Overview The ability of organisms to adapt to new and changing conditions is vital in today's world of climate and other anthropogenic changes. Understanding and predicting this ability to adapt for species is thus key, and is especially relevant in crop species that are needed to sustain food and fuel sources for the world's growing populations. This research project aims to examine the genetic architecture of important traits for adaptation in maize before and after its domestication bottleneck. Two major questions will be addressed: 1) what is the distribution of effect sizes for mutations underlying important phenotypic traits in maize and teosinte, and 2) how do different histories of demography and selection change this genetic architecture over time and impact the species' ability to adapt? Vast resources of genomic and phenotypic data in both teosinte and maize create an ideal system in which to investigate these questions. I propose three objectives to answer these questions using both real and simulated data to create and test theoretical predictions in the system. The first objective is to estimate the distribution of fitness effects of new mutations in teosinte, maize's progenitor. From this result, simulations in objective two will be parameterized and run to simulate the demographic and selective history of maize during its domestication. This creates an expectation for how the genetic architecture of traits in maize, hypothesized to be either under selection during the domestication event or not, should change given its past and can be compared to real data to see if predictions match reality. Objective three then performs further simulations of additional demographic and selective events that varying landraces of maize have experienced to assess if further changes to genetic architecture are detected. This research will be conducted at the University of California, Davis, under the supervision of Dr. Jeffrey Ross-Ibarra and cosupervised by Dr. Kevin Thornton at the University of California, Irvine.

Intellectual Merit The proposed research will greatly expand our understanding of the adaptability of quantitative traits and the relationship this has with the architecture underlying such quantitative traits. Little is known about the genetic architecture of quantitative traits, yet most important traits are quantitative. With the availability of genomic and phenotypic data, this is an area now ripe for further research A standing question in evolutionary biology is whether adaptation generally happens from few loci of large effect or many loci of small effect. This is a difficult question to answer, as the detectability of small effect loci is limited. Traditionally, QTL and GWAS approaches can be used to find important genes underlying traits of interest, but suffer from the aforementioned problem of small effect size loci. This research proposes to resolve this issue and provide deeper insights into the genetic architecture of adaptation. Additionally, conducting this research in the maize/teosinte system informs future research on this important crop and breeding of the species for adapting to future climate change as well as advances our understanding of plant biology and plant genomics. The results could inform crop studies in terms of maintaining genetic diversity for the future in ways that may have much larger long term impacts on the maintenance of diversity as well as on adaptation to changing environments.

Broader Impacts The impacts of this research will span across both the scientific community and the broader public. The research components and results will be made publicly available through online repositories of simulation and analysis code as well as public archiving of the data used. Additionally, I will make an effort to publish results from this work either in open-access journals or on public pre-print servers for other journals. I have a strong record of conference attendance that I will maintain in order to present results to the scientific community, and will also use my presence on Twitter and as a contributor to The Molecular Ecologist online blog to more frequently reach both the scientific and public audiences. Furthermore, I will mentor undergraduate students in the lab and also be able to participate in the educational partnership that has been created between UC Davis and Pioneer High School in Woodland, California. This will allow me to visit high school classrooms to teach about the cutting edge of genomics and evolutionary biology research to young scientists as well as to potentially mentor high school students in the lab or on small-scale projects.

found this here: http://www.dailydemocrat.com/general-news/20140307/uc-davis-partnership-with-pioneer-high-benefits-budding-scientists

didnt know about this. would have to get buy-in from brady/sinha. possible if you're interested, but if students want molecular bnechwork experience doesn't

workshop on maize quant gen would be cool. have collaborators at Langebio in Irapuato where you could present. also could design a youtube video on it? another possibility – i teach undergraduate genetics, and am in process of flipping classroom. you could help design modules (blog posts, videos, etc.) on pop and quant gen for that class, which would potentially could used for all the genetics students at UCD

.

Project Description

A. Introduction

A critical goal of evolutionary biology today is to understand organisms' abilities to adapt to new or changing conditions. This is especially relevant in the face of global climate change and other increasingly common anthropogenic changes to the environment (Easterling, 2000). Many important phenotypic traits are quantitative, and in order to understand the potential for adaptation of these traits and maintenance of variation underlying them, we must understand their genetic architecture and how this architecture may be changed as a result of evolutionary processes such as selection and demography. Such knowledge advances our understanding of the process of adaptation and can further benefit crucial goals such as improving crop yields under changing climates.

