposed research will result in social and/or economic benefits. This summary may be used by Genome Canada to inform the public and other stakeholders about the value of genomics research.

Plants are regularly challenged by a variety of environmental stresses such as drought, flooding, salt, and low-nutrient levels. These stresses negatively affect plant growth and reduce the productivity of our crops. Though wild plants have evolved mechanisms to meet these challenges, many crops are less resilient. To reduce stress-induced yield loss and improve food security, attention has increasingly turned to the development of stress-resistant crops. However, these efforts require an improved understanding of the mechanisms allowing plants to resist these stresses. Such knowledge will enable the development of crops capable of growing in previously unsuitable habitat and thus feed a rapidly growing population in the context of an increasingly variable climate, particularly as marginal farmlands are brought into production.

We propose to investigate the molecular and physiological basis of drought, flooding, salt, and low-nutrient stress resistance in cultivated sunflower and reproductively compatible, stress-adapted wild species that are potential donors of beneficial resistance traits. Sunflower is an ideal system for the proposed research because this $20 billion oilseed crop is clearly limited by such stresses, while wild relative species are adapted to a variety of extreme environments.

Ultimately, the proposed research would achieve several objectives: (1) identify and fully characterize the genetic basis of stress resistance traits in sunflower; (2) create germplasm resources that would enable sunflower breeders to put stress-resistant, high-yield cultivars in the field within four years of project end; (3) develop crop yield models to predict likely yields of the new stress resistant sunflower cultivars in different soil and climate conditions across Canada; and (4) develop strategies for reducing barriers to R&D caused by international treaties. Expanded sunflower production made possible by the new cultivars is expected to yield ~$12 million USD annually within five years of the project end date and up to ~$230 million USD annually after ten years. Stress-resistant cultivars will stabilize production in the face of environmental stresses in Sub-Saharan Africa, reducing the potential for famine. Worldwide impacts will be substantial, as no other oilseed can maintain stable yields across as wide a range of environmental conditions as can improved sunflower cultivars.

VI RESEARCH SUMMARY

In a **maximum of one (1) page**, summarize the proposed research, including integrated GE^{3}LS research activities. Describe the deliverables expected at the end of the project and the social and/or economic benefits anticipated from their practical application.

Food production must increase by 70 to 100% by 2050 to keep pace with predicted population growth and changes in diet. This task is exacerbated by ongoing changes in climate and heightened competition for land and water. To meet this challenge, crops must be developed that combine high yield with resistance to biotic and abiotic stress, and require lower inputs. Here we propose to identify traits and alleles that will allow such cultivars to be developed and to deliver these alleles to breeding programs. Our focus is on sunflower, a globally important oilseed with production valued at $20B/year, but we will apply new genome editing methods to assess whether our discoveries can be applied to the improvement of other oilseeds.

Sunflower is ideal for the proposed research because of its diverse extremophile and cross-compatible wild relatives, which can be exploited for research and breeding. Because sunflower is grown widely in developing countries for food, it is the only oilseed in the Crop Trust's list of 25 priority food security crops. Canada is the world's 13th largest exporter of sunflower, but there is a supply deficit in North America of this healthy vegetable oil. Canadian production is limited by salt and flooding, while low nutrients and drought limit sunflower production in Sub-Saharan Africa (and worldwide). Wild plants have mechanisms to mitigate these challenges, but crops are far less resilient. System-level understanding of stress resistance can facilitate the development of improved varieties that can be grown on marginal farmlands currently unsuitable for crops.

The goals of this project are to: (1) assess resistance to drought, flooding, salt, and low nutrient stress in sunflower and its wild relatives using traditional and high-throughput phenotyping approaches; (2) associate variation in abiotic stress resistance with specific genes, regulatory networks, and/or causal variants; (3) determine the mechanistic basis of stress resistance via physiological and transcriptomic analyses; (4) address major biological questions concerning the types of genes, their network positions, and the nature of physiological trade-offs involved in the evolution of abiotic stress resistance; and (5) identify suitable stress resistance alleles for use in sunflower breeding programs and potentially for improvement of other oilseeds. Our GE^{3}LS research will ensure realization of social and economic benefits of this project by (a) developing yield models for sunflower under different environmental conditions; and (b) addressing negative impacts of international treaties on the use of plant genetic resources.

Deliverables are: (1) "next generation" germplasm resources (resistance alleles with minimal trade-offs in relevant genetic backgrounds), enabling sunflower breeders to put resistant, high-yielding cultivars in the field within four years of project end; (2) a central data mining and analysis resource for sunflower to facilitate research and breeding; (3) crop yield models that will enable predictions of likely yields of new stress resistant sunflower cultivars in different soil and climate conditions across Canada; and (4) strategies for mitigating barriers to R&D (e.g., uncertainties in IP, tech transfer, and profit sharing) caused by international treaties.

Five years out, if just 0.2% of Canadian farmlands affected by related soil problems were dedicated to stress resistant sunflower cultivars, this would add 4,000 hectares to production, with a gross production value ~$12 million USD annually. After ten years, up to $230 million USD in gross annual returns are expected, and an export market for this increase exists. Stress-resistant cultivars will increase production in the face of abiotic stress in Sub-Saharan Africa, reducing malnutrition and associated economic costs, which have been estimated at 5% of GDP in Uganda. Worldwide impacts will be substantial, as no other oilseed can maintain stable yields across as wide a range of environmental conditions as can improved sunflower cultivars.

The PIs are world leaders in plant genomics and have generated most of the genomic resources available for sunflower. End-users involved in the project include sunflower breeders from Canada, the USA, EU, S. America, and Uganda. The GE^{3}LS team has critical expertise developed during the Genome Canada-funded Genomics of Sunflower project.

VII RESEARCH PROPOSAL
Including Research on Ethical, Environmental, Economic, Legal and Social Aspects of Genomics (GE^{3}LS)

Maximum thirty (30) pages, including references, charts, figures and tables. This section must address all relevant evaluation criteria for the competition.

Using a Gantt chart, show project activities, milestones and the timelines for reaching them. Attach the Gantt chart to the end of this section. Please note that the Gantt chart is not included in the page limits above. A template that can be used is attached in Appendix VII.

A. Introduction

Over human history, the domestication and improvement of crop plants has achieved spectacular increases in yield over the wild plants utilized by hunter-gatherers, enough to support a population explosion and the rise of modern civilization[1,2]. However, in striving for maximal yield, the genetic changes caused by domestication were not entirely beneficial. Crop plants often exhibit reduced stress resistance as compared to their wild relatives[3-5], perhaps due to selection for performance in optimal environments and to loss of resistance via trade-offs with productivity under ideal conditions[6], and/or stochastic allele loss during the narrowing of their genetic bases[7,8]. This loss of resistance is especially troubling in the 21st century, as we struggle to increase crop productivity in the face of climate change, rapid population growth, and heightened competition for land and water[9].

To reduce stress-induced yield loss and improve food security, efforts have increasingly focused on the development of stress-resistant crops. However, such crop improvement efforts require knowledge. What is the genetic basis of stress resistance and related traits? Can alleles be found within crop or wild gene pools that confer resistance across broad genetic backgrounds, but with minimal yield trade-offs? What is the influence of factors such as the position of a gene within a regulatory network on the occurrence and severity of possible trade-offs? How often are the same mutations, genes, and gene networks exploited by natural selection in the evolution of stress resistance across different species? We propose to address these basic biological questions through an investigation of the genomic and physiological basis of resistance to drought, flooding, salt, and low nutrient stress in wild and cultivated sunflowers. This is an ideal study system for this work because the productivity of cultivated sunflower is limited by such stresses[10,11], while it retains reproductive compatibility with wild relatives that are adapted to a variety of extreme environments, including desert, swamp, saline, and sand dune habitats.

The availability of cost-effective and efficient genomic technologies sets the stage for linking genomic information to agriculturally relevant phenotypes. However, it remains challenging to identify traits that are most important for stress resistance, and to assess those traits (or proxies) on sufficiently large numbers of plants to permit genetic dissection[12-15]. Another limitation is that, as has long been recognized, there are often trade-offs between plant productivity and stress resistance [16-19]. Inherently fast-growing plants are expected to have relatively little in the way of stress resistance, and this has been hypothesized to be due to trade-offs (genetic constraints) between resource acquisitive traits conferring growth potential and resource conservative traits often associated with stress resistance[20,21]. However, there is evidence that some stress resistance traits have the potential to evolve independently of general resource use and growth potential or yield[17,18,22-24]. The existence of suitable variation in such traits would make possible the development of genotypes that exhibit both high yield potential and stress resistance.

Mechanistically, drought resistance can be achieved by a spectrum of overlapping strategies, including escape, avoidance, and physiological tolerance[25-27]. However, in agricultural situations, the greatest gains are likely to be achieved by selecting primarily for traits that allow plants to avoid declines in plant water status by maximizing water uptake and minimizing water loss per unit carbon gain[11,14,15,28]. These traits include deeper rooting, higher root:mass ratios (root

biomass/total biomass), lower transpirational water loss, and higher water-use efficiency. In addition, osmotic adjustment can facilitate avoidance by enabling continued root growth and access to soil moisture. For cultivated sunflower, there is substantial variation in many of these traits, and no necessary trade-off between yield potential and the yield of stressed crops[11,23,29-38]. Salt and low nutrient stresses have some commonalities with each other, as well as with drought stress. Salt stress includes both osmotic stress (i.e., physiological drought) due to the effect of salts on substrate water potential, and ionic stress caused by the accumulation of potentially toxic Na ions[22,39]. A trait such as osmotic adjustment could contribute to both salt and drought stress resistance, whereas others (e.g., restricting Na uptake, sequestration, or excretion) might specifically mitigate ion toxicity effects[39,40]. Resistance to low nutrient availability may involve the development of more extensive root systems and increased uptake capacity, greater efficiency of N use for photosynthesis, and/or greater ability to extract N from senescing leaves[18,41-43]. The extent of overlap in the genetic architecture of these traits is unknown.

Plant adaptation to flooding stress frequently involves traits that can be antagonistic to drought, salt, and low nutrient stress. Flooding reduces available O_2 and CO_2, impacting respiration and photosynthesis, respectively [44]. If flooding is short in duration – as frequently experienced by cultivated sunflowers in North America – successful adaptive strategies include those that conserve energy such as limited root grown and down-regulation of nonessential processes, adventitious rooting to avoid submergence, as well as metabolic adaptations like anaerobic energy production [45-47]. We will be most interested in the latter two adaptations since they have previously been reported for wild sunflowers from intermittently flooded habitats[48], and such traits seem least likely to negatively impact resistance to other stresses that limit sunflower production.

For large, population-level studies of trait variation, multiple approaches are needed to meet the challenges of phenotyping relevant ecophysiological traits[12,49]. For example, leaf carbon isotope ratio ($\delta^{13}C$) can be used to estimate a leaf level physiological trait, water-use efficiency integrated over the lifetime of the leaf, which is close to impossible to measure for large numbers of plants[14,23,50,51]. Likewise, sensor-based high-throughput phenotyping (HTP) systems have emerged as promising non-destructive approaches[49,52-56]. For field-based applications, for example, sensors to measure canopy reflectance, temperature, and height can be mounted on a high-clearance vehicle and combined with real-time kinematic GPS to provide georeferenced data with cm-level accuracy[12,52,57]. This provides a means of estimating indices related to growth such as the Normalized Difference Vegetation Index (NDVI) and the Crop Water Stress Index (CWSI) for a large number of genotypes. These base-line data also can be used to estimate leaf area index (LAI), canopy biomass, and transpiration, providing insight into more fundamental physiological differences. The combination of such phenotyping strategies with the high-throughput genotypic characterization of experimental populations makes possible the efficient genetic dissection of agriculturally relevant phenotypes[12].

A complementary approach to determining the genetic basis of abiotic stress resistance is the use of comparative population genomic data to identify adaptively important genes[58-62]. Such approaches, which rely on the sampling and genomic characterization of populations across habitat transitions and the association of allelic variation with relevant environmental variables, do not require the up-front collection of phenotypic data. As such, they serve as a useful complement to more intensive, trait-based approaches and provide an avenue for the efficient discovery of putatively adaptive alleles from materials that might otherwise not be evaluated – e.g., range-wide collections of wild species that have been shaped by many generations of natural selection.

While it is clear that crop wild relatives represent important sources of new variation for agriculture, the use of this variation (and innovative germplasm derived from it) is currently limited because of ambiguities in the International Treaty for Plant Genetic Resources in Food and Agriculture. Both genomic and GE^{3}LS research will be needed to facilitate the development and widespread use of environmentally resilient crops.

The goals of this project (Figure 1) are to:

1) Assess abiotic stress resistance (drought, flooding, salt, and low nutrients) in sunflower and compatible wild species using a combination of conventional phenotyping and sensor-based HTP approaches;
2) Associate variation in abiotic stress resistance and related traits in cultivated and wild sunflower with specific genes, regulatory networks, and/or causal genetic variants using genetic map-based and population genomics approaches;
3) Determine the mechanistic basis of stress resistance via in-depth physiological and transcriptional characterization of lines identified as being divergent in stress resistance in our large-scale, population-level phenotypic screens;
4) Identify suitable abiotic stress resistance alleles (i.e., those exhibiting resistance in a variety of genetic backgrounds, ideally with minimal trade-offs) for improvement of sunflower cultivars for Canadian and international markets and subsistence farming – and potentially for improvement of other oilseeds such as soybean; and
5) Develop, characterize, and distribute "next generation" germplasm resources in the form of multi-species, advanced generation intercross (MAGIC) populations that will facilitate the efficient genetic analysis of complex trait variation in *Helianthus* and enable the efficient deployment of exotic alleles in breeding programs.

Our GE^{3}LS research will support these goals by **(a)** developing yield models for sunflower and other Canadian crops that, in combination with existing economic models, will enable predictions of likely yields of new stress resistant cultivars in different soil and climate conditions across Canada; and **(b)** designing strategies for addressing significant negative impacts of international treaties on the use of plant genetic resources by private and public sector breeding programs in Canada and worldwide as part of our GE^{3}LS research.

Our efforts align well with the Request for Applications (RFA – http://www.genomecanada.ca/en/portfolio/research/2014-competition.aspx). Our project addresses:

"challenges that are relevant both in the Canadian context and also to Canada's international development priorities by providing sustainable, genomic-based solutions to problems of agriculture … in developing countries. For example, research on how plants can better respond to abiotic and biotic stresses has the potential for considerable impact on global food security with applications in Canada and the developing world." (RFA, p. 4).

In addition, the genomic information and resistance alleles provided by our project are expected to contribute to the following focal topics of the RFA (p. 5): meeting the caloric demands of population growth, improving crop health, adapting crops to climate change, and lessening their footprint on the environment (by reducing need for external inputs such as fertilizers or water).

Our project was designed in collaboration with end-users (sunflower breeders and researchers) representing four government agencies (Bill May – Agriculture and Agri-Food – AAFC Canada; Nicolas Langlade – the French National Institute for Agricultural Research – INRA; Walter Anyanga – the National Semi-Arid Resources Research Institute – NaSSARI, Uganda; Brent Hulke – the U.S. Department of Agriculture – USDA) and five companies (Andrés Zambelli – Advanta Semilis, Argentina; Marie Coque – Biogemma, France; Jinhua Xiao – Dow AgroSciences, USA; Silke Wieckhorst – KWS Seeds, Germany; and Jim Gerdes – NuSeed Americas, USA) who are also contributing to the project through provision of promising cultivars for inclusion in multi-parent advanced generation intercross (MAGIC) mapping populations (see below), assistance with the crosses needed to develop these populations, field-based phenotyping, and co-funding. The extensive involvement of sunflower breeders, combined with the financial backing of the Global Crop Diversity Trust for the developing country component of

information generated by the project, as well the stress resistance alleles we identify, are efficiently deployed in breeding programs.

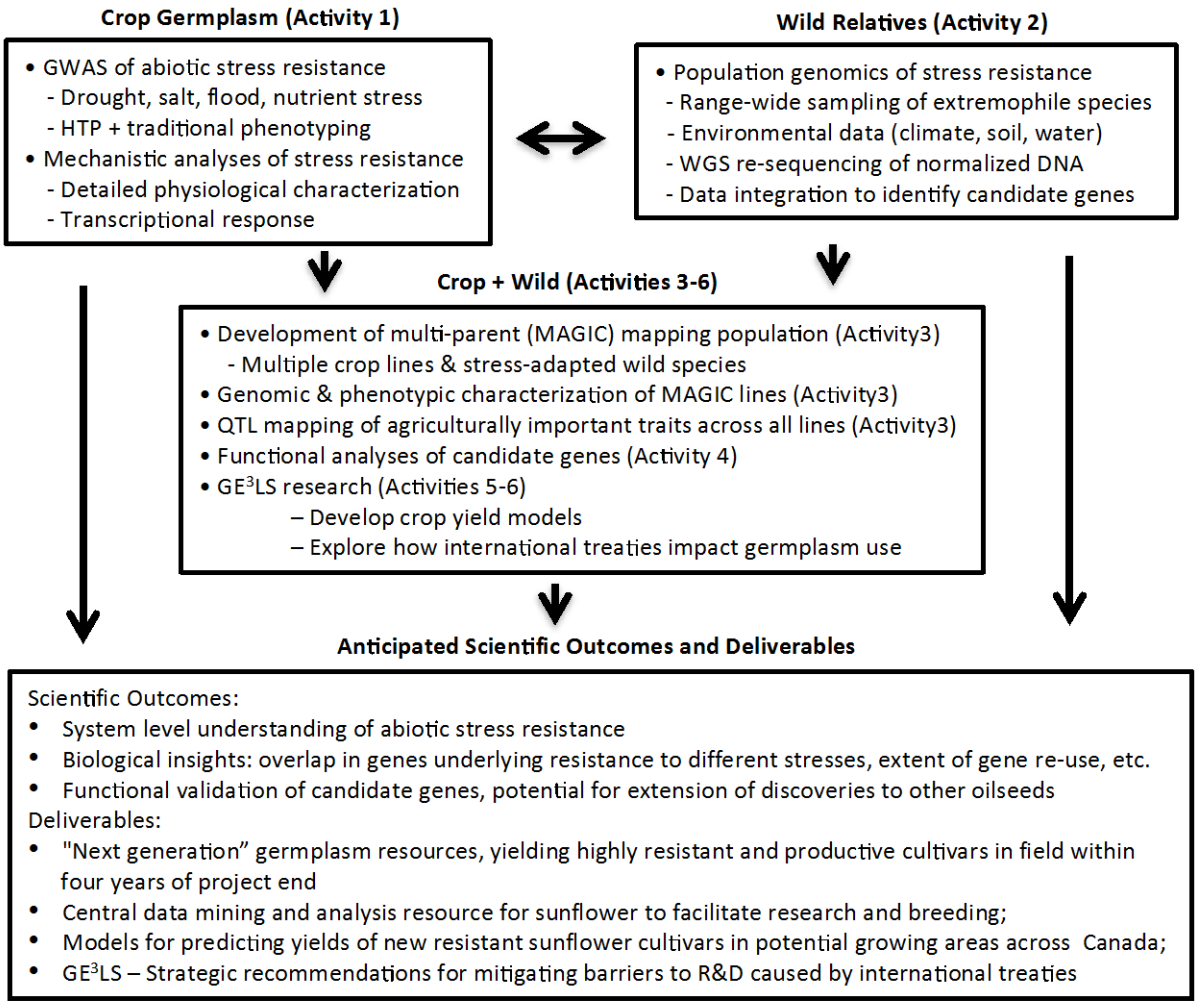

Figure 1. Overview of proposed activities, anticipated scientific outcomes, and deliverables

B. Study System

Sunflower (*Helianthus annuus* L.) is one of the world's most important oilseed crops, with production valued at ca. $20B/year (http://faostat.fao.org). Because sunflower is grown widely in developing countries and used primarily for food, it is the only oilseed included in the Global Crop Diversity Trust's list of 25 priority food security crops (https://www.croptrust.org/crop-diversity-endowment-fund/). Sunflower is also an important ornamental crop and source of confectionery seeds – the current focus of Canadian production. Canadian production is now limited by salt and flooding (yield loss is 100% when flooding or extreme salt stress occurs[63]), while low nutrients and drought limit sunflower production in Sub-Saharan Africa (and worldwide) and will become increasingly important in the southern prairies of Canada with climate change[64].

Sunflower was domesticated from the common sunflower (also *H. annuus*) ca. 4,000 years ago in what is now the east-central USA[65-70]. It is the most economically important member of the genus *Helianthus,* which includes 49 wild species[71], all native to North America. Many of the wild species are cross-compatible with the cultivated sunflower, occur in extreme environments including desert sand dunes, salt marshes, serpentine soils, coastal beaches, etc., and therefore represent important reservoirs of valuable resistance alleles. Moreover, the genus as a whole has emerged as a model for evolutionary and ecological analyses, including studies of hybridization, speciation, and adaptation[41,72-83].