Many ecologically or economically important traits have a complex, quantitative genetic basis, and much heterogeneity has been found among these traits in terms of the number of loci contributing to variation, their effect sizes, and their frequencies both within and among species (Orr, 2001; Slate, 2005). Knowledge of this genetic architecture of such traits is important for understanding how easily or quickly local adaptation may occur, or how long and difficult that process may be (Orr, 2005; Yeaman, 2015; Yeaman and Whitlock, 2011). Such knowledge can therefore greatly contribute to improve breeding and conservation efforts as well as to predicting responses to environmental changes.

Previous studies using genome-wide association (GWAS) and QTL approaches are limited by inability to detect small effect alleles. QTL studies are not able to achieve the resolution necessary for distinguishing the number and effect size of contributing loci, while GWAS studies fail to explain all phenotypic variance, questionably missing either rare large effect alleles or common small effect alleles (Gibson, 2012; Thornton et al., 2013). This has implications for studies attempting to identify loci important in adaptation which struggle when each locus out of many may only contribute a small amount to the trait (Rockman, 2011; Mackay et al., 2009). Thus, studies that aim to identify loci important in adaptation will benefit from knowledge on the range of mutation effect sizes they may expect to see when performing GWAS or genotype-environment associations.

The genetic architecture of a trait is determined by a number of factors, including population size, mutation, demographic history, and the history of both purifying and positive selection. Population bottlenecks, for example, can lead to the loss of variation or maintenance of deleterious variation (e.g. Renaut and Rieseberg, 2015; Gunther and Schmid, 2010). Strong selection, such as that imposed during crop domestication, can fix large-effect loci (Brown et al., 2011). And recently both theoretical and empirical studies in humans have shown that the types of mutations (Thornton et al., 2013) and the interaction of demographic and selective histories (Fu et al., 2014; Gravel et al., 2011; Henn et al., 2015b) have likely changed the genetic architecture of human phenotypes.

The current study aims to investigate how these factors shape the architecture of quantitative traits in maize. Maize and teosinte are extensively studied and serve as ideal models for this research. Maize (*Zea mays*, ssp. *mays*) was domesticated from its wild teosinte ancestor, *Zea mays*, ssp. *parviglumis*, approximately 9,000 years ago in southwestern Mexico (Matsuoka et al., 2002; Piperno et al., 2009). Thanks to its economic importance and history as a model genetic organism, extensive genomic and phenotypic data exist, making maize an ideal system to in which to study processes affecting genetic architecture. Previous work, for example, has estimated the demographic impacts (Wright et al., 2005) and identified loci under selection (Hufford et al., 2012) during maize domestication. A number of studies have used GWAS to study the architecture of phenotypes of interest in both maize (Wallace et al., 2014) and teosinte (Weber et al., 2009), allowing us to match simulated and theoretical results with empirical estimates for both taxa.

I propose three objectives which make use of the extensive knowledge and data available in the maize/teosinte system in order to answer the question of **how demography and selection change**

the genetic architecture of quantitative traits. Objective one builds a simulation model to match the genetic architecture of phenotypically important traits in teosinte. Objective two then simulates the demographic history of teosinte's domestication into modern-day maize and compares the expectations created from simulation to the genetic architecture of modern maize. Objective three further investigates the role of demography and selection on changing genetic architecture by simulating the histories of multiple landraces of maize that spread across the Americas post-domestication. This research will further our understanding of the impact of demography and selection impacts on phenotypic traits, improving our ability to predict the effects of selection and changing environments as well as to exploit genetic diversity in crops for continued breeding.

B. Research Objectives, Methods & Significance

Objective I: Model the genetic architecture of phenotypes in teosinte

This first objective aims to build a model to simulate the genetic architecture of phenotypes in teosinte (*Zea mays* ssp. *parviglumis*). Such a model will then enable us to study the impacts of demography and selection during domestication (Objective II) and range expansion (Objective IV) on the genetic basis of phenotypic traits.