Sunflower is amenable to both classical and molecular genetic analyses. Individual plants can readily produce over a thousand seeds with a generation time of 3-5 months, and excellent genomic resources (http://www.sunflowergenome.org) are available to support the proposed project. These resources, developed by the PIs as part of previously funded Genome Canada and NSF Plant Genome projects, include numerous genetic mapping populations, SNP and expression arrays, a sequence-based physical map, ultra-dense genetic maps, an extensive collection of EST/transcriptome and gene space sequencing data, and a high quality reference sequence[84-89]. The reference genome for sunflower was generated by merging independent assemblies of 454 and Illumina sequence and then further scaffolding of the merged assembly with mate pair libraries and BAC-end sequences at progressively more local scales (using our genetic and physical maps) to reduce the likelihood of mis-assembly. The resulting integrated assembly is similar to the expected genome length (3.64 Gb versus 3.6 Gb expected), with an N50 of 210Kb. The genome includes >98% of CEGMA (Core Eukaryotic Genes Mapping Approach) genes, of which ~90% are full length, indicating that the gene space is well covered. The annotated assembly, which includes ~39k strongly supported protein-coding gene models (excluding transposable elements), is displayed in JBrowse and is accompanied by numerous tools for searching, mapping, and functional analyses. In addition to the reference sequence, the Genome Canada project supported the development of high-density genetic maps and draft genome assemblies of two wild species (*H. argophyllus* and *H. petiolaris*), as well as the whole-genome shotgun (WGS) sequencing of 486 cultivated and wild sunflower genotypes, including individuals from 13 wild *Helianthus* species and our full association mapping population (see below). Lastly, it is also worth noting that cultivated sunflower can be transformed with *Agrobacterium tumefaciens*[90-92], making possible the detailed functional characterization of candidate genes.

Sunflower germplasm collections comprise ~40,000 cultivated and wild accessions globally and have served as the basis for our development of a cultivated sunflower association mapping (SAM) population. This population is composed of 288 inbred lines that capture ca. 90% of the allelic diversity in the sunflower gene pool[93]. Genotypic characterization of this population using a 10k SNP array[85] revealed that linkage disequilibrium (LD; the non-random association of alleles across loci) decays rapidly across much of the genome[94]. Thus, while there are some restricted islands of elevated LD, this population provides us with a powerful tool for investigating the genetic basis of a wide variety of traits in a highly diverse population, and with much higher resolution than typically afforded by traditional genetic map-based approaches.

As mentioned above, we recently completed WGS sequencing of this population, which resulted in the discovery of 72 million high quality SNPs. Of these, 4.1 million were retained after removing SNPs with a minor allele frequency of <10% and <30 % missing data. Notably, preliminary analyses of the SAM population using the 4.1 million SNPs have revealed numerous significant genetic associations for a variety of traits relevant to the present proposal, including salt tolerance (measured as Geometric Mean Productivity growth of reproductive biomass across high and low salt concentrations – see Activity 1), as well as leaf $\delta^{13}C$, leaf N, and leaf mass per area (LMA) measured under optimal conditions (Fig. 2). As we had hoped, in many cases associations can be narrowed down to a single gene (Fig. 2) and can have sizable effects, ranging from 6% PVE for salt tolerance to 30% PVE for LMA.

C. Rationale and Significance

Wild and cultivated plants are challenged by a variety of abiotic stresses on an ongoing basis. Such stresses affect plant growth and development and reduce crop productivity. Though wild plant populations have evolved a number of mechanisms that mitigate the effects of variable and suboptimal growing conditions, many crop plants exhibit less resilience. This is likely due to the occurrence of unacceptable (from an agricultural perspective) trade-offs between resistance traits and overall growth and productivity as well as the chance loss of adaptive variation during the domestication and breeding process. An improved understanding of the mechanisms

underlying abiotic stress resistance is thus needed as we seek to develop crops capable of feeding a rapidly growing population in the context of an increasingly variable climate, particularly as marginal lands are brought into production.

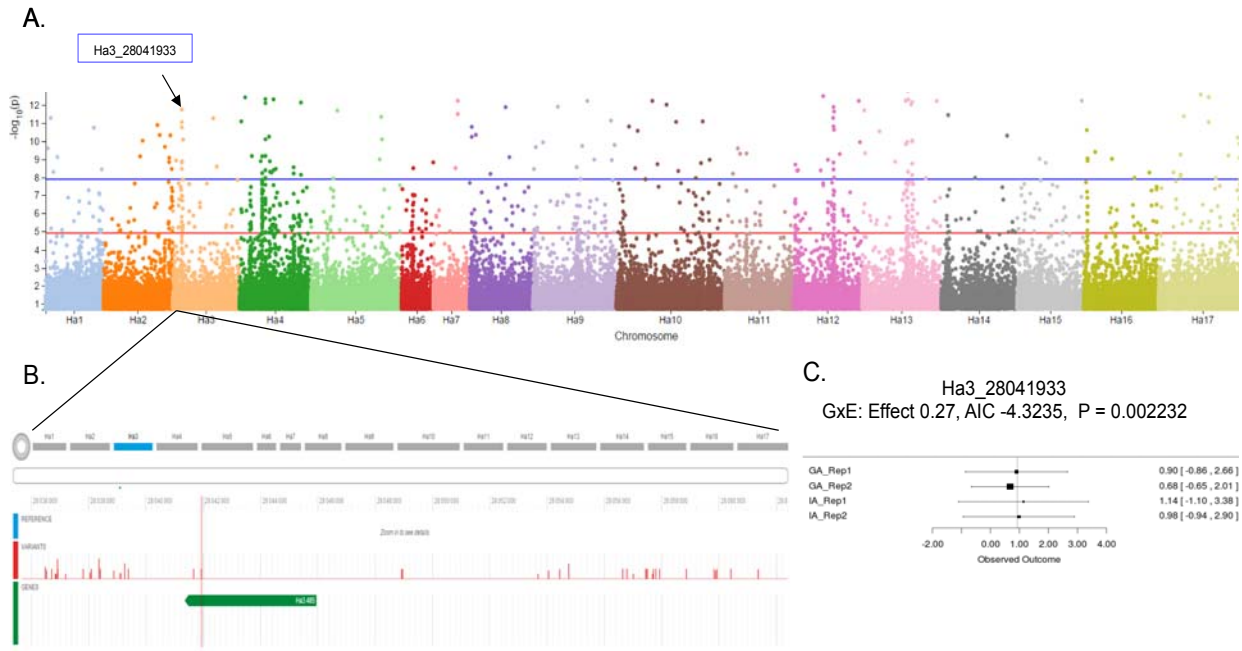

Figure 2. Association mapping of leaf mass per area (LMA) using 4.1M SNPs genotyped across the SAM population. Phenotyping was conducted in Georgia (GA) and Iowa (IA): A) Manhattan plot for association meta-analysis across replicates and environments. Horizontal lines correspond to adjusted p-value using Bonferroni (blue) and simpleM (red) methods at significance level of α=0.05. B) Zoom-in on chromosome Ha3 at position 28041933bp using the genome browser tool developed with SAP. Vertical red bar corresponds to the position of gene Ha3_28041933 (Ubiquitin system component Cue protein), red track corresponds to alternative allele frequency in the SAM population, green track corresponds to genes in that region. C) Forest plot for the effect of the environment at Ha3_28041933. A mixed linear model was used with the environment as a fixed effect and replicates as random effects. Effect = environment effect at SNP in Z units. Left column is the environment and replicate name; central column is the effect of Ha3_28041933 in each environment plus SE, vertical dashed line is the average across environments; left column is the effect of Ha3_28041933 in each environment (in Z units) and CI (95%).

Here, we propose to investigate the genomic and physiological basis of drought, flooding, salt, and low nutrient stress resistance in cultivated sunflower and reproductively compatible, stress-adapted wild species that represent potential donors of beneficial alleles. This work, which makes heavy use of resources developed under prior Genome Canada support, will involve genotypic and phenotypic characterization of a diverse collection of cultivated sunflower lines to enable genome-wide association studies (GWAS), detailed physiological and transcriptomic analyses of resistant and susceptible lines to investigate the mechanistic basis of variation in stress resistance, and population genomic analyses of related species to identify natural variants that confer stress adaptation in the wild. This knowledge will move us closer to a systems-level understanding of the interactions between plants and their environment, will enable the development of smarter and more efficient strategies for breeding environmentally resilient cultivars in sunflower, and has the potential to positively impact other crops.

Throughout the course of our research, we will study fundamentally important biological issues that will both improve our understanding of the natural world and inform ongoing crop improvement efforts. This will include knowledge of the genes, traits, and regulatory networks that are responsible for variation in resistance to the focal stresses, and identification of genomic

factors that influence the nature and extent of physiological trade-offs between stress resistance and performance (i.e., yield) under ideal conditions. One might expect, for example, that changes at so-called "hub" genes (i.e., those with more central positions within networks) would be more likely to incur physiological trade-offs than genes located near the periphery of a network, similar to expectations under the centrality-lethality hypothesis that has been developed in the context of protein interaction networks[95]. We will also determine the origin(s) and extent of evolutionary re-use of the same genes/alleles across species in pursuit of solutions to similar environmental challenges. This latter point will allow us to address evolutionary questions regarding the repeatability of genotypic evolution and the role of hybridization in evolution, and will also provide critical insight into the degree to which such genetic/physiological solutions to abiotic stress might apply across species.

In addition to addressing fundamental scientific questions, we will develop and characterize "next generation" germplasm resources in the form of multi-species, advanced generation intercross populations, as well as bioinformatics tools for exploiting these resources. These populations and associated bioinformatics tools will not only facilitate the efficient genetic analysis of complex trait variation in *Helianthus*, but they also will enable the efficient deployment of exotic alleles in breeding programs. Using a backcross breeding design and marker-assisted selection, we expect new sunflower cultivars to be in the field within four years of project end, with significant economic benefits through mitigation of year-to-year risk due to flooding and drought within the existing sunflower growing regions, expansion of sunflower production onto marginal lands, and important social benefits through increased food security in Uganda and other developing countries. Oilseeds such as sunflower represent a concentrated source of energy and essential fatty acids, but per capita availability in developing countries is approximately 40% of the minimum recommended by the Food and Agriculture Administration of the United Nations[96].

Concomitantly with the genomic research, we will work closely with scientific researchers, the Secretariat of the International Treaty for Plant Genetic Resources in Food and Agriculture, industry stakeholders, and legal and policy experts to develop strategies for navigating and resolving the ambiguities in interpretation of the Treaty. In addition to benefiting this project, the GE^{3}LS work will also be of considerable interest in the legal/policy community as these are issues that face stakeholders for other crops.

D. Experimental Approaches

Activity 1: Investigating abiotic stress resistance in cultivated sunflower

To assess the genetic and physiological basis of abiotic stress resistance in cultivated sunflower, we will phenotype stress resistance and related traits in separate, large-scale screens for drought, flooding, salt, and low nutrient stress, followed by GWAS to determine the genetic basis of variation in those traits. We will then select lines that exhibit notable differences in resistance to each stress for further characterization in a series of additional experiments designed to better understand the mechanistic basis of the stress resistance under multiple stress levels relevant to field conditions.

Activity 1.1 – Phenotyping in large-scale screens

We will perform large-scale screens of cultivated sunflower (SAM population, 288 lines) for stress resistance and associated traits in separate experiments for drought, flooding, salt, and low nutrients. As far as possible, replicate screens will be conducted in the field and greenhouse (or growth room) for each stress. Greenhouse and field studies are complementary. Greenhouse conditions reduce environmental variance and allow greater precision in stress levels applied, whereas field conditions provide a more robust link between stress resistance and yield under more realistic agricultural conditions. For each screen (drought, flooding, salt, or low nutrients), we will implement two treatments (control and stressed). The intensity and timing of the stress will be chosen based on the results of preliminary studies that identify a level of stress that will

reduce growth and/or yield by an average of ~30% and provides the ability to distinguish performance among lines. For each screen, the relative stress resistance of each line will be characterized as Geometric Mean Productivity (GMP; sqrt(Y_{CO*} Y_{ST})), where Y_{CO} is yield or biomass in the control (i.e., non-stressed) treatment, and Y_{ST} is yield or biomass in the stressed treatment[33]. For each screen, the mean trait value of each line in each treatment, and stress resistance calculated across treatments, will be used in the association mapping (below). Additional analyses such as regressions, principle components analyses, and structural equation modeling will be used to assess the relationships among traits and the relationships of traits to stress resistance for each screen.

Drought stress

The SAM population will be assessed for drought resistance in field-based evaluations in years 1 and 2 at the University of California Desert Research & Extension Center (DREC; http://drec.ucanr.edu). DREC is an experimental farm that provides normal cultural practices such as land preparation, planting, irrigation, fertilization, cultivation, and pest control. The average precipitation at DREC of 15 mm from April to August is well below the 600-1000 mm generally required by sunflower[10]. This dry climate, combined with our extensive experience with irrigation management and soil moisture monitoring, ensures that we can produce the necessary conditions needed to assess the differential responses of lines within the population. This experiment will have 2 watering treatments (well watered, mild drought), 288 lines, and 4 replicate plots per line. The experimental design will be a split plot design, where replications are in complete blocks nested within the drought stress and control treatments, respectively. Within each replication in flood and control, we will have a 17x17 partially balanced lattice design, such that 17 entries will be within each incomplete block and some permutations of the lattice will not be used. Plots will be two rows, 6.2 m in length with 1.5 m alleys between plots, planted at 150% of intended population density. Row width will be 0.75 m. Dates of emergence will be recorded. After the V2 stage is reached on 100% of the plots[97], plants will be thinned to a consistent stand of approximately 58 plants/plot (62,500 plants ha^{-1}). In each replicate and treatment, extra plots will be added for a subset of 20 lines that represent the diversity in the SAM population for destructive plant sampling to validate sensor measurements. We will plant at the beginning of April and impose a mild drought from 30 days after planting onward. Based on prior experience, we anticipate that an appropriate level of drought will be achieved by irrigating at 60% of potential evapotranspiration, but this will be confirmed by preliminary trials. Irrigation will be applied within plots so that a high-clearance tractor for phenotyping can be driven to take data at any time. Irrigation will be controlled with gated pipes, and water meters will record irrigation amounts. Plants will be phenotyped with high-throughput, sensor-based approaches as well as traditional methods (see below). At physiological maturity, all plots will be harvested. Several representative heads from each line and replicate plot will be used to estimate yield, achene size estimated as 100-achene weight, and seed oil content[98,99]. Yield will be used as the integrated measure of performance for estimating stress resistance, calculated as GMP.

For the field trials at DREC, sensor-based HTP approaches will be used to evaluate leaf and canopy traits using a tractor-mounted array of sensors developed by engineers from the Precision Agriculture Program at the University of Arizona[57]. The sensors include Pulsar dB3 ultrasonic transducers and Blackbox 130D level controllers to measure canopy height; Apogee SI-121 infra-red radiometers of 36 degree view angle to measure canopy temperature to determine CWSI; and active, multispectral radiometers (model ACS-470; Holland Scientific) for measurement of canopy light reflectance in three bands (670, 720, and 820 nm) for determination of NDVI and CWSI. Additional instrumentation in the field platform includes two Campbell Scientific CR-3000 data loggers for data acquisition, and a Hemisphere GPS ultra-precise real-time kinematic receiver model A320 for data geo-referencing. Starting from three weeks after germination, sensor measurements will be taken weekly to estimate canopy height, temperature, NDVI, CWSI, and LAI for each line[12,52,57]. Settings for platform travel speed and instrumentation sampling frequency will be selected to achieve a spatial resolution of at least half the in-row plant spacing, currently planned to be 20 cm. Bi-weekly destructive plant

sampling from the sensor-validation plots will be used to measure plant height, leaf area, and biomass to characterize plant growth and validate their relationships with sensor measurements. Data from destructive sampling and sensor measurements will then be used to estimate integrated traits such as stress degree-days, leaf area duration, and the rate of height increase.

The traits estimated from sensors will be complemented by visual assessment of dates of phenology (emergence, bud formation, anthesis, and maturity), traditional phenotyping on individual plants for leaf traits, and a rapid visual characterization of root architecture[100,101]. After new leaves have been produced for drought-stressed plants, and parallel to the phenomics, one representative leaf (recently fully expanded) per line, treatment, and replicate plot will be destructively sampled for leaf traits related to plant vigour and growth strategies, including LMA, leaf thickness, integrated photosynthetic water-use efficiency estimated from leaf carbon isotope ratio ($\delta^{13}C$), and leaf N (CHN analysis), leaf area and shape (ImageJ)[14,16,18,21,50,51,102-105]. Due to cost, the leaves will be bulked by line and treatment for $\delta^{13}C$ and N.

The evaluation of drought resistance at DREC will be complemented by evaluation of seedling drought resistance in growth rooms at University of Georgia (UGA) for all 288 lines with 3 biological replicates per line and 2 treatments (control and stressed). This custom protocol, developed jointly by personnel in the Burke and Donovan labs as a modification of other systems using solid substrates and PEG[106,107], has been found to produce a differential drought response across sunflower lines. Newly germinated seedlings will be individually transplanted into 50 mL Falcon tubes of sand saturated with either water (control) or PEG 6000 solution inducing a mild osmotic stress of -0.25 MPa (determined using a vapour pressure osmometer, Wescor). Treatments will be applied for 10 days (sufficient time to produce new leaves), after which seedlings will be harvested and assessed for biomass allocation (roots, stems, and leaves), number of root laterals, stem height and diameter, number of true leaves, and leaf N and $\delta^{13}C$ (bulked by line and treatment). Biomass will be used as the integrated measure of performance for calculating stress resistance.

Flooding, salt, and low nutrient stress

Field-based evaluations of the SAM population for flooding, salt, and low nutrient stress (Table 1) will be conducted in years 1 and 2 in Fargo (North Dakota, USA), Indian Head (Saskatchewan, Canada), and NaSARRI (Serere District, Uganda), respectively. We will employ the same basic experimental design, planting, and thinning strategy as that detailed above for the drought tolerance screens at DREC: two treatments (control, stress), 288 lines, 4 replicate plots per line, split block design, thinning at the V2 stage, and a final density of 58 plants/plot. During year one we will employ conventional phenotyping approaches, while in year 2 we will build on our experience at DREC to extend the HTP approaches to the Fargo and Indian Head sites.

For flooding stress, co-PI Hulke has identified several flat, homogenous fields near Fargo that can be irrigated by sprinkler. Following thinning at the V2 stage, water will be added to the stress treatment until the soil is fully saturated. Growth stage at the onset of stress will be recorded for each plot as described by Schneiter and Miller[97]. The stress treatment will be maintained for two weeks, and the following traits will be measured using conventional methods: percent survivorship, extent of adventitious root growth, height of growth, leaf traits (same as measured for drought tolerance, listed above), visual assessment of dates of phenology (anthesis and maturity), as well as yield and oil traits. In addition, plants will be assessed for malate accumulation: malate concentrations are very low in flooded sunflowers from dry habitats due to cessation of respiratory mitochondrial activity[48].

For salt stress, we will target fields near Indian Head for which soil salinity has been previously shown to reduce sunflower yields (B. May, pers. comm.), as well as similar nearby fields that lack soil salinity problems. Plants will be screened for the same set of traits targeted for flooding tolerance (above), except that we also will examine leaf ion characteristics (Na, K, Mg, Ca). For both Fargo and Indian Head, yield will be directly determined on the entire plot using a Kincaid

8XP small plot combine (Kincaid Mfg., Haven, KS, USA) with a Harvestmaster High-Capacity Grain Gage (Juniper Systems, Logan, UT, USA), which measures total harvest weight and grain moisture using onboard probes, and records data directly on a laptop or handheld device. Oil content will be determined by the Nuclear Magnetic Resonance (NMR) method using an Oxford MQC (Oxford Instruments, Abingdon, Oxfordshire, UK) calibrated with appropriate standards to determine oil content on a 10% moisture basis with low error.

The screens for low nutrient stress will be conducted on sandy soils near National Semi-Arid Resources Research Institute (NaSARRI) in Uganda, where nutrient stress is known to limit sunflower production. Here the stress treatment will have no nutrients added, whereas for the control we will add a standard top dressing of fertilizer, equivalent to 90 kg N per ha. We will assay largely the same set of traits targeted for the flooding and salt treatments, except for adventitious roots. Also, we will analyze leaf N (bulked by line and treatments) instead of leaf ion concentrations. Yield and oil will be determined in a manner similar to the drought tolerance studies

Table 1. Large-scale phenotyping of SAM population.

Screen(s)	Location[a]	Phenotypes
Drought stress	Field, DREC[a]	Leaf and canopy traits, phenological traits, leaf N and $\delta^{13}C$, yield traits, seed oil traits, and root architecture
Drought stress[b]	Growth rooms, UGA	Biomass allocation (roots, stems, and leaves), root architecture, stem height and diameter, number of true leaves, and leaf N and $\delta^{13}C$
Flood stress	Field, Fargo, ND	Germination, survivorship, adventitious rooting, malate concentration, relative height growth, leaf traits, phenological traits, yield traits, seed oil traits
Flood stress	Greenhouse, UBC	Germination, survivorship, adventitious rooting, malate concentration, leaf characteristics, leaf N, biomass allocation
Salt stress	Field, Indian Head, SK	Germination, survivorship, relative height growth, adventitious rooting, leaf traits, leaf ion concentrations (Na, K, Mg, Ca), phenological traits, yield traits, seed oil traits
Salt stress[b]	Greenhouse, UGA	Relative height growth, leaf characteristics, leaf N, biomass allocation, and leaf ion concentrations (Na, K, Mg, Ca)
Low nutrient stress	Field, NaSARRI[c]	Relative height growth, leaf traits, leaf N, phenological traits, yield traits, seed oil traits
Low nutrient stress[b]	Greenhouse, UGA	Relative height growth, leaf characteristics, leaf N, biomass allocation, N resorption proficiency and efficiency

[a]DREC = California's Desert Research & Extension Center.
[b]Two screens will be performed, one at the seedling stage and one at the late vegetative stage.
[c]NaSARRI = National Semi-Arid Resources Research Institute, Uganda.

The SAM population will also be assessed for flooding, salt, and low nutrient stress resistance in in the greenhouse in years 1 and 2 (Table 1), with biomass as the integrated measure of performance for calculating stress resistance. For salt and low nutrient stress, there will be two screens, one at the seedling stage (through first full leaf), and one at the late vegetative stage, whereas for flooding stress there will be a seedling screen only since flooding stress typically only occurs early in the growing season. For each stress and developmental stage, the SAM lines will be exposed to two treatments (control and stressed) in a randomized split-plot design with a minimum of three replicates per line. For the flooding treatment a fourth replicate will be added to allow destructive sampling. Flooding stress will be imposed by propagating seedlings in fully saturated soils, whereas the salt stress level will be 75 mM NaCl, which, based on preliminary studies, results in differential growth for cultivated *H. annuus* with little lethality[108,109]. For the low nutrient stress, preliminary studies with a subset of SAM lines will be used to

determine the level of fertilizer for the low nutrient stress treatment. Traits to be measured on all plants in all of the screens include seed germination and survival, relative height growth, leaf characteristics of a recently expanded leaf (leaf chlorophyll (SPAD-502 meter, Konica Minolta), NDVI (PS-300 spectroradiometer, Apogee Inst.), area, shape, LMA, toughness (penetrometer), and N (bulked by line and treatment), and biomass allocation at harvest (roots, stems, leaves). In addition, plants will be assessed for advantageous root growth and malate accumulation in the flooding resistance screen. For the salt stress and low nutrient screens, plants will be assessed additionally for leaf ion concentrations (Na, K, Mg, Ca)[110,111] and leaf N resorption proficiency and efficiency[42,104].

Activity 1.2: Genome-wide association studies of abiotic stress resistance and related traits

Following the phenotypic characterization of the 288 cultivated lines in the SAM population, we will use GWAS to identify alleles associated with variation in abiotic stress resistance and related traits in sunflower, and to determine their potential suitability for use in marker-assisted and/or genomic selection. This phase of our research will make use of the phenotypic data collected in our large-scale field and greenhouse studies of drought, salt, and low nutrient stress (above), as well as genotypic information extracted from available WGS re-sequencing data (details below). We favour this approach for developing a more complete understanding of the genetic basis of intraspecific variation in abiotic stress resistance, because association mapping offers far greater resolution than traditional QTL-based approaches, and also allows us to simultaneously investigate the effects of multiple alleles per locus across a range of genetic backgrounds.

The actual resolution afforded by association mapping ultimately depends on the structure of LD in the population of interest, which itself can be influenced by a number of genetic and non-genetic factors[112]. In sunflower, LD is known to decline quite rapidly across much of the genome[94,113,114]. As such, we anticipate being able to map functional variation with a high level of precision, down to the level of one or a few genes in many cases (e.g., Fig. 2). Of course, this rapid decline in LD means that we will need very high marker density. As part of the previous Genome Canada project, we re-sequenced all individuals within the focal population to a minimum of 8-10x depth. Using custom bioinformatics pipelines developed in collaboration with the software company SAP, we extracted 4.1 million SNPs with a minor allele frequency of >10% (see **study system**), which can be used for association mapping. Detection of structural variants (SVs) and copy number variants (CNVs), which will also be employed for association mapping, is underway. The most time-consuming algorithms in the processing pipeline, from raw reads to variant calling, are implemented in SAP HANA's in-memory engine and optimized for performance (Fig. 3). After reads are trimmed and cleaned, they are aligned against the sunflower reference genome. A first set of variant candidates is called using a maximum-posterior-probability-genotyping algorithm based on the alleles found in the reference genome and the reads overlapping the variant candidates. To remove false positives due to false alignment calls, the reads are fed into the GATK indel realigner (https://www.broadinstitute.org/gatk/). Using only reads that overlap the variant candidates reduces the realignment time dramatically. Afterwards, the variants are called based on the realigned reads and finally processed using Beagle[115,116] to impute missing calls and to phase haplotypes. Our long-term vision is to eliminate third party tools in the pipeline and replace them by native functionality to allow consistent and high- performance data processing in an end-to-end pipeline.

Our traits of interest will include drought, flooding, salt, and low nutrient resistance and related traits phenotyped in the large-scale phenotypic screens under both field and greenhouse/growth room conditions described above (Table 1). These analyses will be relatively straightforward, though care will be taken to minimize false positives. The primary concern in this regard relates to population structure, which can produce marker/trait correlations in the absence of physical linkage, thereby resulting in spurious associations[117,118]. Statistical approaches have, however, been developed to minimize this problem by adjusting for both population structure and

relatedness amongst individuals (i.e., kinship) based on genotypic data from a collection of 'background' markers[119-121]. As currently envisioned, our analyses will make use of both individual SNPs and CNVs, as well as multi-locus haplotypes (the latter have the potential to increase power and reveal novel associations[122-124]) and will be performed using the R packages GWASTools[125] for the association model, and SNPRelate[126] to calculate population stratification and relatedness among accessions. We will use both linear regression and likelihood ratio association tests with the additive gene action model to detect regions of interest in the genome. In the regression model, population stratification (eigenvectors) and relatedness (identity by state) will be included separately as covariates to reduce the inflation in *p*-values which can lead to spurious associations. To choose the best model, we will calculate the inflation factor (lambda) for each analysis and compare results to ensure that we use the best model in each analysis. To reduce false positive calls due to multiple comparisons in the association tests, all *p*-values will be adjusted and corrected accordingly. Our current implementation of association analysis performs >50 times faster compared with other tools tested. This will facilitate an iterative analytical process to make certain the best results are reached. We will, however, re-evaluate the available options on an ongoing basis to ensure that we are using the most appropriate and powerful analytical approaches. In addition to the use of existing tools, we will develop genome-wide association mapping algorithms to incorporate genotype x environment interactions, which are a key component of the present proposal. This will give us the option of including phenotypic data from multiple sites and from both stress and control treatments into a single model. These algorithms will be implemented in the SAP HANA database to gain both analytical flexibility and speed.

Beyond the identification of genes/genomic regions underlying variation in abiotic stress resistance, we will investigate the occurrence and nature of genetic correlations between traits.

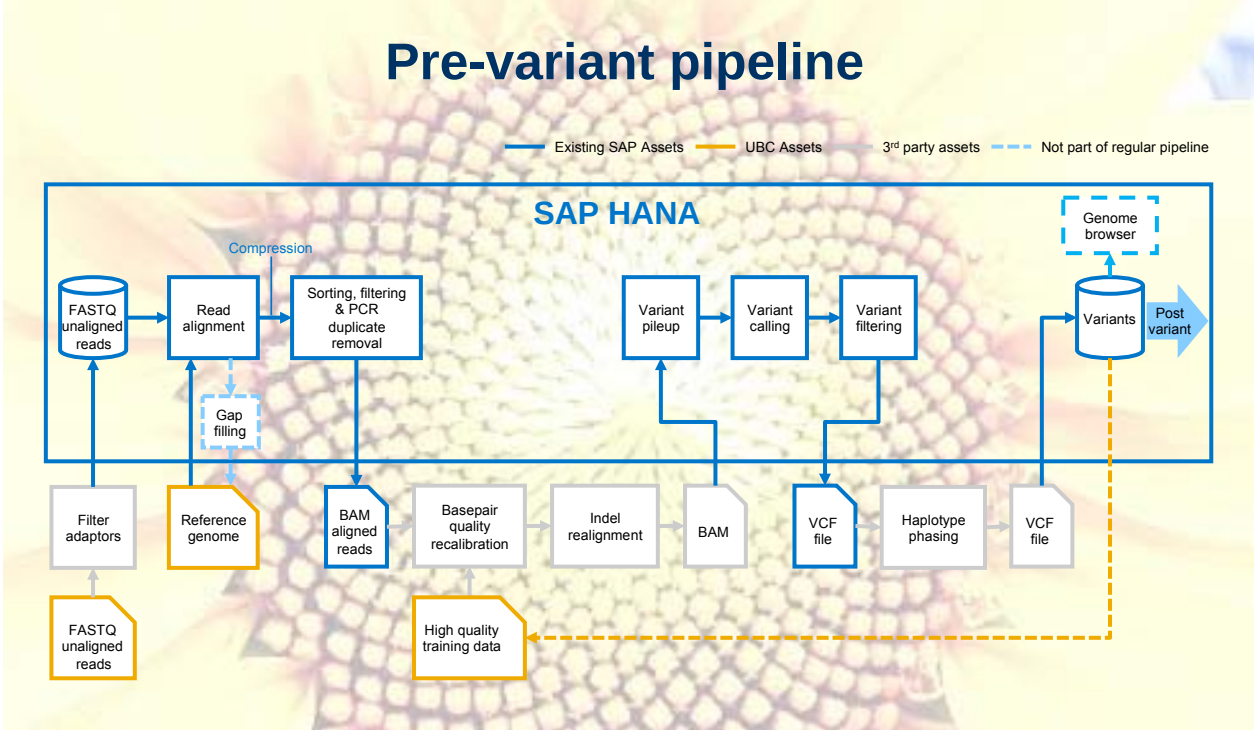

Figure 3. Alignment and variant calling pipeline we have developed in collaboration with SAP.

From a breeding perspective, the goal will be to identify abiotic stress resistance alleles that lack severe antagonistic correlations. Such correlations could result in unacceptable trade-offs with resistance to other stresses or performance under ideal conditions.

Activity 1.3 – Detailed ecophysiological characterization

In order to better understand the mechanistic basis of variation in stress resistance, subsets of cultivated lines that have been ranked for resistance to each stress (based on results from Activity 1.1) will undergo a further series of experiments including detailed ecophysiological characterization, characterization with an automated phenotyping system, and transcriptomic analyses.

For ecophysiological characterization, we will carry out a series of greenhouse and growth chamber studies at UGA (drought, salt, and low nutrient stress) and at UBC (flooding stress) using multiple phenological stages and multiple levels of stress. For each stress, the most and least resistant lines (15 of each) will be replicated within treatments using experimental designs appropriate for the stress applications (i.e., randomized complete block or split plot design). For drought stress resistance, we have the capacity to control soil moisture with several novel methods, including controlled irrigation to hold soil moisture at predetermined levels with an automated irrigation system that utilizes a soil moisture sensor in each of 200 pots[127], and tall pots (120 cm tall x 10 cm dia.) made of clear plastic (permitting root observations) outfitted with multiple irrigators that are removed from the top downwards to produce a declining soil moisture front in a drought treatment. For each stress (drought, flooding salt, low nutrients), the divergent stress resistance lines will be characterized for a core set of traits including relative growth rate, gas exchange (photosynthesis, stomatal conductance, instantaneous water-use efficiency), leaf lifetime, and specific root length[41,72,128-130]. Additional traits that are specific to each stress will be measured. Traits specific to drought stress include stem hydraulic efficiency and safety, leaf venation, root growth rates and rooting depth rate under well-watered conditions, the ability to respond to soil drying with greater allocation to roots, osmotic adjustment, turgor loss point, and maintenance of leaf N and green leaf area (i.e. stay-green)[14,23,39,131-135]. Additional flood stress traits include adventitious rooting, and malate accumulation[46-48]. Additional salt stress traits include leaf succulence, and the ability to exclude or sequester Na ions in root, stems, and leaves[111,136]. Additional low nutrient stress traits include relative growth rate, nutrient uptake rates, and the ability to respond to low nutrient conditions with greater allocation to roots[104,137]. Univariate and multivariate approaches will be used to assess the relationships among traits and the relationship of traits to stress resistance, in order to provide insight to the mechanistic basis of stress resistance at a scale intermediate between genes and phenotypes assessed in large-scale screens.

Activity 1.4 – Phenotyping with the Heliaphen platform

A subset of the SAM lines phenotyped using the tractor-mounted HTP in DREC field studies (years 1 and 2) will be characterized using Heliaphen's automated phenotyping platform (https://www.youtube.com/watch?v=VZSvgeWuhlwan), during years 3 and 4 of the project. We will target 15 lines of each type (drought resistant and susceptible based on results from Activity 1.1) with 2 treatments (stressed and control) in 3 replicates (180 individuals/year). This aspect of our work provides characterization of physiological differences of the divergent lines and provides an opportunity for collaboration between two different research groups at the forefront of developing sensor-based HTP approaches. Plants will be grown outdoors by our collaborators at INRA [138] in Toulouse, France during the sunflower growing season in 15L pots equipped with individual pot shelters. A 7,000 sq ft outdoor platform provides homogeneous environmental conditions for the plants to be tested. Throughout an experiment, environmental conditions are automatically recorded (temperature, wind, precipitation, evaporative demand) and a robot waters plants individually according to a pre-defined drought scenario. This scenario will be chosen to reproduce the stress observed in the DREC field experiments. Pot weight is measured automatically and plant water status (i.e., the fraction of transpirable soil water) is calculated on a daily basis. The Heliaphen platform also phenotypes the plants. Plant growth is assessed daily using ultrasonic sensors (plant height), laser scanner (stem diameter), and visible and near-infrared imaging associated to image segmentation (leaf area). These measurements will be compared to the plant height measurements and vegetative indices measured with the field HTP system at DREC. Together with pot weights, leaf area estimations allow approximation of the transpiration rates of the different lines in various water stress conditions. These estimations will

be confirmed by regular “manual” measurements of leaf area to avoid problems of organ orientation. Finally, the transpiration rates of the different lines in each treatment will be determined and compared to the data acquired through canopy temperature using the field HTP system.

Activity 1.5 – Transcriptomic analyses

In addition to the extensive ecophysiological characterization described above, we will compare the transcriptional responses of the lines that are most and least resistant to each stress. Our main goals are to: (1) identify genes that are differentially expressed in response to the four stresses; (2) determine overall patterns in the number, direction, and timing of gene expression changes; (3) develop gene co-expression networks for each stress; and (4) assess the extent of overlap in co-expression networks between stresses.

The evaluation of transcriptional responses will be conducted in growth rooms at UGA (for drought, salt, and low nutrient stress) and UBC (flooding stress) using four lines from the SAM population (two resistant and two susceptible), three biological replicates per line, four stresses (drought, salt, nutrient, flood), two treatments (control and stressed), three time points, and two tissues (576 samples total). The scale of the experiment necessitates that transcriptional responses be measured at the juvenile stage only. Also, note that the same four lines are unlikely to be chosen for analysis across the various stress treatments, and that the relevant stresses will be applied as described previously. Stress treatments will be applied starting at the four-leaf stage and will continue for 10 days. Whole seedlings will be harvested 24 hours after stress application to monitor expression changes due to acute stress, after 10 days to identify chronically expressed transcripts, and five days after recovery to assess whether stress-induced expression changes are fixed or plastic. Note that a pre-stress time point is not required because we will assay stressed and non-stressed individuals at each time point. All samples will be harvested at the same time of day to minimize the confounding effects of circadian variation.

We will generate strand-specific RNAseq libraries[139] from leaf and root tissue from each of the 576 seedlings and sequence them on an Illumina HiSeq (75 bp reads, v4 chemistry, 70-80x depth). There are multiple, rapidly maturing packages for analyzing RNAseq data. We currently use the Bioconductor package DESeq[140] for identifying expression level differences[141]. Gene co-expression network analysis will be employed to identify co-expression modules that are associated with specific stress responses, as well as modules that respond to multiple stresses[142,143]. This will provide insight into gene regulatory networks[144] and potentially reveal other genes that influence responses to stress[145]. The availability of these co-expression networks will also provide a framework for the interpretation of results from other aspects of the proposal. This will include insights into the putative functional modules within which our candidate stress resistance genes reside and their network positions and degree of connectedness to other members of a given module.

Activity 2: Population genomic analyses of stress adaptation in the wild

To address major evolutionary questions, and to further dissect genomic regions underlying abiotic stress resistance, we will search for associations between genotypic variation (SNPs and haplotypes) and important ecological variables (climate and soil characteristics) in natural populations of wild species contributing to our multi-species mapping population (see below). The challenge with this approach is distinguishing locus-environment correlations caused by selection from correlations arising from neutral processes (e.g., unrecognized population structure) and/or small sample sizes[58]. Simulations indicate that neutral correlations between genotypic and ecological variation arising from population structure can be minimized by sampling from geographically proximate, but ecologically divergent populations[146]. False positives can be further reduced by separate analyses of environmental correlations in multiple species, since the same locus is unlikely to be repeatedly associated with a given environmental variable due to chance alone.

Activity 2.1 – Collections & soil characterization

We will obtain soil samples and seeds from 50 populations across the full range of each of three extremophile species: *H. annuus* (tolerance to drought, flooding, and salt stress), *H. argophyllus* (drought, flooding, salt, and low-nutrient stress), and *H. petiolaris* (drought and low-nutrient stress). These three species are diploid annuals, with an outcrossing mating system. *Helianthus annuus* and *H. petiolaris* have broadly overlapping distributions across much of central and western North America. In contrast, *H. argophyllus* has a more linear geographic distribution, occurring for several hundred km along the Texas coastal plain at both coastline and inland sites. As far as possible, we will use the paired population sampling strategy described above to reduce neutral correlations caused by population structure. This will be straightforward for salt, low nutrient, or flooding stress because populations of the same species can be geographically proximal, but occur in soils that differ strongly in salinity, organic matter, and nitrogen availability (or occur at sites that are prone to spring flooding or not). Paired population sampling is more challenging for drought, but for the two widespread species (*H. annuus* and *H. petiolaris*) we will exploit both longitudinal and latitudinal precipitation gradients across the central and western USA to disrupt neutral correlations. Our sampling approach will build on the extensive seed collections and soil analyses previously made by the PIs, with the goal of filling gaps in our earlier collections. Note that samples from all populations, along with relevant collection information, will be deposited into the USDA sunflower germplasm collection. As detailed in her letter of collaboration, Dr. Laura Marek has agreed to oversee the distribution and long-term maintenance of these materials.

Activity 2.2 – Genotyping of wild populations

For genotypic characterization, we will conduct WGS re-sequencing of normalized libraries[147] from 1,500 genotypes (50 populations x 10 individuals per population x 3 species) to circa 6x depth. These sample sizes will allow us to detect essentially all strongly selected loci (s=0.1) and 20-90% of weakly selected loci (s=0.005), depending on demographic history[146]. Sequencing from normalized libraries represents a cost-effective strategy for comprehensively sampling the gene space in sunflower, as well as providing sufficient SNPs for haplotype reconstruction. In sunflower, for example, the genome is reduced from 3.6 Gb to circa 1 Gb of gene space. Such an approach is cost-competitive with a large SNP genotyping array, while avoiding ascertainment biases that are inherent to genotyping array methods, which can lead to erroneous conclusions concerning demographic history and natural selection (e.g. Lachance and Tishkoff[148]). The pipeline we previously developed with SAP will enable us to process the data from raw reads to SNP calls efficiently and accurately and will permit filtering of raw SNP calls to account for the minimum minor allele frequency, heterozygosity, and minimum depth of coverage to reduce the error rate in the data.

Activity 2.3 – Analyses of population genomic data

To identify locally adapted alleles, we will search for correlations between SNPs and/or haplotypes and spatially explicit environmental variables, including soil fertility data from each site and climatic variables from the WorldClim database[149]. Analyses will be performed for each species independently using a Bayesian statistical method implemented in the program Bayenv[58,59], which controls for the neutral correlations due to population structure. This program uses a null model for how allele frequencies co-vary among populations at putatively neutral loci and then identifies unusual correlations between allele frequency and ecological variables given this null model. In addition, we will search for loci with unusually large differences in allele frequency between populations using outlier detection methods[150-152], as such outlier loci may contribute to adaptive differences[153]. Population structure can also result in false positives in outlier tests[151,154]. To reduce the false positive rate, we will compare our simulated F_{ST} distribution to the empirical F_{ST} distribution to determine if the simulated mean and variance of F_{ST} is similar to the 50% quartiles of the empirical F_{ST} distribution[154]. As our main interest is in loci with major effects, we will focus on consecutive outliers (using sliding window or HMM).

However, the most compelling evidence of selection will be the discovery of the same unusual environmental correlations and/or outlier loci in multiple species.

The results from these analyses will allow us to ask fundamental biological questions, including the extent of overlap in genes under selection for various stresses, the extent of gene re-use during adaptation in the different species, the types of genes under selection (regulatory vs. structural), and the positions of these genes within regulatory networks. We also will be able to assess the role of hybridization in adaptation at the gene level by asking if the same haplotype is under selection in multiple species; this would suggest a role for hybridization/introgression as opposed to parallel responses from standing variation or new mutations. Extensive range overlap and frequent hybridization between *H. annuus* and *H. petiolaris* make them ideal for addressing this question[81].

The population genomic data and association/QTL mapping data from the SAM and MAGIC populations (below) are highly complementary. Because of low LD in the natural populations, we should be able to identify the specific gene (and site in some instances) that is targeted by selection. On the other hand, the phenotypic effects of this variation may be unclear. With association mapping/QTL data, we will have confidence in the phenotypic effects of a given locus, but it may contain multiple genes and SNPs. A combination of the two data sets will allow us to more confidently and efficiently link abiotic stress resistance to causal variants. An example of the potential power of this approach comes from studies of alcohol dehydrogenase (ADH). Allozyme studies showed that an allele of ADH was associated with flooded habitats[48], and this locus has repeatedly appeared as a top outlier in our transcriptome scans[88,155].

Activity 3: Development and characterization of multi-species mapping populations

To facilitate the efficient genetic analysis of complex trait variation in *Helianthus* and to produce materials containing exotic alleles that can be readily deployed in breeding programs, we will develop modified Multiparent Advanced Generation Inter-Cross (MAGIC) mapping populations[156] that include both cultivated and wild sunflower donors. Such populations combine the high power to detect QTLs offered by biparental populations with the ability to assay a broader spectrum of diversity and improved resolution afforded by association mapping.