In order to model the architecture of quantitative traits, we first need to understand the effects of mutation. The distribution of fitness effects (DFE) describes the consequences of mutations in terms of their impacts on an organism's fitness. Taking advantage of published (Chia et al., 2012) and new whole-genome sequencing data in teosinte (see Data section below), we will estimate the DFE in teosinte with the software DoFE (Keightley and Eyre-Walker, 2007; Stoletzki and Eyre-Walker, 2011), using estimates of the number of nonsynonymous and synonymous substitutions in genes. The resulting estimates of the fitness effects of new mutations can then be parameterized in terms of their effect on a fitness-related quantitative trait (Keightley and Hill, 1988; Eyre-Walker, 2010) such as yield, plant height, or flowering time.

The estimated DFE, combined with prior information on mutation (**?**) and recombination rates (**?**) will then be used as input in simulation models of quantitative trait evolution. Additional parameters necessary for our model (including dominance and the correlation between a trait and fitness) will be estimated using using an Approximate Bayesian computation approach (**?**): simulation parameters will be drawn from a prior distribution, and results compared to observed data to estimate the parameters' posterior distributions. We will take advantage of the library fwdpy (a Python implementation of fwdpp Thornton 2014 available at) to write simulation code that explicitly models a quantitative trait evolving under a model of stabilizing selection. From each simulation, we will then perform genome-wide-association analysis in order to compare with observed data from published () and on-going (see Data below) analysis of 16 phenotypic traits in a natural population of teosinte. The ABC approach should allow robust estimation of the necessary parameters from a set of several million such simulations.

Obtaining the DFE for teosinte provides a distribution of effect sizes for loci across the genome, information that is not widely available in many systems. Actual estimates of this distribution contribute to future studies aiming to realistically simulate any distribution of mutation effects by further describing the variability of this distribution across species, and such information adds to an important bodies of work such as understanding the genetic basis of complex diseases, among others (**?**).

kevin: other params?

cite github

with the url? `https://github.com/molpopgen/fwdpy`

cite Weber

do we need more detail on fwdpy? size of region etc? you had this and i commented out.

happy to leave out for space-saving, NSF doesn't need this level of detail, right?

Objective II: Model quantitative genetics of maize domestication

The model developed in Objective I will enable simulation of quantitative traits under stabilizing selection. Here we will modify this model in to incorporate both demographic change and direc-

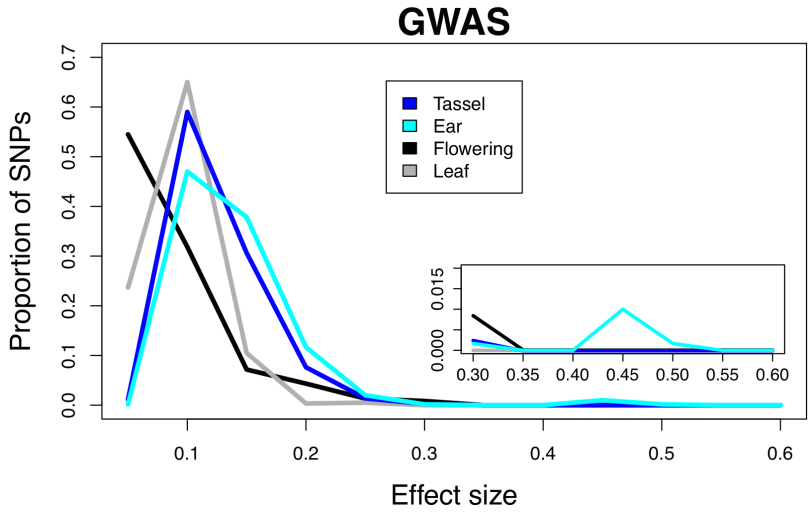

Figure 1: Effect sizes from GWAS analyses in maize, grouped by trait category. Inset shows the largest effects. Ear traits, showing the largest effect sizes, were likely selected during maize domestication while the others were not. Figure from Brown et al. 2011.

tional selection, and use it to investigate the impact of domestication on the genetic architecture of phenotypes of interest. Both demographic change associated with a domestication bottleneck and directional selection are expected to change the architecture of quantitative traits, but current theory includes only relatively simple models, and a detailed understanding of how these two forces interact is lacking.