Activity 3.1 - Population development

Because sunflower is a hybrid crop, we will develop separate male and female populations. For each population, our crossing design will involve four elite breeding lines (E) and four wild (W) donors. Each wild donor will be successively combined with three elite lines to produce four different intermated populations (IM1-4; see Figure 4, below) that will ultimately be intercrossed in all pairwise combinations (with a target of seed from 50 crosses from each of the six pairs, IM1 x IM2, IM1 x IM3, […] IM3 x IM4; 300 crosses total). We will then develop 600 RILs with equal representation from each of the six intercross pairs via multiple rounds of self-pollination.

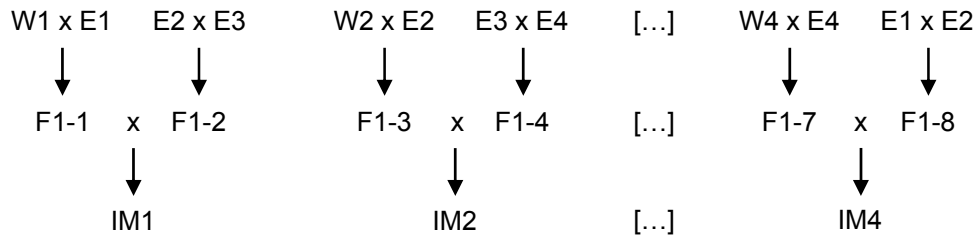

Figure 4. Overview of initial crossing design for production of multi-species mapping populations.

The elite donors were chosen in consultation with our public and private partners and were selected to represent a broad cross-section of cultivated sunflower diversity, including key

resistance traits and both oil and confectionary varieties (Table 2). The wild donors, which have likewise been identified in consultation with our partners, include drought, salt, flooding, and dune (low nutrient) adapted populations. Initial crosses were made at UBC during the summer and fall of 2014, prior to the initiation of funding and IM1xF1 lines are currently being produced.

Table 2. MAGIC population donors.

Female population		
Elite Donors	**Source**	**Traits**
LR1	INRA	Broomrape tolerance
CM447	Agrifood Canada	Early maturing, high oil, rust resistance, large seeds, high yield
HA336	INRA	Mildew resistance
HA467	USDA	Sclerotinia and Phomopsis resistant, high oleic, herbicide resistance
Wild Donors		
H. annuus – salt flats	Wild collected	Salt and drought tolerance
H. annuus – stabilized dunes, western Nebraska	USDA	Drought and low nutrient tolerance
H. annuus, southern Manitoba	USDA	Early flowering, flooding tolerance
H. argophyllus, Texas Barrier Islands	Wild collected	Salt and low nutrient tolerance
Male Population		
Elite Donors	**Source**	**Traits**
RHA 426	USDA	Herbicide resistance, Sclerotinia stalk rot resistant
RHA 455	USDA	High oleic, Sclerotinia resistant
BH13-714	USDA	Early Flowering, parent of promising Canadian hybrid
HE13EE024294.002	Mycogen/Dow	Unknown
Wild Donors		
H. exilis – Serpentine soils, N. California	Wild collected	Heavy metal tolerance
H. petiolaris – Dune habitat, Colorado	Wild collected	Low nutrient tolerance
H. paradoxus – Saline wetland, New Mexico	Wild collected	Salt and flooding tolerance
H. maximiliani – Great Plains, Oklahoma	The Land Institute	Drought tolerance, perennial life history

Subsequent generations will be produced by our partners. The use of winter nurseries will speed population development, and we anticipate RIL availability by year 3 of the project. If development goes more slowly than planned because of challenges associated with flowering time variation and hybrid incompatibilities, we will evaluate the RILs after fewer rounds of selfing (evaluation of S4 RILs would probably be the worst case scenario). Note that because the F1 lines are crossed into cultivated material, we expect most incompatibilities to be efficiently eliminated by selection, which will reduce issues with recombination suppression in the RILs. On the other hand, we are interested in maintaining flowering time variation in the population, as early flowering is especially valuable at higher latitudes such as Canada.

Activity 3.2 – Genotyping & Phenotyping of MAGIC populations

As currently envisioned, normalized libraries of the 16 parental genotypes will be sequenced to circa 15x depth using the Illumina Hi-Seq platform. The 600 RILs from each population will be

characterized using a genotyping by sequencing (GBS) approach[157] we have optimized for sunflower. GBS is inexpensive, and based on our experience with wild and cultivated sunflower, should deliver >50,000 SNPs, but levels of missing data can be high. Missing markers can, however, be efficiently and accurately imputed in multi-parent mapping populations[158]. We will re-visit our genotyping strategy when the populations are ready for analysis and will use the most suitable and cost-effective methods available at that time. Due to the size and the timing of the availability of the MAGIC populations, evaluation of both populations under both field and greenhouse conditions is beyond the scope of the present proposal, though such work will be a priority going forward. Instead, we will grow and evaluate the female population under field conditions using the same general design and phenotyping methods (traditional as well as HTP) outlined above for the SAM population: two treatments (control, stress), 600 lines, 2 replicate plots per line, and an incomplete block design. The incomplete block design will be nested in two tiers to eliminate spurious variation arising from the large land area within each replicated complete block: each full-sib population will be a 10x10 simple lattice design nested within a sets-in-reps design among full-sib populations. Such an incomplete block design allows us to eliminate spurious "noise" variation due to unknown field trends, and allows us to measure, compare, and map within full-sib families with optimal precision. Note that we have reduced replication relative to the SAM population (from 4 to 2 replicates) because of the larger size of the MAGIC population, and we will conduct power analysis on the SAM population data to verify that this is sensible *a posteriori* and make adjustments as needed. The drought stress trials will be conducted in year 3 at DREC, whereas the evaluation of flooding, salt, and low nutrient stress will be carried out at Fargo, ND, Indian Head, SK, and NaSARRI, Uganda, respectively.

Map construction and QTL detection will employ R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines[159]. This work will thus result in the production and baseline characterization (including the mapping of QTLs underlying a variety of agriculturally important traits) of a highly valuable germplasm resource for the sunflower research and breeding community. This next generation germplasm resource and the accompanying genotypic information can be directly employed by our end users to efficiently move desired traits and alleles into their breeding programs through marker-assisted and/or genomic selection.

Activity 3.3 - Detailed ecophysiological characterization

Though wholesale, population-level stress testing will necessarily be left to the future, we will use data from the field experiments to rank the lines for each focal stress. For each stress, the 25 most resistant MAGIC lines and the 25 least resistant MAGIC lines will be subjected to in-depth phenotypic characterization under control and stressed conditions in year 4. Stress resistance will be tested using the treatments (control, drought, flooding, and low nutrients), experimental design, and trait measures as described for the greenhouse / growth chamber evaluation of the SAM lines (Table 1). A subset of 30 putatively drought resistant/susceptible MAGIC lines (15 of each) will also be characterized by our collaborators at INRA in year 4. This work will follow the same general methods described above for the Heliaphen-based evaluation of SAM lines.

Note that all lines from this population will be made freely available via deposition in the USDA sunflower germplasm collection as soon as sufficient seed is available. As detailed in her letter of collaboration, Dr. Laura Marek has agreed to oversee the distribution and long-term maintenance of this population.

Activity 4: Functional validation of candidate stress resistance genes

Building on information from Activities 1-3, we will target a subset of genes that appear to have large impacts on drought, salt, or low nutrient resistance for functional validation. In the short run, this information will confirm the agronomic value of the alleles that are likely to be the focus of marker-assisted breeding efforts and ensure that appropriate markers are deployed for marker-assisted selection. In the longer run, this information will allow us to potentially increase

trait efficacy using genome editing approaches and potentially extend our discoveries to other oil crops (see below).

Activity 4.1 – Functional analyses of candidate genes in sunflower

We will prioritize circa 15 genes that: (1) exhibit strong associations with resistance to one or more abiotic stresses; (2) show unusually strong correlations with the relevant environmental parameter(s) in natural populations of two or more wild species; (3) occur in regions of the genome with low LD, where trait associations can be confidently mapped to individual genes; (4) display notable differences in sequence or gene expression; and/or (5) exhibit the signature of divergent natural selection. Note that for the first several quarters, we will analyze candidate genes identified in previous projects, such as an allele of alcohol dehydrogenase that was previously shown to be associated with tolerance to flooding (see Activity 2.3).

The Rieseberg lab has developed an effective protocol for transformation and RNAi in sunflower. Stable transgenics with RNAi constructs designed to silence selected candidate genes will be generated using an actin-2 promoter, or the candidate gene's own promoter, since we and others have found that the standard 35S promoter is not well tolerated in sunflower. T_1 plants will be assayed for silencing of the YFP reporter gene and their phenotypes recorded. Phenotypic modifications correlated with the segregation of the T-DNA and the silencing phenotype will be confirmed in the T_2 generation. To minimize faulty inferences due to off-target gene silencing, we will test alleles for complementation in homozygous knock-out lines in *Arabidopsis* whenever possible, and employ genome editing approaches to locate causal mutations[70,160]. The latter will mainly be conducted in *Aradidopsis* because of its many experimental advantages. However, Rieseberg is currently collaborating with a seed company to optimize the CRISPR/Cas9 system for sunflower with the goal of enhancing the efficacy of a crop protection trait. Thus, it should be possible to employ such an approach in sunflower for the proposed project as well, if necessary.

Activity 4.2 – Genome editing in soybean

For mutations displaying expected effects in *Arabidopsis*, we will introduce them into other crops as time and technology permits. Our initial focus will be on soybean since it is the world's leading oilseed crop, has a similar climate suitability envelope in Canada, and transformation efficiencies are high. The CRISPR/Cas9 system is particularly promising for the proposed work since a variety of mutations/alleles can be made. Also, the mutations are permanent, so the transgene need not be retained after the cleavage event[161], and collaborator Parrott's lab brings experience with the CRISPR/Cas9 system in soybean[162]. Parrott's group has developed a number of CRISPR family vectors for soybean that can be easily modified for use in the present project (http://www.addgene.org/browse/article/7014/). A milestone will be the introduction of 3-4 of the most promising mutations into soybean homologs.

E. GE^{3}LS Research

The GE^{3}LS research includes two activities, both of which are intended to help ensure that the social and economic benefits of this project can be realized. The goal of the first GE^{3}LS activity is to develop crop models that will allow us to predict likely yield gains from new highly resistant sunflower cultivars on primary, secondary, and marginal lands in Canada and in the context of climate change. The goal of the second GE^{3}LS activity is to address significant negative impacts of international treaties on the use of plant genetic resources by private and public sector breeding programs in Canada and worldwide.

Activity 5: Development of crop yield models

We will conduct a broad-scale assessment of the potential growing areas and yields of the new sunflower cultivars, as well as competitor crops, across Canada. Large-scale assessments have previously taken a "crop suitability" approach [163-165]. While such analysis can usefully delineate potential growing areas for different crops as well as potential shifts from climate change, the

interpretation of results is challenging as suitability is a simple index ranging from zero to one and its relationship to crop productivity is unclear. Instead, here we will use a crop modelling approach to directly simulate crop yields.

The development of global crop models in recent years enables the simulation of crop yields using basic ecological principles, a few crop growth and phenological parameters, and large-scale climate and soil data sets. The advantage of this approach over using more detailed crop models is that they require fewer input data sets and are capable of doing spatial simulations over large scales. In this project, we will use the Agro-IBIS model, which was developed to assess global crop yield response to climate change and changes in crop management practices. Agro-IBIS was developed by representing agro-ecosystems within the IBIS global ecosystem model. IBIS has process-based representations, within a single consistent framework, of land-surface processes (exchanges of energy, water, and momentum between the vegetation and the atmosphere), plant physiology (photosynthesis and stomatal conductance), vegetation phenology, carbon balance (including soil biogeochemistry), and vegetation dynamics (see 166,167 for more details). Agro-IBIS introduced representations of crop management (planting/harvest dates, cultivar selection, fertilizer application, and tillage) and crop phenology (thermal time and crop growth stages)[168]. Adapting the Agro-IBIS model to simulate the yields of sunflower as well as its competitor crops will involve the following tasks.

Activity 5.1. Model development

We will develop the Agro-IBIS model to simulate sunflower and other crops important in the Canadian context. Agro-IBIS already includes representation of maize, spring wheat, and soybean. We will include model representations of sunflower, canola, and barley. In developing these model parameterizations, we will derive parameter values by consulting the literature as well as detailed crop growth models that are part of the DSSAT system http://ecobas.org/www-server/rem/mdb/dssat.html; [169] and the EPIC model http://ecobas.org/www-server/rem/mdb/epic.html; [170].

Activity 5.2. Input data gathering and development

To conduct spatially-explicit model simulations across Canada, we will need spatial input data on climate, soils, irrigation, and fertilizer use. Climate data in gridded format (i.e., spatially-explicit data on a raster grid) are available from the Climatic Research Unit of the University of East Anglia (http://www.cru.uea.ac.uk/high-resolution-gridded-datasets), while soils survey data are available from the Soil Landscapes of Canada data provided by the Canadian Soil Information Service (http://sis.agr.gc.ca/cansis/). For irrigation and fertilizer use, we will perform several different model simulations based on alternate scenarios of application rates, as explained in the next section.

Activity 5.3. Model simulations and analysis. We will simulate the yield responses of all crops to different irrigation and fertilizer application rates, for both current cultivars as well as new cultivars developed by this project. For irrigation, as in Deryng et al.[171], we will perform model simulations of rain fed conditions as well as varying degrees of irrigation (i.e., irrigation water application rates ranging from zero to that required to maintain non-water-stressed conditions). For fertilizer application, we will perform an initial model simulation using current crop-specific fertilizer application rates from Mueller et al.[172]. We will then perform simulations with different scenarios of both increased and decreased fertilizer application rates. The different yield responses of the old and new cultivars to varying management conditions will be explored. We will further simulate yield changes as a function of various climate change scenarios[173]. Comparison of these simulations will reveal the competitive advantage of sunflower relative to other crops as well as shifts in these with climate change. The crop yield models will be further linked to existing economic models, such as those developed by collaborator May[174], which will allow predictions of locations where particular cultivars are likely to be successful and to compare predicted productivity and net returns of different crops (and cultivars) at a given site.

Activity 6: International treaties and use of plant genetic resources

This activity focuses on the Convention on Biological Diversity (CBD) and International Treaty for Plant Genetic Resources in Food and Agriculture (Treaty) in two inter-related work streams: (1) to clarify potential impacts and opportunities for the outputs of this project; and (2) to engage in policy discussions with the Treaty Secretariat and other relevant organizations to facilitate uptake of innovations generated by agricultural genomics projects such as this one. To do so, co-PI Marden will work closely with the project team as well as engaging with other stakeholders from academia, governmental bodies, industry and the Treaty Secretariat.

At present, the CBD and Treaty present potential hurdles to uptake of innovations in agricultural genomics. The CBD, developed to address access and benefit-sharing of indigenous biological resources in developing nations, allows national regimes to control access to these resources on a number of grounds, sometimes resulting in barriers to research and development. The Treaty was developed, in part, to alleviate these barriers by facilitating access to covered plant genetic resources under uniform mutually-agreed upon terms. The signatories to the Treaty, a group that includes Canada, Europe, and most developed and developing countries, commit *inter alia* to make accessions of plant genetic resources for 64 key crops (including sunflower) available to all users according to the terms of a Standard Material Transfer Agreement (SMTA). The SMTA, negotiated as part of the Treaty, includes potentially perennial benefit sharing obligations arising "from the use, including commercial, of plant genetic resources for food and agriculture [covered under the Treaty]." The United States is not a signatory to the Treaty and thus materials in its National Plant Germplasm System (NPGS) are not subject to the SMTA; nonetheless, the U.S. has committed to honouring obligations under the SMTA (P. Bretting, USDA-ARS National Program Leader, pers. comm.).

The Treaty has opened access to many resources. However, there are multiple layers of ambiguity arising from the Treaty and its application that may hamper innovation. First, the scope and impacts of SMTA obligations remain unclear: there is no guidance in the Treaty or SMTA, for example, on what use of Treaty resources triggers benefit sharing obligations, or how such obligations are modified by combination of Treaty resources with other non-Treaty germplasm. Further, it remains unclear whether the SMTA obligations apply to derivative or improved germplasm or to germplasm modified via new genome editing technologies. Even more, there is ongoing ambiguity as to whether genomic data generated from Treaty resources (i.e. plant genomes) or used to modify Treaty resources are also subject to the SMTA. The U.S. position adds a further layer of complexity, as it becomes necessary to track the origins and combinations over time.

The role of the Treaty vis-à-vis genomics data is also unclear. As required by Article 17 of the Treaty, the Treaty Secretariat is working toward the creation of a Global Information System (GIS) to facilitate sharing of information not only about Treaty resources but also plant genetic resources and genomic data more broadly. Work on this endeavour has begun, and while the potential benefits of a funded mechanism are clear, there are unresolved questions of what proprietary interests can or should be attached to resources entered into the GIS, and how such interests will be tracked and enforced. Such questions have potential far reaching implications for a broad range of stakeholders.

The net result of these uncertainties is that some companies avoid use of any genetic resources subject to the SMTA (H. Dempewolf, Global Crop Diversity Trust, pers. comm.), and may not access innovative germplasm such as that proposed by this project. The Rieseberg lab recently witnessed this when a multinational corporation expressed interest but ultimately declined to use pre-bred germplasm generated by the Rieseberg lab, because the germplasm was subject to the Treaty SMTA. Conversations with parties in industry suggest that there is a need for greater clarity and transparency in how germplasm and data are shared.

Activity 6.1. Clarify potential impacts and opportunities

Co-PI Marden will work closely with the project team to: (1) identify the source of externally resourced germplasm and data; (2) identify project use and potential modifications, derivation, or improvements – in legal terms - to any such germplasm or data; (3) define intended outputs of the project and relationship to any identified uses, modifications, derivation or improvements of germplasm or data; and (4) coordinate potential sharing or out-licensing of such outputs. This work will commence with meetings and interviews with team members to identify the elements above. In addition, Marden will identify and review terms and conditions related to any germplasm or data utilized by the project team.

In conjunction with this work, Co-PI Marden will conduct a legal analysis to assess explicit obligations and ambiguities in the CBD and Treaty, as well as in pertinent UBC, Genome Canada and other institutional frameworks. Ultimately, based on the factual information gathered from the project team and the analysis of the relevant legal frameworks, the GE^{3}LS team will map out actual and potential impacts on project outputs in terms of resources used and potential outputs. Marden will also identify areas where the legal frameworks are ambiguous and areas where there are opportunities to interpret the frameworks in a beneficial manner. These analyses will be shared with the project team in an iterative manner in order to refine potential issues and to ensure that any changes are addressed.

Activity 6.2. Facilitate Policy Impacts

In addition to directly advising the project team, Co-PI Marden intends to bring the identified issues into broader policy discussions with the aim of facilitating further innovation in agricultural genomics. This effort will take place through a number of avenues. In each case, Marden will discuss developments with the project team to ensure proposals are consistent with the outputs of the project.

First, Marden will act as a contributing expert to the two Treaty Secretariat Working Groups and have the potential to modify interpretations of the Treaty (textual changes to the Treaty or the SMTA itself are unlikely in the near term). These Working Groups include: (1) the Article 17 Global Information System (GIS) on Plant Genetic Resources for Food and Agriculture and, (2) to the degree relevant, the Ad Hoc Open-ended Working Group to Enhance the Functioning of the Multilateral System of Access and Benefit-sharing. In particular, the development of the GIS offers the opportunity to discuss how – if at all – the Treaty will be interpreted to apply to data arising or derived from SMTA material, as well as how data from non-Treaty sources might interact with SMTA material. Participation in the Working Groups has the potential to result in actual modifications in how the Treaty is interpreted and in new mechanisms through which to share genomics data.

Separately, in her role as the governance expert on the Steering Committee of the DivSeek Initiative (http://www.divseek.org), Marden will engage with plant genomics researchers, breeders, funders, genebank managers and other stakeholders to devise a governance mechanism for sharing plant genomic resources that respects certain proprietary interests but facilitates innovation. This position gives Marden a role at the center of determining how agricultural genomics data will be shared going forward.

Finally, in an effort to generate independent consensus policy options arising from issues identified in this project, Marden will convene a working group of stakeholders from industry, CGIAR centers, seed banks, academia and the Treaty Secretariat in a neutral forum to propose how ambiguities in the CBD and Treaty could be interpreted to address issues of innovation in agricultural genomics. Through her past and current research, Marden has contacts with the Treaty Secretariat, the CG Bioversity Centre, the USDA Agricultural Research Service, industry, genebank managers and other relevant stakeholders.

F. Scientific Outcomes and Deliverables

Our project will identify strategies, mechanisms, traits, and alleles that can be used to produce sunflower lines that are more resistant to drought, flooding, salt, and low nutrient stress, but with

minimal impact on productivity under ideal conditions. We will also determine the types of genes (and their network positions) that are most likely to confer resistance across broad genetic backgrounds and with minimal yield trade-offs. This knowledge will move us closer to a systems-level understanding of the interactions between plants and their environment and will enable the development of smarter and more efficient strategies for breeding environmentally resilient cultivars in sunflower and has the potential to positively impact other crops.
Knowledge of the mechanisms and genetic architecture of abiotic stress resistance, coupled with identification and functional validation of key resistance alleles (and associated genotypic information), will lead to our main deliverables:

(1) "Next Generation" germplasm resources in the form of multi-species, advanced generation intercross populations for use by sunflower breeders. This germplasm will be available in year 3 of the project and will enable end-users to efficiently deploy resistance alleles with minimal yield trade-offs in sunflower breeding programs using marker-assisted and/or genomic selection. New highly resistant and productive cultivars are expected in the field within four years of project end.

(2) A central data mining and analysis resource for sunflower that will enable researchers and breeders to more rapidly process, analyze, and manage large genomic and phenotypic data sets. The mining and analysis tools will be released throughout the course of the project.

(3) Crop yield models that will enable predictions of likely yields of new stress resistant sunflower cultivars (and competitor crops) in different soil and climate conditions across Canada. These models, in combination with existing economic models, will allow seed companies and farmers to determine where particular cultivars are likely to be successful and to compare predicted productivity and net returns of different crops (and cultivars) at a given site. Crop yield models will be delivered in year 4.

(4) Strategies for mitigating barriers to public and private breeding programs (e.g., uncertain IP and profit sharing obligations) resulting from international treaties; and Co-PI Marden will continue to work closely with the CGB/Treaty to move forward refinements to existing treaties to reduce such barriers. Strategies and suggestions for refinements will be made throughout the course of the project.

We expect the new sunflower cultivars to be in the field within four years of project end, with significant economic benefits through mitigation of risks in contemporary sunflower production in North America and elsewhere, expansion of sunflower production onto marginal lands, and important social benefits through increased food security in Uganda and other developing countries. This project team encompasses most of the major participants in scientific discovery in sunflower as it pertains to crop improvement and evolutionary biology. The team will use the results of this project to integrate into other related projects to ensure rapid adoption of the technology.

G. Research Expertise

Investigators: Rieseberg and Burke are international leaders in plant evolutionary genomics, with expertise in genetic mapping, QTL analyses, GWAS, genome assembly, population genomics, bioinformatics, and various reverse genetic approaches. Donovan is a leading expert in plant ecophysiology, especially adaptation to drought, salt, and low nutrient stress. Langlade has expertise in high throughput phenotyping, gene network analyses, and molecular physiology, especially in relation to drought stress. Yeaman is a rising star in population genetic theory, population genomics, and bioinformatics, with latter skill set developed as part of the Genome Canada-funded AdapTree project. Hulke has expertise in marker-assisted and genomic selection, as well as flood stress phenotyping. The GE^{3}LS team PIs, Marden and Ramankutty, have complementary expertise relating to the Treaty/sharing of agricultural data and land use, respectively. As mentioned above, Marden's expertise in these areas was developed in part

during the Genome Canada-funded Genomics of Sunflower project. Ramankutty is an international leader in global land use change and its implications and brings expertise in crop yield modeling to the project.

Collaborators: Bill May is a Canadian agronomist with expertise in phenotyping, breeding, and economic model development. Bali provides expertise in irrigation and drought stress management, whereas Andrade is an agricultural engineer with expertise in high throughput phenotyping (HTP). Laura Marek runs the USDA's sunflower germplasm collection and brings expertise in the collection and management of wild sunflower germplasm. Parrott is a world leader in the implementation of genome editing approaches, especially in soybean, and Kubach manages a highly skilled team of software engineers at SAP who are developing bioinformatics tools for plant genomics. Our other collaborators, who are also end users, bring expertise in sunflower population development and both conventional and molecular breeding approaches to the project.

H. Research Support

Sequencing will be conducted by the Genome Sciences Centre (GSC) in Vancouver, which operates 13 Illumina HiSeq 2500 v4 and 3 MiSeq instruments**.** The GSC is one of Genome Canada's GIN nodes. Bioinformatics will be conducted at several locations. Our collaborators at SAP AG will provide access to their in-memory database (SAP HANA), as well as to the SAP Cloud, which will be useful for the processing and analyses of the very large genomic datasets to be generated by this project. Yeaman and Rieseberg's groups also have access to Westgrid, which offers high-performance computing to academic institutions in Western Canada, the UBC Biodiversity Centre's computing cluster, and to the large-scale and cost effective storage capacity offered by UBC's Botany department. This is in addition to significant computing power available in the investigator's own labs.

Phenotyping will be carried out at several locations, including UBC, Indian Head (SK), U. Georgia, USDA-Fargo, DREC, INRA, and NaSSARI. The required infrastructure for greenhouse/growth room phenotyping and ecophysiological characterization is available at UBC, U. Georgia, and INRA. High-throughput phenotyping capability exists at DREC (e.g. tractor-mounted sensor arrays) and INRA (the Heliaphen system). The critical infrastructure necessary to apply field-based HTP approaches is also present at USDA-Fargo and Indian Head, including GPS-RTK guided planters and small plot combines. Conventional phenotyping capability exists at NaSSARI, and the present project will introduce more automated phenotyping approaches to this location.

Functional analyses will be performed in the laboratories of Burke, Rieseberg, and collaborator Parrott. All three labs are well-equipped for basic molecular analyses and both forward and reverse genetics approaches. Parrot and Rieseberg also maintain transformation and tissue culture labs.

I. References

1. Diamond J. 2002. Evolution, consequences and future of plant and animal domestication. *Nature* **418**:700-707.
2. Gepts P, 2014. The contribution of genetic and genomic approaches to plant domestication studies. *Curr Opin Plant Biol* **18**:51-59.
3. Smedegaard-Petersen V *et al.* 1985. The limiting effect of disease resistance on yield. *Ann Rev Phytopathol* **23**:475-490.
4. Koziol L *et al.* 2012. Reduced drought tolerance during domestication and the evolution of weediness results from tolerance-growth trade-offs. *Evolution* **66**:3803-3814.
5. Mayrose M *et al.* 2011. Increased growth in sunflower correlates with reduced defences and altered gene expression in response to biotic and abiotic stress. *Mol Ecol* **20**:4683-4694.
6. Mayrose I *et al.* 2011. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. *Mol Biol Evol.* **28**:759-770.

7. Plucknett DL *et al.* 1983. Crop germplasm conservation and developing countries. *Science* **220**:163-169.
8. Tanksley SD *et al.* 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* **277**:1063-1066.
9. McCouch SR *et al.* 2013. Agriculture: feeding the future. *Nature* **499**:23-24.
10. Anonymous. 2014. FAO <http://faostat.fao.org/%3E>.
11. Rauf S. 2008. Breeding sunflower for drought tolerance. *Comm Biom Crop Sci* **3**:29-44.
12. Cobb JN *et al.* 2013. Next-generation phenotyping. *Theor Appl Genet* **126**:867-887.
13. Sinclair TR. 2011. Challenges in breeding for yield increase for drought. *Trends Plant Sci* **16**:289-293.
14. Richards RA *et al.* 2010. Breeding for improved water productivity in temperate cereals *Funct Plant Biol* **37**:85-97.
15. Salekdeh GH *et al.* 2009. Conceptual framework for drought phenotyping during molecular breeding. *Trends Plant Sci* **14**:488-496.
16. Wright IJ *et al.* 2004. The worldwide leaf economics spectrum. *Nature* **428**:821-827.
17. Chapin FS *et al.* 1993. Evolution of suites of traits in response to environmental stress. *Am Nat* **142**:S78-S92.
18. Reich PB *et al.* 2003. The evolution of plant functional variation: Traits, spectra, and strategies. *Int J Plant Sci* **164**:S143-S164.
19. Grime JP *et al.* 2012. *The Evolutionary Strategies That Shape Ecosystems*. (Wiley).
20. Vasseur F *et al.* 2012. A common genetic basis to the origin of the leaf economics spectrum and metabolic scaling allometry. *Ecol Letters* **15**:1149-1157.
21. Donovan LA *et al.* 2011. The evolution of the worldwide leaf economics spectrum. *TREE* **26**: 88-95.
22. Munns R. 2011. Plant adaptations to salt and water stress: differences and commonalities. *Adv Bot Res* **57**:1-32.
23. Sadras VO *et al.* 2014. Improvement of crop yield in dry environments: benchmarks, levels of organization, and the role of nitrogen. *J Exp Bot*, doi:10.1093/jxb/eru1061.
24. Poorter H *et al.* 2005. A genetic analysis of relative growth rate and underlying components in *Hordeum spontaneum*. *Oecologia* **142**:360-377.
25. Ludlow MM. 1989. in *Structural and Functional Responses to Environmental Stress* (eds KH Kreeb, H Richter, & T M Hinckley), 269-281 (SPB Academic Publishing).
26. McKay JK *et al.* 2003. Genetics of drought adaptation in *Arabidopsis thaliana*: I. pleiotropy contributes to genetic correlations among ecological traits. *Mol Ecol* **12**:1137-1151.
27. Chaves MM *et al.* 2003. Understanding plant responses to drought - from genes to the whole plant. *Funct Plant Biol* **30**:239-264.
28. Stone LR *et al.* 2002. Water depletion depth of grain sorghum and sunflower in the central High Plains. *Agron J* **94**:936-943.
29. Conner DJ *et al.* 1997. in *Sunflower Technology and Production* (ed AA Schneiter) 113-182 (Soil Science Society of America, Madison, WI).
30. Fereres E *et al.* 1986. Genetic variability in sunflower cultivars under drought. 1. Yield relationships. *Australian J Agr Res* **37**:573-582.
31. Maury P *et al.* 2000. Leaf water characteristics and drought acclimation in sunflower genotypes. *Plant Soil* **223**:153-160.
32. Baldini M *et al.* 1998. Agronomic and physiological assessment of genotypic variation for drought tolerance in a cross of *H. annuus* and *H. argophyllous*. *Agricol Meda* **128**:232-240.
33. Darvishzadeh R *et al.* 2010. Evaluation of the reaction of sunflower inbred lines and their F(1) hybrids to drought conditions. *Spanish J Agr Res* **8**:1037-1046.
34. Gimenez C *et al.* 1986. Genetic variability in sunflower cultivars under drought. 2. Growth and water relations. *Australian J Agr Reh* **37**:583-597.
35. Chimenti CA *et al.* 2002. Osmotic adjustment and yield maintenance under drought in sunflower. *Field Crops Res* **75**:235-246.
36. Fulda S *et al.* 2011. Physiology and proteomics of drought stress acclimation in sunflower (*Helianthus annuus* L.). *Plant Biol* **13**:632-642.

37. Serraj R *et al.* 2002. Osmolyte accumulation: Can it really help increase crop yield under drought conditions? *Plant, Cell Env* **24**:333-341.
38. Darvishzadeh R *et al.* 2011. Path analysis of relationships between yield and related traits in sunflower under well-watered and water-stressed conditions. *Aust J Crop Sci* **5**:674-680.
39. Verslues PE *et al.* 2006. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J* **45**:523-539.
40. Munns R 2002. Comparative physiology of salt and water stress. *Plant, Cell Env* **25**:239-250.
41. Brouillette LC *et al.* 2014. Adaptive differentiation of traits related to resource use in a desert annual along a resource gradient. *New Phytol* **201**:1316-1327.
42. Aerts R *et al.* 2000. The mineral nutrition of wild plants revisited: a re-evaluation of processes and patterns. *Adv Ecol Res* **30**:1-67.
43. Omoigui LO *et al.* 2007. Response of low-N pool maize population to nitrogen uptake and use efficiency after three cycles of full-sib recurrent selection. *J Agri Sci* **145**:481-490.
44. Bailey-Serres J *et al.* 2012. Waterproofing crops: Effective flooding survival strategies. *Plant Physiol* **160**:1698-1709.
45. Bailey-Serres J *et al.* 2008. Flooding Stress: Acclimations and genetic diversity. *Ann Rev Plant Biol* **59**, 313-339.
46. Colmer TD *et al.* 2009. Flooding tolerance: suites of plant traits in variable environments. *Funct Plant Biol* **36**:665-681.
47. Kennedy RA *et al.* 1992. Anaerobic metabolism in plants. *Plant Physiol* **100**:1-6.
48. Torres AM *et al.* 1981. Sunflower alcohol-dehydrogenase genotypes--germination rates and response to flooding. *Env Exp Botany* **21**:35-44.
49. Furbank RT *et al.* 2011. Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci* **16**: 635-644.
50. Donovan LA *et al.* 1994. Potential for selection on plant for water-use efficiency as estimated by carbon isotope discrimination. *Amer J Bot* **81**:927-935.
51. Farquhar GD *et al.* 1989. Carbon isotope discrimination and photosynthesis. *Ann Rev Plant Physiol Plant Mol Biol* **40**:503-537.
52. White JW *et al.* 2012. Field-based phenomics for plant genetics research. *Field Crops Res* **133**:101-112.
53. Berger B. *et al.* 2010. High-throughput shoot imaging to study drought responses. *J Exp Botany* **61**:3519-3528.
54. Araus JL *et al.* 2014. Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci* **19**:52-61.
55. Ul Haq T *et al.* 2013. Reliability of ion accumulation and growth components for selecting salt tolerant lines in large populations of rice. *Funct Plant Biol* **41**:379-390.
56. Asch F *et al.* 2000. Leaf K/Na ratio predicts salinity induced yield loss in irrigated rice. *Euphytica* **113**:109-118.
57. Andrade-Sanchez P. *et al.* 2012. Use of a moving platform for field deployment of plant sensors. *ASABE Annual International Meeting* Paper Number 121337985.
58. Coop G *et al.* 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**:1411-1423.
59. Gunther T. *et al.* 2013. Robust identification of local adaptation from allele frequencies. *Genetics* **195**:205-220.
60. Keller SR *et al.* 2011. Local selection across a latitudinal gradient shapes nucleotide diversity in Balsam Poplar, *Populus balsamifera* L. *Genetics* **188**:941-948.
61. Keller SR *et al.* 2012. Local adaptation in the flowering-time gene network of Balsam Poplar, *Populus balsamifera* L. *Mol Biol Evol* **29**:3143-3152.
62. Yoder JB *et al.* 2014. Genomic signature of adaptation to climate in *Medicago truncatula. Genetics* **196**:1263-1275.
63. Hulke BS *et al.* 2014. in *Yield Gains in Major U.S. Field Crops* (eds Smith S *et al.*), 433-458 (Amer Soc Agronomy, Crop Sci Soc America, Soil Sci Soc Amer).
64. Weber M *et al.* 2003. A regional analysis of climate change impacts on Canadian agriculture. *Canadian Public Policy* **29**:163-180.

65. Heiser CB. 1969. The North American sunflowers. *Memoirs Torrey Bot Club* **22**:1-218.
66. Smith BD. 1989. Origins of agriculture in eastern North America. *Science* **246**:1566-1571.
67. Blackman BK *et al.* 2011. Sunflower domestication alleles support single domestication center in eastern North America. *PNAS* **108**:14360-14365..
68. Crites GD. 1993. Domesticated sunflower in 5th millennium Bp temporal context - new evidence from middle Tennessee. *Amer Antiquity* **58**:146-148.
69. Harter AV *et al.* 2004. Origin of extant domesticated sunflowers in eastern North America. *Nature* **430**:201-205.
70. Blackman, B.K., *et al.* 2010. The role of recently derived FT paralogs in sunflower domestication. *Current Biol* **20**:629-635.
71. Timme RE *et al.* 2007. High-resolution phylogeny for *Helianthus* (Asteraceae) using the 18s-26s ribosomal DNA external transcribed spacer. *Amer J Botany* **94**:1837-1852.
72. Mason CM *et al.* 2013. Ontogeny strongly and differentially alters leaf economic and other key traits in three diverse *Helianthus* species. *J Exp Botany* **64**:4053-4808.
73. Vekemans X. 2010. What's good for you may be good for me: evidence for adaptive introgression of multiple traits in wild sunflower. *New Phytol* **187**:7-9.
74. Rieseberg LH *et al.* 1995. Interspecific pollen competition as a reproductive barrier between sympatric species of *Helianthus* (Asteraceae). *Amer J Botany* **82**:515-519.
75. Renaut S *et al.* 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Comm* **4**, article 1827.
76. Donovan LA *et al.* 2009. Phenotypic selection on leaf ecophysiological traits in *Helianthus*. *New Phytol* **183**, 868-879.
77. Donovan LA *et al.* 2010. Are hybrid species more fit than ancestral parent species in the current hybrid species habitats? *J Evol Biol* **23**:805-816.
78. Sapir Y *et al.* 2007. Patterns of genetic diversity and candidate genes for ecological divergence in a homoploid hybrid sunflower, *Helianthus anomalus*. *Mol Ecol* **16**:5017-5029.
79. Lai Z *et al.* 2005. Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**:291-303.
80. Ludwig F *et al.* 2004. Selection on leaf ecophysiological traits in a desert hybrid *Helianthus* species and early-generation hybrids. *Evolution* **58**:2682-2692.
81. Rieseberg LH *et al.* 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**:1211-1216.
82. Lexer C *et al.* 2005. Genetics of species differences in the wild annual sunflowers, *Helianthus annuus* and *H. petiolaris*. *Genetics* **169**:2225-2239.
83. Donovan LA *et al.* 2014. Ecological and evolutionary lability of plant traits affecting carbon and nutrient cycling. *J Ecol* **102:**302-314.
84. Heesacker A *et al.* 2008. SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet* **117**:1021-1029.
85. Bachlava, E. *et al.* 2012. SNP discovery and development of a high-density genotyping array for sunflower. *Plos One* **7**, e29814.
86. Bowers JE *et al.* 2012. Development of an ultra-dense genetic map of the sunflower genome based on single-feature polymorphisms. *Plos One* **7**, e51360.
87. Kane N C *et al.* 2013. Sunflower genetic, genomic and ecological resources. *Mol Ecol Resources* **13**:10-20.
88. Renaut S. *et al.* 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Comm* **4**, article 1827.
89. Baute GJ *et al.* 2015. Genome scans reveal candidate domestication and improvement genes in cultivated sunflower. *New Phytol* **206:**830-838.
90. Rao KS *et al.* 1999. Agrobacterium-mediated transformation of sunflower (*Helianthus annuus* L.): A simple protocol. *Annals Botany* **83**:347-354.
91. Malone-Schoneberg J *et al.* 1994. Stable transformation of sunflower using *Agrobacterium* and split embryonic axis explants. *Plant Sci* **103**:199-207.
92. Rieseberg, L. H. unpub.

93. Mandel JR *et al.* 2011. Genetic diversity and population structure in cultivated sunflower and comparison to its wild progenitor. *Theor Appl Genet* **123**:693-704.
94. Mandel JR *et al.* 2013. Association mapping and the genomic consequences of selection in sunflower. *Plos Genetics* **9,** e1003378.
95. Raman K *et al.* 2014. The organisational structure of protein networks: revisiting the centrality–lethality hypothesis. *Systems Synthetic Biol* **8**:73-81.
96. Salunkhe DK *et al.* 1992. *World Oilseeds* (Springer).
97. Schneiter AA *et al.* 1981. Description of sunflower growth-stages. *Crop Sci* **21**:901-903.
98. Burke JM *et al.* 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* **171**:1933-1940.
99. Tang S *et al.* 2005. Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. *Crop Sc* **46**:721-734.
100. Trachsel S *et al.* 2013. Maize root growth angles become steeper under low N conditions. *Field Crops Res* **140**:18-31.
101. Trachsel S. *et al.* 2011. Shovelomics: high throughput phenotyping of maize (*Zea mays* L.) root architecture in the field. *Plant Soil* **341**:75-87.
102. White JW *et al.* 2005. Variation in parameters related to leaf thickness in common bean (*Phaseolus vulgaris* L.). *Field Crops Res* **91**:7-22.
103. Howard AR *et al.* 2007. Helianthus nighttime conductance and transpiration respond to soil water but not nutrient availability. *Plant Physiol* **143**:145-155.
104. Brouillette LC *et al.* 2011. Relative growth rate and functional traits of a hybrid species reflect adaptation to a low-fertility habitat. *Int J Plant Sci* **172**, 509-520.
105. Lambrides C *et al.* 2001. in *13th Australian Sunflower Assoc Conf Proc*.
106. Blum A. 2011. in *Plant Breeding for Water-Limited Environments* (Springer).
107. Caruso A *et al.* 2008. Physiological characterization and identification of genes differentially expressed in response to drought induced by PEG 6000. *J Plant Physiol* **165**:932-941.
108. Ishibashi CDA. unpub.
109. Rey JI. 2011. *Silverleaf sunflower (*Helianthus argophyllus*) in sunflower breeding: from comparative mapping to the genetics of salt tolerance in the species.* PhD thesis (U. Georgia)
110. Glenn E *et al.* 1984. Relationship between salt accumulation and water content of dicotyledonous halophytes. *Plant, Cell Environ* **7**:253-261.
111. Donovan L *et al.* 1997. Nutrient relations of the halophytic shrub, *Sarcobatus vermiculatus*, along a soil salinity gradient. *Plant Soil* **190**:105-117.
112. Gaut BS *et al.* 2003. The lowdown on linkage disequilibrium. *Plant Cell* **15**:1502-1506.
113. Liu AZ *et al.* 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **173**:321-330.
114. Kolkman JM *et al.* 2007. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* **177**:457-468.
115. Browning SR *et al.* 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Human Genet* **81**:1084-1097.
116. Browning BL *et al.* 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of unrelated individuals. *Am J Human Genet* **84**:210-223.
117. Buckler ES *et al.* 2002. Plant molecular diversity and applications to genomics. *Current Op Plant Biol* **5**:107-111.
118. Pritchard JK *et al.* 2000. Association mapping in structured populations. *Am J Human Genet* **67**:170-181.
119. Aranzana MJ *et al.* 2005. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *Plos Genet* **1**:531-539.
120. Yu JM *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genet* **38**:203-208.
121. Price AL *et al.* 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**:904-909.

122. Shim H *et al.* 2009. Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes. *BMC Proc* **3**, S35.
123. Hamblin MT *et al.* 2011. Factors affecting the power of haplotype markers in association studies. *Plant Genome* **4**:145-153.
124. Lorenz AJ *et al.* 2010. Performance of single nucleotide polymorphisms versus haplotypes for genome-wide association analysis in barley. *Plos One* **5,** e14079.
125. Gogarten SM *et al.* 2012. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* **28**:3329-3331.
126. Zheng XW *et al.* 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**:3326-3328.
127. Nemali KS *et al.* 2006. An automated system for controlling drought stress and irrigation in potted plants. *Scientia Hort* **110**:292-297.
128. Espeleta JF *et al.* 2002. Fine root demography and morphology in response to soil resources availability among xeric and mesic sandhill tree species. *Funct Ecol* **16**:113-121.
129. West JB *et al.* 2003. Root longevity and phenology differences between two co-occurring savanna bunchgrasses with different leaf habits. *Funct Ecol* **17**:20-28.
130. Cornelissen JHC *et al.* 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. *Australian J Botany* **51**:335-380.
131. Sack L *et al.* 2013. How do leaf veins influence the worldwide leaf economic spectrum? Review and synthesis. *J Exp Botany* **64**:4053-4080.
132. Bartlett MK *et al.* 2012. Rapid determination of comparative drought tolerance traits: using an osmometer to predict turgor loss point. *Methods Ecol Evol* **3**:880-888.
133. Bartlett MK *et al.* 2012. The determinants of leaf turgor loss point and prediction of drought tolerance of species and biomes: a global meta-analysis. *Ecol Letters* **15**:393-405.
134. Sperry JS *et al.* 1988. A method for measuring hydraulic conductivity and embolism in xylem. *Plant, Cell Environ* **11**:35-40.
135. Rosenthal DM *et al.* 2010. Contrasting drought tolerance strategies in two desert annuals of hybrid origin. *J Exp Botany* **61**:2769-2778.
136. Dodd GL *et al.* 1999. Water potential and ionic effects on germination and seedling growth of two cold desert shrubs. *Amer J Bot* **86**:1146-1153.
137. Brouillette LC *et al.* 2006. Testing hypothesized evolutionary shifts toward stress tolerance in hybrid *Helianthus* species. *W N Amer Nat* **66**:409-419.
138. Anonymous. 2014. *INRA* <http://www.toulouse.inra.fr/en%3E>.
139. Zhong S *et al.* 2011. High-throughput Illumina strand-specific RNA sequencing library preparation. *Cold Spr Har Protocols* **2011:**940-949.
140. Anders S *et al.* 2010. Differential expression analysis for sequence count data. *Genome Biol* **11,** article R106.
141. Moyers BT *et al.* 2013. Divergence in gene expression is uncoupled from divergence in coding sequence in a secondarily woody sunflower. *Int J Plant Sci* **174**:1079-1089.
142. Langfelder P *et al.* 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, article 559.
143. Li J. *et al.* 2014. GPLEXUS: Enabling genome-scale gene association network reconstruction and analysis for very large-scale expression data. *Nucleic Acids Res* **45**, e32.
144. Marchand G *et al.* 2014. Bridging physiological and evolutionary time scales in a gene regulatory network. *New Phytol* **203:**685–696.
145. Priest HD *et al.* 2014. Analysis of global gene expression in *Brachypodium distachyon* reveals extensive network plasticity in response to abiotic stress. *Plos One* **9**, e87499.
146. Lotterhos KE, Whitlock MC. 2015 The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol* **24**:1031-1046.
147. Matvienko M *et al.* 2013. Consequences of normalizing transcriptomic and genomic Lbraries of plant genomes using a duplex-specific Nuclease. *Plos One* **8**, e55913.
148. Lachance J. *et al.* 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. *Bioessays* **35**:780-786.

149. Hijmans RJ *et al.* 2005. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol* **25**:1965-1978.
150. Beaumont MA *et al.* 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* **13**:969-980.
151. Excoffier L *et al.* 2009. Detecting loci under selection in a hierarchically structured population. *Heredity* **103**:285-298.
152. Foll M *et al.* 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**:977-993.
153. Lewontin RC *et al.* 1973. Distribution of gene frequency as a test of the theory of the selecive neutrality of polymorphism. *Genetics* **74**:175-195.
154. Lotterhos KE, Whitlock MC. 2014. 2014. Evaluation of demographic history and neutral parameterization on the performance of Fst outlier tests. *Mol Ecol* **23**:178-2192.
155. Renaut S *et al.* 2014. Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Mol Ecol* **23**:311-324.
156. Cavanagh C *et al.* 2008. From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. *Curr Opin Plant Biol* **11**:215-221.
157. Elshire RJ *et al.* 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* **6**, e19379.
158. Huang BE *et al.* 2013. Selecting subsets of genotyped experimental populations for phenotyping to maximize genetic diversity. *Theor Appl Genet* **126**:379-388.
159. Huang BE *et al.* 2011. R/mpMap: a computational platform for the genetic analysis of multiparent recombinant inbred lines. *Bioinformatics* **27**:727-729.
160. Gaj T *et al.* 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotech* **31**:397-405.
161. Voytas DF *et al.* 2014. Precision Genome engineering and agriculture: Opportunities and regulatory challenges. *Plos Biol* **12**, e1001877.
162. Jacobs JZ *et al.* 2014. Implementation of the CRISPR-Cas9 system in fission yeast. *Nature Comm* **5**, article 5344.
163. Cramer WP *et al.* 1993. Climatic classification and future global redistribution of agricultural land. *Climate Res* **3**:97-110.
164. Ramankutty N *et al.* 2002. The global distribution of cultivable lands: current patterns and sensitivity to possible climate change. *Gl Ecol Biogeography* **11**:377-392.
165. Zabel F *et al.* 2014. Global Agricultural Land Resources – A high resolution suitability evaluation and its perspectives under climate change conditions. *PLoS ONE* **9**, e114980.
166. Foley JA *et al.* 1996. An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Gl Biogeochem Cycles* **10**:603-628.
167. Kucharik CJ *et al.* 2000. Testing the performance of a dynamic global ecosystem model: water balance, carbon balance, and vegetation structure. *Gl Biogeochem Cycles* **14**:795-825.
168. Kucharik CJ. 2003. Evaluation of a process-based agro-ecosystem model (Agro-IBIS) across the U.S. cornbelt. *Earth Interactions* **7**:1-33.
169. Jones JW *et al.* 2003. The DSSAT cropping system model. *Eu J Agron* **18**:235-265.
170. Gassman PW *et al.* 2005. Historical development and applications of the EPIC and APEX Models (Center for Agricultural and Rural Development, Iowa State U, Ames).
171. Deryng D. *et al.* 2011. Simulating the effects of climate and agricultural management practices on global crop yield. *Gl Biogeochem Cycles* **25**, article GB2006.
172. Mueller ND *et al.* 2012. Closing yield gaps through nutrient and water management. *Nature* **494**:390-390.
173. Deryng D *et al.* 2014. Global crop yield response to extreme heat stress under multiple climate change futures. *Environ Res Letters* **9**, e034011.
174. May WE *et al.* 2010. Adaptation of oilseed crops across Saskatchewan. *Can J Plant Sci* **90**:667-677.

Genomics of Abiotic Stress Resistance in Wild and Cultivated Sunflowers

Loren H. Rieseberg & John M. Burke

ACTIVITY	PLANNED START (01/10/15)	PLANNED DURATION (#Quarters)	ACTUAL START (dd/mm/yy)	ACTUAL DURATION (#Quarters)	% COMPLETE	QUARTER - 3 months Start Date 10/15															
						1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Activity 1: Abiotic stress resistance in cult. sunflower																					
Activity 1.1 - Phenotyping in large-scale screens	01/10/15	8																			
Milestone – Phenotyping of SAM population for 4 stresses completed	01/10/17	1																			
Activity 1.2 - GWAS of abiotic stress resistance	01/10/16	8																			
Milestone – GWAS of SAM population for 4 stresses completed	01/10/18	1																			
Activity 1.3 - Detailed ecophysiological characterization	01/10/16	8																			
Milestone – Ecophysiological characterization of 15 most & least resistant phenotypes completed	01/10/18	1																			
Activity 1.4 – Heliaphen phenotyping	01/10/17	8																			
Milestone – Heliophen phenotyping completed	01/07/19	1																			
Activity 1.5 - Transcriptomic analyses	01/10/16	8																			
Milestone – Transcriptome sequencing of 576 samples completed	01/10/18	1																			
Activity 1.5 – Analyses & publication	01/10/18	4																			
Deliverable – Data mining resource made publicly available	01/10/17	1																			
Activity 2: Population genomics - stress adaptation in wild																					
Activity 2.1 – Collections & soil characterization	01/10/15	4																			
Milestone – Seeds & soil samples obtained from 150 sites & soil samples	01/10/17	1																			

characterized																						
Activity 2.2 – Genotyping of wild populations	01/10/16	4																				
Milestone – 1,500 individuals from 150 populations sequenced to 6x depth	01/10/17	1																				
Activity 2.3 -– Analyses of population genomic data	01/10/17	8																				
Milestone – Loci under selection in wild populations identified	01/07/19	1																				
Activity 2.4 – Preparation of manuscripts for publication	01/10/18	4																				
Activity 3: Development & characterization of MAGIC populations																						
Activity 3.1 - Population development	01/10/15	10																				
Milestone – MAGIC population development (2 x 600 RILs) completed	01/01/18	1																				
Activity 3.2 – Genotyping & Phenotyping of MAGIC population	01/04/17	3																				
Milestone – Genotyping & Phenotyping of female MAGIC population for 4 stresses completed	01/04/19	1																				
Activity 3.3 - Detailed ecophysiological characterization	01/10/18	4																				
Milestone – Ecophysiological characterization of 15 most & least resistant phenotypes completed	01/07/19	1																				
Activity 3.4 – Analyses & publication	01/10/18	4																				
Deliverable –Next generation germplasm released	01/01/18	1																				
Activity 4: Functional validation of stress resistance genes																						
Activity 4.1 – Functional Analyses of candidate genes in sunflower	01/10/15	8																				
Milestone – Functional analyses of 5 genes completed	01/10/17	1																				
Milestone – Functional	01/07/19	1																				

analyses of additional 10 genes completed																					
Activity 4.2 – Genome editing in soybean	01/10/17	8																			
Milestone – Introduction of 3 mutations into soybean	01/07/19	1																			
Activity 4.3 – Preparation of manuscripts for publication	01/10/17	8																			
Activity 5: GE^{3}LS - Development of crop yield models																					
Activity 5.1 Model development	01/10/15	8																			
Milestone – Yield models completed	01/10/17	1																			
Activity 5.2 Input data gathering and development	01/10/16	4																			
Milestone – spatial input data on climate, soils, irrigation, and fertilizer use gathered	01/10/17	1																			
Activity 5.3 Model simulations and analysis	01/10/17	8																			
Milestone – Simulations under stressed & un-stressed conditions & with climate change completed	01/07/19	1																			
Activity 5.4 Preparation of manuscripts for publication	01/10/17	8																			
Deliverable – Yield models for sunflower & competitor crops delivered	01/07/19	1																			
Activity 6: GE^{3}LS – International treaties and use of plant genetic resources																					
Activity 6.1 Clarify potential impacts and opportunities	01/10/15	12																			
Milestone – Report to project team identifying impacts & opportunities relating to SMTA	01/10/17	1																			
Milestone – Publication identifying impacts & opportunities relating to SMTA	01/10/18	1																			
Activity 6.2 Facilitate policy Impacts	01/10/15	16																			
Milestone – Publication on governance of data sharing developed in the DivSeek setting	01/10/17	1																			

Deliverable – Refinements to Treaty with respect changes in policy & SMTA	01/07/19	1																				
Activity 7: Project Management and Administration																						
Project launch	01/10/15	2																				
Personnel hiring	01/10/15	5																				
ROC & Team F2F meetings	01/01/16																					
Team management meetings (teleconference)	01/10/15	16																				
Progress and financial reports	01/10/15	16																				
Close-out reporting	01/07/19	1																				

Current Quarter Highlight **Plan** **Ongoing** **Complete**

VIII EXPECTED SOCIAL AND/OR ECONOMIC BENEFITS

Maximum of eight (8) pages, including charts, figures and tables. This section must address all relevant evaluation criteria for the competition. Include a plan for knowledge translation and development of benefits, which outlines the next steps of how the deliverables from the research will be transferred, disseminated, used, and/or applied to realize social and/or economic benefits.

A. Deliverables

This project will deliver the following resources, tools, and strategies that will allow our end-users to access innovative germplasm; analyze and exploit large genomic data sets for research, breeding, and other applications; develop new highly resistant and productive sunflower cultivars; and predict where such cultivars will be maximally productive in Canada:

(1) "Next Generation" germplasm resources in the form of multi-species, advanced generation intercross populations for use by sunflower breeders. This germplasm will be available in year 3 of the project and will enable the efficient deployment of alleles that provide resistance to drought, flooding, salt, and low nutrient alleles, but with minimal yield trade-offs, to sunflower breeding programs. New highly resistant and productive cultivars are expected in the field within four years of project end.

(2) A central data mining and analysis resource for sunflower that will enable researchers and breeders to more rapidly process, analyze, and manage large genomic and phenotypic data sets. The mining and analysis tools will be released throughout the course of the project.

(3) Crop yield models that will enable predictions of likely yields of new stress resistant sunflower cultivars (and competitor crops) in different soil and climate conditions across Canada. These models, in combination with existing economic models, will allow seed companies and farmers to determine where particular cultivars are likely to be successful and to compare predicted productivity and net returns of different crops (and cultivars) at a given site. Crop yield models will be delivered in year 4.

(4) Strategies for mitigating barriers to public and private breeding programs (e.g., uncertain IP and profit sharing obligations) resulting from international treaties; and Co-PI Marden will continue to work closely with the CGB/Treaty to move forward refinements to existing treaties to reduce such barriers. Strategies and suggestions for refinements will be made throughout the course of the project.

B. Expected benefits

Genome Canada has highlighted numerous challenges to crop production in its sector strategy on the agri-food industry, including extreme environments, climate change, competition for land, cost and environmental impacts of external inputs such as fertilizer and water, food security due to increasing global population size, limited crop genetic diversity, and regulatory impediments to the use of innovative technology (http://www.genomecanada.ca/medias/PDF/EN/Agri_Food_EN.pdf). Our deliverables address each of these issues and will yield environmentally resilient sunflower cultivars that require fewer external inputs and that can be grown in marginal and degraded crop lands in Canada and developing countries, thereby contributing to global food security. Additional benefits will derive from the crop yield models, which will reduce the "hit and miss" aspects of trying new cultivars (and their associated costs), and the treaty strategies for mitigating barriers to R&D (e.g., uncertainties in IP, tech transfer, and profit sharing).

Benefits to Canada: Sunflower is currently a minor crop in Canada because of a lack of well-adapted cultivars[1], so the greatest economic gains will likely result from expansion of sunflower production to marginal lands. Canada has ~38 million ha of marginal farmlands[2], of which ~22.3 million ha is grassland habitat in the Southern Prairies (Fig. 1). These lands are currently unsuitable for annual food crops for a variety of reasons, including climate, topography, adverse soils (low nutrients, salt), and other factors such as poor drainage. Approximately 18% (4 million ha) of such marginal lands are, however, potentially suitable for improved sunflower cultivars, and an additional ~4 million ha are expected to become suitable with climate change (Fig. 1). While this land has other uses such as livestock grazing and biomass production, these uses typically provide lower marginal returns than could be achieved through the cultivation of

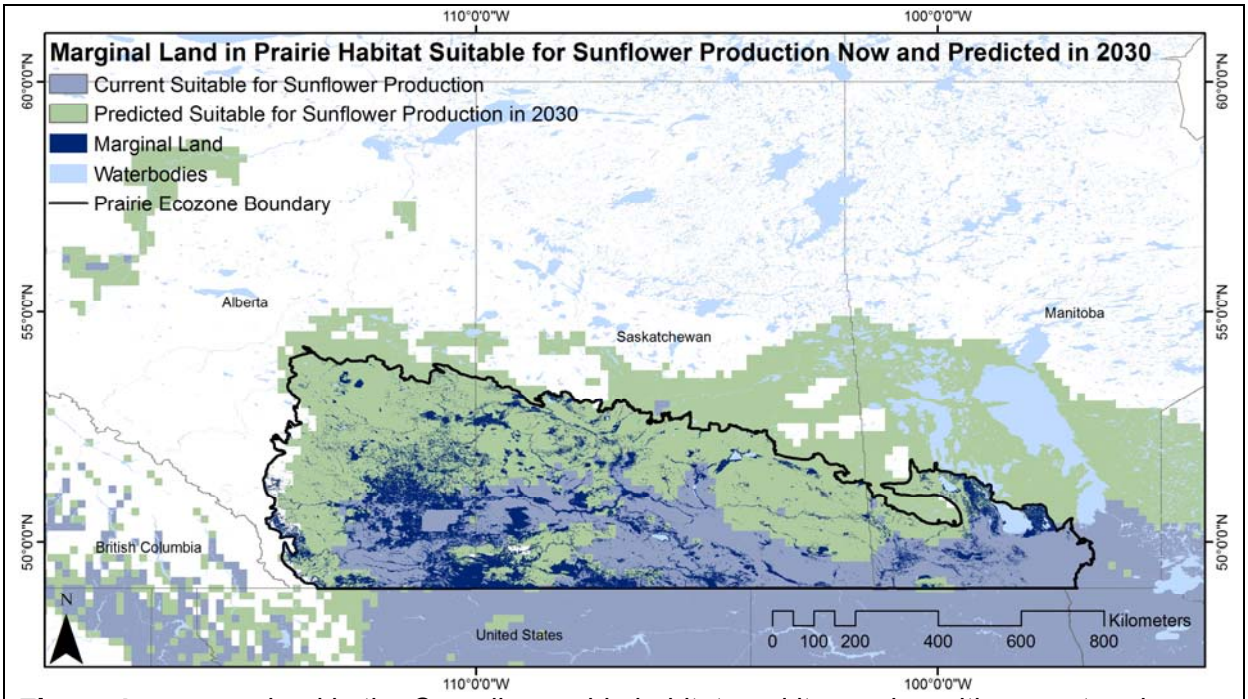

Figure 1. Marginal land in the Canadian prairie habitat, and its overlap with current and predicted climate suitability (temperature & precipitation) for sunflower production. Sunflower is at the northern limit of its production range, so all Canadian land currently within the sunflower climate suitability envelope will continue to be so with climate change.

resistant and highly productive sunflower cultivars[3,4]. Other potential impediments include up front costs associated with converting marginal lands to crop production, and increased transportation costs for marginal lands away from ongoing crop production[2].

We expect that resistant sunflower cultivars will have a competitive advantage on marginal lands relative to earlier sunflower cultivars or competitor crops, as most cannot be productively grown on these lands. Nonetheless, we anticipate a relatively slow uptake of the improved sunflower cultivars on "new" land because it has not previously been employed for sunflower cultivation. If we conservatively assume that within 5 years of the project end, approximately 0.1-0.2% of currently suitable marginal lands would be planted with improved cultivars, this would yield a gross production value of 6-12 million USD annually (Fig. 2), and a net production value of 3-7 million USD based on the most recent available (2012) gross and net production values for sunflower in Canada[5]. Assuming the new cultivars perform well in marginal soils, then after ten years, we expect ~2-4% of these lands will be planted, adding 0.08-0.16 million ha to Canadian production, with gross and net returns of ~115-230 million USD/year and ~70-140 million USD/year, respectively.

Stress resistant cultivars will also increase productivity on land where sunflower is currently grown in Canada. Sunflower fields typically contain pockets of flooded or salty soils; resistant

cultivars will enable cultivation of the full field (W. May, pers. comm.). Best estimates of production increases due to better use of such fields in Canada is circa 20% over present-day productivity, which would lead to additional gains of ~1-2 million USD years five years out and ~$11 million annually in net returns to farmers within 10 years of the project end date (or a total of up to $150 million in additional value when including marginal land use).

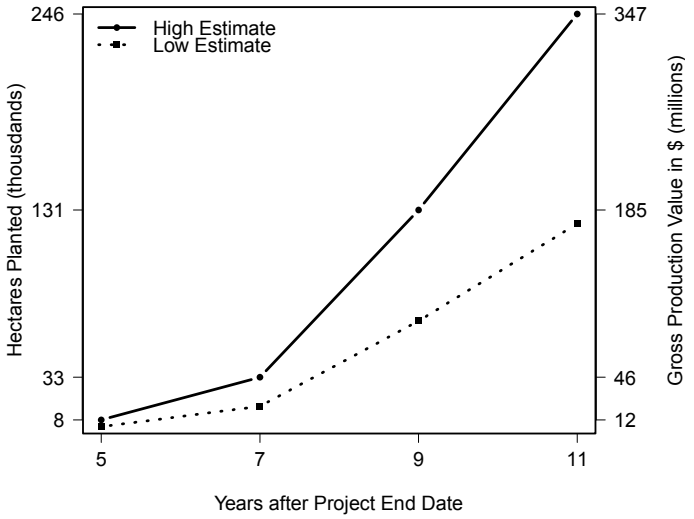

Figure 2. Predicted number of hectares and gross production value of new resistant sunflower cultivars on marginal prairie lands in Canada.

A market for this increase already exists: North America can now absorb an additional 350,000 metric tons of oilseed sunflower per year, which would require 0.2 million ha. (J. Sandbakken, Natl. Sunflower Assoc., pers. comm.). The public is also demanding higher oleic content in vegetable oil (lower saturated fat), which sunflower can provide, and the FDA is moving close to a ban on partial hydrogenated oils in the US (http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm372915.htm), the major dietary source of *trans* fat, opening even more opportunities for sunflower oil (which does not need hydrogenation). Best US industry projections are that, by 2016, an additional 0.7-0.8 million ha need to be seeded in oil-type sunflowers to meet demand – and 90-95% of US imports come from Canada (J. Sandbakken, Natl. Sunflower Assoc., pers. comm.). The new cultivars will be non-GMO, easing grower and market acceptance.

Benefits to developing countries: Internationally, the impact of the proposed research on food security is potentially substantial, especially in developing countries. Sunflower is one of the world's most important crops, ranking 12th in terms of area harvested[5], with a gross value of $20 billion per year, and global use is increasing rapidly[6]. Sunflower is grown widely in developing countries and used primarily for food. As a consequence, it is the only oilseed included in the Global Crop Diversity Trust's list of 25 priority food security crops (https://www.croptrust.org/crop-diversity-endowment-fund/). The new cultivars to be produced by project end users are especially well-suited for agricultural environments in Sub-Saharan Africa, since they are expected to combine high yield with resistance to abiotic stress, require lower inputs, and produce less environmental harm. In the present proposal, we have targeted sunflower breeding and production in Uganda since this will allow us to extend an ongoing and productive collaboration of the project team with sunflower breeders at the National Semi-Arid Resources Research Institute (NaSARRI).

A recent analysis of food security in Uganda reported that 6.3% of households were food insecure, and another 21.3% were moderately food insecure and at risk of becoming food insecure if conditions were to deteriorate (http://documents.wfp.org/stellent/groups/public/documents/ena/wfp085649.pdf). As a consequence, malnutrition is a significant problem in Uganda where it contributes to disease and early death, with economic costs estimated at 5% of GDP (http://www.wfp.org/news/news-release/hunger-costs-uganda-56-cent-gdp). Iron deficiency, in particular, affects almost 75% of children under five and 50% of women of childbearing age (http://www.fao.org/docrep/017/aq011e/aq011e.pdf). To reduce iron deficiency, the Ugandan government is promoting increased production and use of foods rich in iron such as sunflower seeds (100g serving contains 28% of daily needs).

Factors contributing to food insecurity and malnutrition in Uganda include unreliable climatic conditions (particularly droughts and floods), low productivity, and declining soil fertility (http://documents.wfp.org/stellent/groups/public/documents/ena/wfp085649.pdf). The information and germplasm from this project will allow rapid development of resistant and productive sunflower cultivars that should mitigate these problems. In a best-case scenario, germplasm from this project will increase sunflower yields in Uganda to those currently enjoyed by Canadian farmers (Fig. 3). Considering only acreage currently devoted to sunflower production, and current yields[3], this would add >200 calories per day to the diets of each Ugandan (current caloric intake is 1,970 calories, which is below minimum requirement of 2,200 calories), positively affecting the lives of the > 37 million Ugandans. Even if yield gains were lower, and brought sunflower yields in Uganda to average sunflower productivity levels world-wide, which we believe to be readily achievable, this would result in an addition of ~146 calories per day per Ugandan, and these calculations do not consider possible gains from expanding sunflower cultivation to marginal lands in Uganda. Similar gains are estimated for other sunflower growing regions in sub-Saharan Africa. Most importantly, stress-resistant cultivars will stabilize production in the face of abiotic stress, thereby reducing the potential for malnutrition and its social and economic costs.

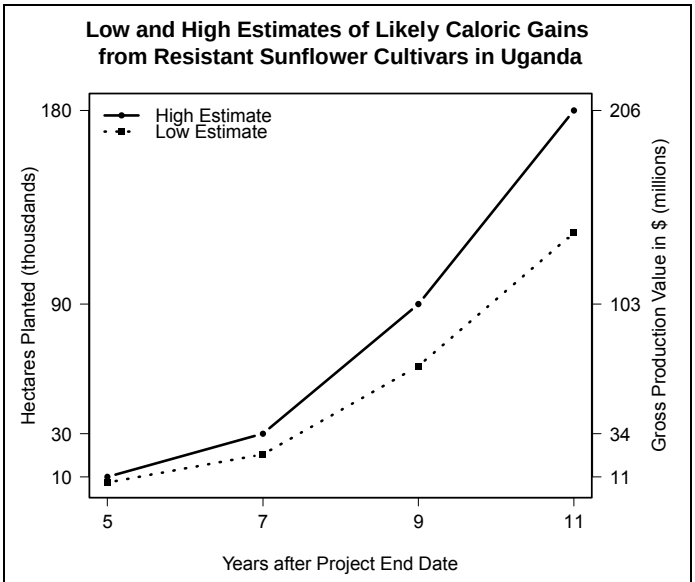

Figure 3. Anticipated daily caloric gains of new resistant and sunflower cultivars in Uganda.

C. Plan for Knowledge Translation and Development of Benefits

Strategy for Realizing Benefits: The main socioeconomic benefit from this project will derive from the development and commercialization of new highly resistant and productive sunflower cultivars based on the information and germplasm we provide (Table 1). Assuming the existence of major QTLs for key traits, a backcross breeding design, and marker-assisted selection, it should be possible to have our first inbred lines by the end of the project, and the first hybrid cultivars by the first year after the project ends. Assuming three years of field-based evaluation of the hybrid cultivars, this would allow commercial release of new cultivars within four years of the project end date. Note that some breeding companies have technology that reduces sunflower breeding cycles relative to those assumed here, which could speed up commercial release. On the other hand, if QTL effects are small (which does not appear to be the case), and conventional and/or genomic selection must be employed, commercial release could be delayed.

Table 1. Steps and timeline for commercial release of improved cultivars based on genetic information and germplasm generated by this project.

Project Timeline	Release of data and/or germplasm	End users	IP
Year 2	SNP markers & QTLs from SAM population	Start using genetic info for backcross assisted trait introgression, MAS	
Year 3	Next generation germplasm – MAGIC RILs	Germplasm enhancement, MAS,	

Year 4 project completion	SNP markers & QTLs from MAGIC population	First backcross derived inbreds developed, trait stability tested in small experiments	
Project +1 years		Begin hybrid testing of backcross derived lines	
Project +4 years		Multiplication and commercial release	Obtain plant variety rights or plant patents for cultivars

The following steps will be taken to ensure that this and other benefits from this project are realized: (1) engagement of end users (plant breeders in both the public and private sector) in both planning and carrying out the proposed research; (2) training of end users in the interpretation and use of the genomic and phenotypic data generated by the project; (3) partnering with the computing industry to deliver state-of-the art and user friendly data mining and data analytic methods; (4) development of crop yield models that reduce risk associated with the conversion of marginal land to sunflower production; (5) clarification of CBD and Treaty obligations to maximize uptake of the germplasm generated by this project; and (6) broad dissemination of our results and recommendations.

(1) Engagement of end users – We canvassed the sunflower genetics and breeding community in Canada, major sunflower markets in Europe, Argentina, the USA, and Sub-Saharan Africa to identify the abiotic stresses presenting the greatest limitation to sunflower production. We then designed our project in collaboration with sunflower breeders from four government agencies (AAFC, INRA, NaSARRI, and the USDA) and five companies (Advanta, Biogemma, Dow AgroSciences, KWS Seeds, and NuSeed Americas), who have contributed advice, valuable germplasm, and co-funding and are committed to population development and phenotyping (see co-funding letters). Thus, they have 'skin in the game' and are highly likely to incorporate valuable germplasm and alleles into their breeding programs.

(2) Training of end users – The use of genomic information in plant breeding can be challenging for breeders who typically lack expertise in genomics and bioinformatics. Therefore, we will continue to offer our sunflower genomics and bioinformatics workshop in San Diego each January following the Plant and Animal Genome meetings. This workshop includes updates on sunflower genomic resource and tool availability and hands-on guidance in their use. It is well attended. Training will also occur through online tutorials and advice from the project team.

(3) Partnering with industry to provide a data mining and analysis resource – To provide genomics support for sunflower breeders, we have established a collaboration with the software company SAP and several development teams at the headquarters in Waldorf, Germany and their branch in Vancouver, Canada. Apart from processing the data using the SAP HANA in-memory database, we will also provide a cloud instance (i.e., computing network) running an interactive user interface similar to SAP's ProteomicsDB[7], which provides the proteomics mass spectra of the human proteome and analytics functionality on top of the processed data. This cloud instance would serve as a central data mining and analysis resource for sunflower breeders and researchers that would speed up breeding by facilitating marker-assisted and/or genomic selection. Different genomics aspects of the reference genome, elite varieties and wild relatives of sunflower would be publicly accessible including analytical tools (e.g. association mapping), data mining (e.g., sequences, variants, annotations, expression profiles) and interactive visualization tools (e.g., genome browser and many plotting functions). Using an interactive HTML5 UIs in the cloud instance will provide an independent and efficient platform for browsing and analyzing complex datasets in a simple and friendly environment. SAP is contributing to the project through provision of the cloud instance, as well as through algorithm development and data processing. SAP Canada is considering this project and collaboration as of high strategic value, which will enable the company to reach new markets and opportunities,

and they have recently released a media pitch describing our collaboration (http://www.24news.ca/tech-science/74406-fast-track-to-food-security).

(4) Development of crop yield models – The financial risk associated with converting marginal land to sunflower production could be a significant impediment to realizing the benefits from this project. We will reduce this risk by providing cultivar yield models that can be used in combination with existing economic models such as those developed by collaborator May[1] to select the most promising land for sunflower production using new cultivars to be developed by our end users. We will make a user-friendly version of the model available on the sunflower data mining site (above) so it can be efficiently exploited by breeders, crop associations, and farmers.

(5) Clarifying CBD and Treaty obligations to maximize uptake of the germplasm generated by this project – Ambiguities in the CBD and Treaty represent impediments to the use of innovative germplasm. Co-investigator Marden will work closely with the Treaty Secretariat to clarify outstanding issues and facilitate innovation.

(6) Broad dissemination of our results and recommendations – Germplasm developed by this project will be made publicly available by the USDA's sunflower germplasm repository. In addition, we will provide sufficient seed for immediate distribution to interested parties. Sequence and expression data will be deposited in GenBank and Gene Expression Omnibus, respectively. Dissemination of genomics, physiological, and evolutionary knowledge, as well as data analytic tools, will occur via the central data mining and analysis resource we are planning for the SAP Cloud, the Sunflower Genome Database (http://www.sunflowergenome.org) and Heliagene (https://www.heliagene.org). Project findings, including those from the GE^{3}LS research, will also be disseminated through publication in high-profile journals and presentations at conferences, universities, and companies. Our GE^{3}LS team will also produce internal reports directed at the project team for managing CBD and Treaty obligations to maximize uptake, and publications and presentations to stakeholders, scholars and policymakers, and other crop research teams about these recommendations. Co-investigator Marden will engage relevant bodies, including the Treaty Secretariat, the Global Crop Diversity Trust, the Canadian Senate Standing Committee on Agriculture and Forestry, and Plant Gene Resources of Canada to move forward refinements to reduce barriers and facilitate innovation. This will be disseminated in a jointly authored publication outlining potential policy options generated from high level stakeholder discussions that is aimed at further policy discussions about innovation in agricultural genomics.

Timing: We expect breeding programs both in Canada and internationally to begin to benefit from this project by year 2 as new information about the genetics and physiology of abiotic stress resistance, new genomic tools and resources, and new germplasm, begin to increase the efficiency of ongoing breeding programs (Table 1). Production benefits from these improved breeding programs will begin to be felt in Canada and internationally shortly thereafter: within four years after the project ends.

Project benefits from the crop models and Treaty work will begin in year 2. However, fully parameterized yield models will be delivered in year 4. We are optimistically targeting year 4 for refinements to the Treaty, but we recognize that this deliverable may take longer to achieve because of political impediments

Government approvals: The release and use of the germplasm developed by this project does not require government approval, as no GMOs or previously unknown traits will be involved in the process. That is, our end users will employ the new germplasm to improve the efficacy of existing traits rather than to develop new traits that would be regulated by the Canadian Food Inspection Agency.

Risks and challenges: As far as we are aware, there are no public efforts underway in the sunflower research community to study the genomics of abiotic stress resistance, at least on the scale proposed here. Likewise, we are unaware of public sector plans to develop MAGIC populations for sunflower or to develop tools and resources for genomics data mining and

analyses in sunflower (outside of those already developed by the project team). There are smaller studies underway by a number of public groups to evaluate the drought or salt tolerance of various sunflower cultivars, and to map QTLs in biparental crosses, but these studies do not compete directly with the proposed work.

On the other hand, there has been a long history of pre-breeding in sunflower, carried out mainly by the USDA, in which genetic material from wild species has been introgressed into cultivated lines. However, the USDA has focused mainly on providing new sources of disease resistance and cytoplasmic male sterility, whereas we are interested in abiotic stress resistance. Also, conventional pre-breeding approaches are ill suited for hybrid crop development or genetic studies, and are less efficient than modern population designs for evaluating the effects of numerous wild alleles from multiple donors in multiple genetic backgrounds. Thus, we think the next generation germplasm developed by the current project will be much more valuable and widely used than the more conventional germplasm generated previously

Intellectual property (IP) strategy: The preference of our industry and government partners is for the data, results, and germplasm generated by our project to be considered pre-competitive. Indeed, they indicated that they would otherwise be reluctant to use the germplasm. However, our end users will be able to obtain legal protection of cultivars they develop from germplasm released by the project through plant variety rights or plant patents (Table 1). Most breeding companies prefer the latter since it provides stronger protection. The main restriction on use of the germplasm itself will be the standard SMTA as required by the Treaty. There will be no restrictions on the use of data and knowledge generated by the project.

There are two exceptions to this general rule: (1) While legal protection of natural allelic variants, such as those identified by the project, is weak, if such variants are modified using gene editing to enhance trait efficacy, then strong patent protection becomes possible. Indeed, Rieseberg's lab is collaborating with a major seed company to patent modified natural variants underlying a valuable crop protection trait. A similar strategy is likely to be employed here, as a means of combining valuable mutations that are validated in Activity 4, although the gene editing and patenting would occur in a follow-up project. Likewise, we may patent soybean alleles that are modified by this project, if such modified alleles are agronomically valuable. (2) A second exception relates to bioinformatics algorithms developed by SAP. UBC has previously signed a collaborative agreement with SAP such that SAP owns the rights to bioinformatics algorithms developed in collaboration with UBC, but not to any of the sunflower genomic data or results, which are made publicly available with no restrictions on their use.

D. Expertise for Realizing Benefits

Our project team was carefully constructed to ensure that the proposed socioeconomic benefits are realized. The PIs have a long track record of working closely with industry and in realizing both public and private benefits from genomic tools and resources. For example, Rieseberg, Langlade, and Burke run a Sunflower Genomic Resources Consortium and workshop that are supported by industry, and the genomic and bioinformatics tools we have developed are widely used. We also have ongoing trait-development projects with various industry partners and have a strong history of developing and releasing pre-bred germplasm to public and private breeding programs. Lastly, Rieseberg has developed a productive collaboration with SAP to develop data analytic tools for plant genomics and to analyze sunflower genomic data, a collaboration that will be extended in the current project.

We also have included seven sunflower breeders as PIs or collaborators on the project, who will use the information and germplasm in their breeding programs. Co-PI Hulke and collaborator May spearhead the only public sector breeding program that is actively developing and evaluating oilseed sunflower cultivars for the Canadian environment. Hulke and his colleagues at the USDA have released numerous cultivars over the past several decades with no restrictions on their use. As a consequence, USDA germplasm has been incorporated into breeding programs worldwide. Our collaborators at NuSeed Americas and Dow Agrosciences

mainly target the North American Sunflower market, including Canada, whereas Biogemma and KWS have a European focus. Advanta focuses on the South American market, especially Argentina, whereas collaborator Walter Anyanga at NaSSARI runs the main public breeding program in Uganda. We are currently collaborating with Anyanga on other evaluation and breeding projects, and we have been pleased and impressed with his ability to phenotype very large populations at very low cost and with the performance of the hybrid cultivars he has successfully developed to date.

The GE^{3}LS team PIs, Marden and Ramankutty, have complementary expertise relating to the Treaty/sharing of agricultural data[8] and land use[9], respectively. Marden's expertise in these areas was developed in part during the Genome Canada-funded Genomics of Sunflower project. As the GE^{3}LS PI on the Genomics of Sunflower project, Marden engaged with industry (Syngenta), the CG Centres (Bioversity) and the Treaty Secretariat on issues of sharing. In addition, as a member of the initial Legal/Governance team supporting DivSeek (http://www.divseek.org), Marden has engaged with senior researchers in agricultural genomics as well as other Genome Canada funded GE^{3}LS researchers, including Peter Phillips of the VALGEN project.

Ramankutty is a Tier 1 Canada Research Chair in Global Environmental Change and Food Security at UBC. Ramankutty's research concerns global land use change and its implications, focusing on agricultural practice and the implications for environmental change and food security. He employs data and models to address the question of how to feed 9-10 billion people while reducing agriculture's environmental footprint. His research interests include global agriculture and food security, land use and cover change, global environmental change, global climate change, earth system science, ecosystem services, climate-vegetation interactions, and global biogeochemical cycles. He also advises end-users in government agencies (e.g., the USDA) and NGOs (e.g., CGIAR) on policy developments stemming from these models.

E. References

1. May, W. E. *et al.* 2010. Adaptation of oilseed crops across Saskatchewan. *Canadian Journal of Plant Science* **90**, 667-677.
2. Milbrandt, A. *et al.* 2009. Assessment of biomass resources from marginal lands in APEC economies. *Energy Working Group*.
3. Liu, T. T. *et al.* in *1st International Conference on Agro-Geoinformatics (Agro-Geoinformatics).* 650-655 (Ieee, 2012).
4. Anonymous. 2013. *Guidelines for estimating crop production costs 2013 in Western Manitoba*, <http://www.gov.mb.ca/agriculture/business-and-economics/financial-management/pubs/copcropproductioncosts2013_westernmb.pdf%3E.
5. http://faostat3.fao.org/home/E. 2015.
6. Khoury, C. K. *et al.* 2014. Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4001-4006, doi:10.1073/pnas.1313490111.
7. Wilhelm, M. *et al.* 2014. Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582-+, doi:10.1038/nature13319.
8. Marden, E. *et al.* 2015. *Innovation in agricultural genomics: overcoming complexities in the intellectual property-regulatory complex*.
9. Mueller, N. D. *et al.* 2013. Closing yield gaps through nutrient and water management (vol 490, pg 254, 2012). *Nature* **494**, 390-390, doi:10.1038/nature11907.

IX MANAGEMENT

Maximum of four (4) pages including charts, figures and tables. This section must address all relevant evaluation criteria for the competition. Please include an organization chart and project management plan.

A. Management Team

Members of the management team have clear and unambiguous responsibilities and reporting/supervision roles. The project leaders are Rieseberg and Burke, who will be working with project manager Staton to oversee the project. Each co-PI leads one or more activities. The management roles are set out in Fig.1 and in Table 1 (below).

Table 1. *Responsibilities of Project Management Team*

	Responsibility	Reporting to	Oversight of
Rieseberg	Project leader and lead of activity 4 (functional analyses)	Genome BC / Genome Canada / ROC / end users	Project manager, co-PIs, PDFs, collaborators,
Burke	Co-Project leader and lead of activity 1.2 (GWAS)	Genome BC / Genome Canada / ROC / end users	Project manager, co-PIs, PDFs, collaborators,
Staton	Project manager	Rieseberg & Burke	N/A
Donovan	Co-lead, activity 1 (stress resistance). Lead, sub-activities 1.1 (large-scale phenotyping), 1.3 (ecophysiology)	Rieseberg & Burke	Collaborators, PDFs, Techs, GRAs
Langlade	Co-lead, activity 1 (stress resistance). Lead, sub-activities 1.4 (Heliaphen phenotyping), and 1.5 (transcriptomics)	Rieseberg & Burke	GSC, INRA staff, PDFs, GRAs
Yeaman	Lead, activity 2 (population genomics)	Rieseberg & Burke	GSC, Bioinformatician, Bioinformatics PDFs
Hulke	Lead, activity 3 (MAGIC population)	Rieseberg & Burke	Collaborators, GSC, PDFs, GRAs
Ramenkutty	Lead, activity 5 (GE^3LS – crop yield models)	Rieseberg & Burke	PDF
Marden	Lead, activity 6 (GE^3LS – treaties & germplasm)	Rieseberg & Burke	RA

B. Governance and Critical decision-making

Overall Project governance will occur through the Research Oversight Committee (ROC) and Genome BC who will review any change in scientific, technology or funding realignment. Rieseberg, as overall project head, will be responsible for critical scientific decisions in each of the activity areas. These decisions will be informed by input from several sources, including (1) co-leader Burke, (2) the activity leads who will have quarterly teleconference meetings with Rieseberg and Staton; (3) the ROC and Genome BC; and (4) our end users and collaborators, who have helped us develop the MAGIC population design and contributed germplasm.

Decisions will be communicated to the whole team by the project manager and discussed at the quarterly management team meetings.

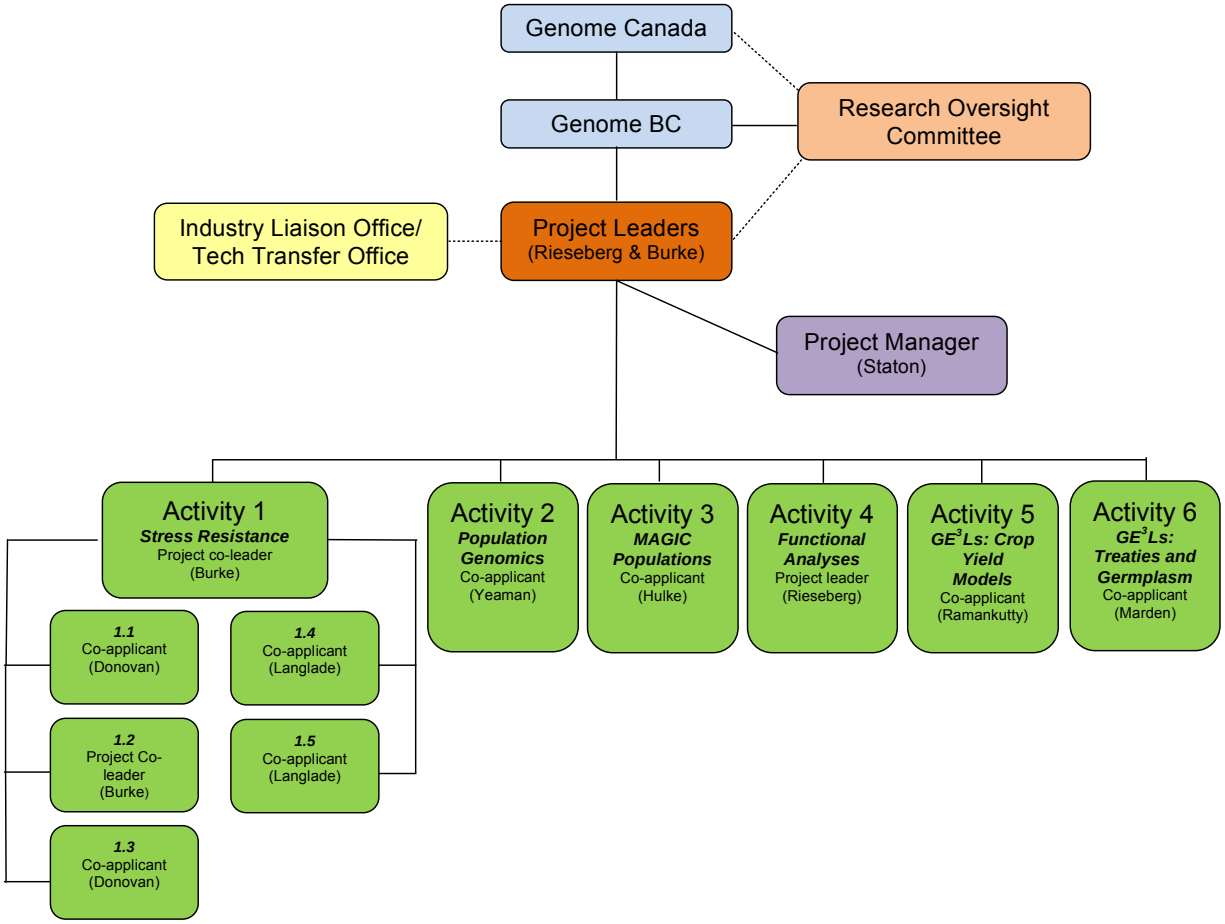

Figure 1. Organization Chart

C. Financial management and control

The financial management and controls are described in detail in section X (FINANCIAL INFORMATION)**.** Project Leader Rieseberg faculty and Project Manager Staton will have signing authority for purchases. The Botany Department at UBC will administer the project and keep records of all transactions. Staton will review the budget statements produced by UBC on a monthly basis. Rieseberg and Staton will review the amount spent relative to the budget allocated and adjust spending accordingly. Quarterly financial reports will be provided to Genome BC.

D. Deployment plan

Deployment is relatively straightforward. There are no large items of equipment to order or install as the equipment intensive parts of the project are to be contracted out, for instance the Illumina sequencing is to be done at the BC Genome Sciences Centre (GSC). Research personnel hires are the main deployment challenges. However, the major co-funding projects will be underway for several months before the start date of the present project, so personnel for these projects will likely be in place by Q1. At UBC, the project manager (Louisa Staton), GE^{3}LS Experts (Marden and Ramankutty), and one bioinformatics postdoctoral fellow (PDF Evan Staton) are in place, conditional offers have been made for two additional PDFs (Gina Conte & Dylan Burge), and one graduate student research assistant (GRA) has been identified (Ada Roman). Nonetheless, we have staggered the starting dates of UBC and U. Calgary personnel in accordance with project needs, which will also ease the burden of recruiting and hiring, with the

field phenotyping crew for the work at Indian Head to be hired in Q2, with the GE^{3}LS personnel to be hired in Q3, and the PDFs for the population genomics work to be hired in Q5. Additional recruiting (if needed) will begin as soon as we receive word of funding and will be done with the assistance of the experienced Faculty of Science Human Resources team and comparable team at U. Calgary, as well as through inquiries with colleagues. We have a ramp down phase as well, with several technical staff positions ending twelve and nine months prior to the end of project. However, we do not have a ramp down for positions related to computational analyses, because the need for personnel skilled in bioinformatics, data analyses, and manuscript preparation will be greatest towards the end of the project.

E. Partnerships

Mechanisms for the formation and coordination of partnerships are briefly described below:

(1) We will continue to hold our sunflower genomics workshop in San Diego each January after the Plant and Animal Genome (PAG) Conference. The workshop includes team members, our public and private partners, and end-users. As such, it provides an ideal setting for updating partners and end-users with progress on our project, as well as hands-on training in using data and tools developed by the project.
(2) We will provide continuous updates through our public and private websites (see below).
(3) Announcements/updates will also be made by regular emails from the desk of the project leader (Rieseberg) to the entire project via an email list coordinated by the project manager (Staton).
(4) The proposed project would likely be highlighted at the next International Sunflower Conference—a gathering of the sunflower producing community, which offers an opportunity to discuss priorities with end users and connect with new partners. The next Conference will be in 2016.

We have ongoing and overlapping collaborations with essentially all of our partners through our sunflower genomics resources consortium (INRA, U. Georgia, Dow, KWS, Advanta, Biogemma, Pioneer, Syngenta, and BASF), MAGIC population development consortium (INRA, USDA, Advanta, KWS, Dow, Nuseed, and Biogemma), data analytics project (SAP), and a pre-breeding/agrobiodiversity project funded with the Global Crop Diversity Trust, NASSARI, and SOLTIS. The proposed project will strengthen these collaborations and potentially lead to their merger when appropriate.

F. Accessibility of research results

As described in more detail in Appendix VI (Data and Resource Sharing Plans), we will employ several different strategies to make our results accessible to the research community. The germplasm resources produced under this award (i.e., the multi-species MAGIC population) will be deposited in the USDA sunflower germplasm collection at the North Central Regional Plant Introduction Station in Ames, IA for ongoing maintenance and distribution.

Dissemination of genomics, physiological, and evolutionary knowledge, as well as data analytic tools, will occur via the central data mining and analysis resource we are planning for the SAP Cloud, as well the Sunflower Genome Database (http://www.sunflowergenome.org), which also serve as home to the proposed project. In addition to being made available via the project website, information will be freely shared via GenBank, Phytozome at the Joint Genome Institute (http://www.phytozome.net/), the Gene Expression Omnibus (http://ncbi.nlm.nih.gov/geo/), and Dryad (http://datadryad.org/).

Project findings, including those from the GE^{3}LS research, will also be disseminated through publication in high-profile journals and presentations at conferences, universities, and companies. Data will be released without restrictions (and prior to publication) after it has passed our quality controls.

G. Arrangements with technology service providers

We have arrangements with two service providers: (1) the Genome Sciences Center (GSC), which is a Genome Canada GIN node; and (2) the USDA-ARS, Northern Crop Science Laboratory.

All of our sequencing and expression analyses (RNASeq) will be conducted by the GSC, which is Genome BC's genomics platform. The GSC operates 14 Illumina HiSeq 2500 instruments, with each instrument capable of generating one TB of sequence data per run. Library construction for Illumina sequencing is available for whole genome and transcriptome (ssRNA-seq) analyses. The GSC has constructed over 50,000 Illumina libraries to date, including many libraries for the members of the project team.

Our field-based phenotyping of flooding stress will be conducted by the USDA-ARS, Northern Crop Science Laboratory. The USDA-ARS laboratory has access to land without tile drainage systems and with access to irrigation to selectively flood a portion of a field. They also have extensive practical experience evaluating responses to flooding stress, as well as the necessary equipment and expertise for deploying sensor-based, field high-throughput phenotyping platforms.

H. Experience in managing large-scale projects

Project leader Rieseberg has previously directed the US NSF-funded Compositae Genome Project (CGP) and the Genome Canada/BC-funded Genomics of Sunflower projects, which were similar in scale and complexity to the present proposal (see Appendix I CVs) and had a translational component. These projects have provided perspectives on what works well and what works less well. Likewise, co-PIs Langlade and Burke have successfully directed large genomics projects funded by government and industry and have collaborated with Rieseberg to form two public/private consortia (see above) that leverage genetic and genomic data to facilitate breeding.

I. Highly qualified personnel

The HQP needs for this project include GRAs, PDFs, and technicians with expertise in a wide range of topics including ecophysiology, bioinformatics, population genetic theory, and high-throughput phenotyping. However, we do not anticipate challenges in recruiting the necessary personnel, as there is high demand to get into our labs and we receive numerous inquiries monthly. In addition, we have already identified several highly talented HQP, who will be available in Q1. Nonetheless, if we do run into difficulties in finding HQP with the required skill set, we will first email colleagues to inquire about the availability of HQP from their groups. If this approach fails, we will work with the Human Resources teams at UBC and U. Calgary to recruit more widely.

X FINANCIAL INFORMATION

This section must address all relevant evaluation criteria for the competition and must include:

- a description of the financial and budgetary controls (e.g., processes for authorizing purchases, payments and budget adjustments), and
- a justification for the main budget items including a summary of principal financial assumptions or explanations. If applicable, please include justification and assumptions for the calculation of a general consumable rate per FTE, for consumables commonly utilized in most laboratories. Please refer to budget line number (ref.#) when providing additional explanations. A narrative description of all budget lines is not required.

Budget

- Provide a budget request for **up to four years** using the budget template provided in Excel 2007 Macro-Enabled (*.xlsm) format.
- It is expected that applicants will work with Genome Centre staff to ensure that the budget meets all requirements outlined in the *Guidelines for Funding Research Projects*.
- Please ensure that the research activities are consistent between the research proposal, budget and Gantt chart.
- The budget and supporting documents (e.g., supplier quotes, statements of work (SOWs) from service providers) must be included in Appendix II.

A. Control Processes

Management of Research Funds:

As this project is undertaken within the University of British Columbia there are already in place established processes and controls for the handling of grant and project funds.

Project Leader (faculty) and staff will have the authority to incur expenses on behalf of the Institution. Both the spending member of faculty or staff and, if required, the senior individual approving payment are responsible for ensuring that claims for expenses are in accordance with the institution's procedures and are for project purposes only. Original signatures are required and all original vouchers and invoices are kept for audit purposes.

Research accounts are monitored at four levels:

1. Laboratory/Project Managers are responsible for reconciling all purchases with expenses appearing on the ledgers;
2. Project Leaders (PL) are responsible for monitoring the ledgers to ensure all expenses fall within guidelines established by the funding agencies, and that the purchases comply with the Institution's purchasing procedures and guidelines;
3. Project Managers monitor the research ledgers on a monthly basis and advise the researchers of potential spending concerns;
4. Financial Services ensures that all expenses comply with the granting agency's guidelines.

UBC has in place a number of strict policies and procedures developed to ensure compliance with a variety of funding agency requirements including Tri-Council (SSHRC, NSERC & CIHR) and National Institutes of Health (NIH). In particular UBC has policies governing administration of research funds encompassing purchases, contractors, travel and staff appointments. As well, UBC's Supply Management department oversees procurement and "…whose role is to maximize opportunities, manage risk and enable users to access goods and services at best value…." All project expenditures must be individually approved by investigators in whose name funds are awarded.

Expenditures for the proposed project will be subject to ongoing scrutiny and audit by UBC's internal and external auditors as well as any applicable funding agencies.

Accountability and Reporting:

The PL in conjunction with UBC will submit to Genome BC on a quarterly basis all information and data as prescribed by the Centre in terms of timing, format and content, which will allow for the on-going assessment and monitoring of project performance. It is also the responsibility of UBC to ensure that all the project activity leader(s) participate in this process. In addition, all co-funding expenditures will be reported to Genome BC and Genome Canada on a quarterly basis.

Budget Adjustments:

Any changes to the scientific, managerial or financial conditions of funding initially approved by Genome Canada will require recommendation from Genome BC and the ROC prior to submission to Genome Canada for consideration. All adjustments will be managed according to the principles outlined in Genome Canada's "Guidelines for Management of Changes to Genome Canada Projects", and others as applicable.

Final Reporting:

Within three (3) months of the completion of the project, the PL will submit to Genome BC a final report that includes a description of the accomplishments of the project relative to the approved objectives as well as a detailed financial report in the format prescribed by Genome Canada.

B. Budget Justification

We are requesting $3,215,247 in funds from Genome Canada and $1,752,158 from Genome BC (as co-funding) for this 4-year project. These funds will support all activities of the project.

The largest funding component of the project will support salaries and benefits for the project team. As this project requires a heavy phenotyping and bioinformatics work load, we have allocated funds accordingly.

We have two major fee-for-service arrangements. The first is with the Genome Sciences Centre who will provide the required library and sequencing services for activities 1, 2 and 3 (see Appendix II, document 1, budget lines 22, 23, 26, 29 and 55-61). The second is with the USDA-ARS, Northern Crop Science Laboratory in Fargo, North Dakota to support phenotyping of flooding stress for activities 1 and 3. They are currently the most suitable facility in North America to provide large-scale field phenotyping of flooding tolerance. The USDA-ARS lab has access to land without tile drainage systems and with access to irrigation to selectively flood a portion of a field. They have extensive experience in flooding tolerance evaluation in sunflower and have the necessary equipment to implement high-throughput phenotyping approaches. There are no such facilities in Canada available at the time of this application. Several Canadian government sources currently partner with this facility to support their phenotyping and breeding needs including: (1) Saskatchewan Ministry of Agriculture ($30,000 USD) to support line development; (2) Agriculture and Agri-Food Canada ($170,000 CAD) to conduct phenotyping; and (3) National Sunflower Association of Canada ($10,000 CAD) support for phenotyping in Manitoba. The Fargo USDA-ARS lab is home to the only public sector sunflower line development programme for Canada, and Agriculture and Agri-Food Canada and the Saskatchewan Ministry of Agriculture are cooperating with USDA-ARS to produce inbred lines and experimental hybrid varieties of sunflower for testing in Saskatchewan and Manitoba. Thus, coordination with this programme will be essential for realising the benefits of Genome Canada's investment in abiotic stress resistance in sunflower. The service they will provide (see Appendix II, document 3, budget lines 81-84) is vital to the success of this project and will ensure that we can deliver outputs for the end users.

The project will require funds for two computer servers to process the data (see Appendix II, document 5; budget lines 101,102). One server will be set up at UBC and the other at U Calgary.

We are requesting $1500 to support a survey of other plant genomics project to identify issues vis-à-vis treaty and genomics data sharing. This service will be provided by Peter Phillips at the University of Saskatchewan.

Financial Assumptions:

1. Currency conversions: Our co-funding sources are international, thus currency exchange rates can be difficult to predict. We have conservatively used 1USD=1CAD for all of the US co-funding. For our co-funding from INRA, we have used 1EUR=1.37CAD.

2. Salary and Benefit rates at international institutions: The University of Georgia had direct cost benefits that exceed the 20% limit outlined in the Genome Canada Funding Guidelines. The benefit rates at UGA include: FICA, Retirement, Life Insurance and Health Insurance (see Appendix II, document 4).

3. Lab consumable rate per FTE: We have assigned $12,000 per year for each activity 4 FTE (budget line 114) to support lab supplies. This will support, DNA and RNA extractions, PCR, making libraries; genome editing, enzyme digests and ligations and other tasks associated with sunflower transformations.

XI CO-FUNDING STRATEGY

Maximum of three (3) pages including tables. Refer to ***Section 7.2*** of the ***Guidelines for Funding Research Projects*** for details on co-funding requirements.

Please provide a well-developed and feasible plan which demonstrates the project's potential to secure at least 75% of the co-funding prior to the release of Genome Canada funds. All co-funding must directly support the objectives of the project. Co-funding must also be for eligible costs specifically requested in the Genome Canada budget in order to be considered as an eligible co-funding source.

In the format below, please provide details of the co-funding sources. For each source, include the organization name, amount that directly supports the objectives of the Genome Canada proposal, contribution type (e.g., cash or in-kind), expected receipt date, status of co-funding and a description of how the funds will directly support the objectives of the project. Documentation supporting secured or proposed co-funding must be included in ***Appendix III (Supporting Documentation for Co-Funding)***.

A. Funding Sources

#	Name of organization	Amount	Type	Expected Receipt Date	Status
1	Genome BC	$1,752K	Cash	October 1, 2015	Committed
Genome BC funds will be used to support the large scale phenotypic screens, ecophysiological characterization of abiotic stress tolerance, transcriptome sequencing, population genomic analyses, development and characterization of MAGIC populations, functional analyses of candidate genes, GEL3S research, and project administration (budget items Ref 97-101, and 103-123).					

#	Name of organization	Amount	Type	Expected Receipt Date	Status
2	National Science Foundation (NSF) University of Georgia	$2,482K	Restricted cash	July 1, 2015	Awaiting response
Co-Project Leader John Burke, has applied to NSF for US$3.9M, $2.4M of which is eligible for co-funding. NSF will fund the large-scale greenhouse and phenotypic screens at the University of Georgia (UGA) and DREC, ecopophysiological characterization at UGA, functional analyses of candidate genes at UGA, and will contribute to the transcriptome sequencing and analyses, collections of wild germplasm and soil analyses, genotyping the wild species for the population genomic analyses, and genotyping of the MAGIC populations. These activities will cover all four years of the project and directly support activities 1-4 (budget items Ref 1-41, 124-128). The NSF project is expected to begin July 1, 2015, one full quarter in advance of the Genome Canada start date. We have prorated the funds to include the July-Aug 2015 funds. Exchange rate used 1USD=1CAD.					

#	Name of organization	Amount	Type	Expected Receipt Date	Status

3	Global Crop Diversity Trust	$100K	In-kind	October 1, 2015	Committed

Description of how the funds will directly support the objectives of the project:
The Global Crop Diversity Trust within the context of their program on crop wild relatives 'Adapting Agriculture to Climate Change: Collecting, Protecting and Preparing Crop Wild Relatives", will fund all of the phenotyping and evaluation activities in Uganda (supporting Activity 3) including salaries, equipment, supplies and travel. These funds will cover all four years of the project (budget items Ref 48-53). Exchange rate used 1USD=1CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
4	Institut national de la recherche agronomique (INRA) Toulouse	$140K	In-kind	October 1, 2015	Committed

Description of how the funds will directly support the objectives of the project:
Co-applicant Nicolas Langlade has secured in-kind funding (EUR102,652) through INRA for this project. INRA will be responsible for the Heliophen phenotyping of selected genotypes from the SAM and MAGIC populations, and will contribute to the inference of gene regulatory networks. They will provide in-kind support for salaries and consumables. These activities will cover all four years of the project supporting activity 1 (budget items Ref 129-132). Funding has been secured and all items are eligible. Exchange rate used 1EUR=1.37CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
5	KWS Seeds	$50K	In kind	October 1, 2015	Committed

KWS Seeds has committed in-kind US$50,000 KWS Seeds will receive two pairs of IM1F1 lines. They will make fifty (50) crosses within each pair, and extract and send DNA from five plants from each of the fifty (50) crosses to UBC for genotyping. They will then self one hundred (100) lines from each of the two pairs for a minimum of six generations and conduct seed increases as needed for phenotyping. This directly supports activity 3 and will cover the first three years of the project (budget items Ref 43). Exchange rate used 1USD=1CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
6	Advanta Semillas SAIC	$27K	In kind	October 1, 2015	Committed

Advanta Semillas SAIC will contribute in-kind US$27,000 over 3 years. They will receive two pairs of IM1F1 lines. They will make fifty (50) crosses within each pair, and extract and send DNA from five plants from each of the fifty (50) crosses to UBC for genotyping. They will then self one hundred (100) lines from each of the two pairs for a minimum of six generations and conduct seed increases as needed for phenotyping. This activity will cover the first three years of the project. Advanta Semillas SAIC funding will be used to cover the costs of all salary and benefits for the technician in charge of the crossing and DNA extractions in support of activity 3 (budget items Ref 44). Exchange rate used 1USD=1CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
7	Biogemma	$50K	In kind	October 1, 2015	Committed

Biogemma has committed in-kind US$50,000 over 3 years. They will receive two pairs of IM1F1 lines. They will make fifty (50) crosses within each pair, and extract and send DNA from five plants from each of the fifty (50) crosses to UBC for genotyping. They will then self one hundred (100) lines from each of the two pairs for a minimum of six generations and conduct seed increases as needed for phenotyping. This activity will cover the first three years of the project and directly support activity 3. Biogemma funding will be used to directly support activity 3 (budget items Ref 45). Exchange rate used 1USD=1CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
8	Nuseed Americas	$35K	In kind	October 1, 2015	Awaiting response

Nuseed Americas has committed US$35,000 over 3 years. They will receive two pairs of IM1F1 lines. They will make fifty (50) crosses within each pair, and extract and send DNA from five plants from each of the fifty (50) crosses to UBC for genotyping. They will then self one hundred (100) lines from each of the two pairs for a minimum of six generations and conduct seed increases as needed for phenotyping. This activity will cover the first three years of the project in direct support of activity 3. Nuseed Americas funding will be used to directly support activity 3 (budget items Ref 46). Exchange rate used 1USD=1CAD.

#	Name of organization	Amount	Type	Expected Receipt Date	Status
9	SAP AG	$187K	In kind	October 1, 2015	Committed

SAP AG will provide in-kind a 128GB Cloud Instance for 1 year of the project, US$187,644 (budget item Ref 62) in support of activity 1. Exchange rate used 1USD=1CAD.