First, wewill extend the simulation framework developed in Objective I to include estimates of population size change during the domestication bottleneck and subsequent expansion .Using parameters estimated in Objective I, we will simulate traits in a population undergoing a domestication bottleneck and perform GWAS on the simulated data. This first analysis provides a null expectation of the genetic architecture of traits in the absence of directional selection. We will compare these results to GWAS in both published and our own on-going analyses of maize data (see Data below).

Differences between simulations and observed data are informative of the history of selection on phenotypes, and we will again use fwdpy and our ABC approach to estimate the strength of selection on domestication phenotypes. I can assess how traits of varying heritabilities have changed over the course of domestication. More heritable traits or traits under stronger positive selection may change in different ways than those not under directional selection passing through the bottleneck and I can test if allelic effect sizes have shifted in the distribution's mean or overall shape to more or fewer loci of large effect. I will simulate the same quantitative traits used in Objective I. Traits in maize such as kernel weight and kernel row number are expected to have been selected for improving the species as a crop, while there are no strong reasons to suspect other traits such as total plant height were under selection. This suite of traits thus provides us with *a priori* predictions on which traits should show differences from our simple demographic simulations.

The results of this objective will show the relative importance of demography and selection in determining maize's genetic architecture. Demographic bottlenecks are common during the geographic spread of populations (cite) and can have several effects on the genome including purging of deleterious alleles (recessive alleles become homozygous and are more efficiently removed), or alternatively may lead to an increase of some deleterious alleles through increased random genetic drift (allele surfing, Klopfstein et al. 2005). The effects of these processes varies depending on the degree of population size reduction and the length of time over which populations are bottlenecked (cite bottleneck lit), but also have the potential to interact with the strength of selection during this demographic process. Changes in the distribution of allele effect sizes are not well understood in any system and are controversial in humans (Lohmueller, 2014; Simons et al., 2014; Hancock

i've been writing 'we' but feel free to change to 'i'; shoudl be consistent throughout though

cite wright et al. 2005

do we need to explain extent of bneck in maize at 5%??

cite wallace 2014

i don't think this has ever been done except backwards from archaeological data!

be more specific than 'may change in different ways' if possible. can cite chevin hospital 2008 PMID:18832353 for effect of directional selection on different QTL allele sizes

maybe add that differences for traits not thought to be under selection might be explained by genetic correlations with other traits? we could test this by estimating G matrix in maize.

et al., 2011). Deleterious alleles likely play a large role in many adaptive phenotypes: crop plants have undergone dramatic demographic shifts, usually involving a domestication bottleneck followed by expansion as cultivation spread, and some authors even argue that selection on domestication traits has inadvertently increased the frequency of alleles deleterious for other phenotypes (Gunther and Schmid, 2010). Consistent with this, it has recently been shown that genes associated with a number of quantitative traits in maize are enriched for deleterious alleles compared to randomly chosen genes (Mezmouk and Ross-Ibarra, 2014). Such information is crucial for understanding variation in phenotype, designing breeding strategies, utilizing diversity from wild relatives, or even engineering new traits using biotechnology.

this paragraph is great. not sure if it goes here or in project significance?

Objective III: Validate model predictions in synthetic mapping populations

Furthermore, a validation of the distributions of fitness effects in maize and teosinte can be performed using data from a synthetic cross of maize and teosinte. This *Zea* synthetic results from a mix of 26 maize lines with 12% teosinte genes (from 11 different founders, fully sequenced), 40% B73 (the reference genome line), and 2% from 25 other inbred maize lines. Using this genomic data to again estimate a distribution of effect sizes for the same traits, but now across regions of the genome known to originate from either maize or teosinte, I can test if the respective distributions for maize and teosinte are recovered in the same genetic background. This will identify if our simulations were truly valid.

expand this a bit, cite Sherry's NIL population too

Objective IV: Investigate the impact of various demographic and local adaptation of maize landraces maize across the Americas

do we want to include the SeeDs phenotype/genotype data?

worth a figure? am short on space already

Since domestication, various landraces of maize have spread across Central and South America and adapted into different lowland and highland habitats.These populations have experienced further demographic and selective pressures in addition to the initial domestication bottleneck. South American populations are inferred to have experienced a second severe bottleneck (cite), populations expanding geographically are likely to have experienced serial founder effects that can change allele frequencies in unexpected ways due to allele surfing (Klopfstein et al., 2005), and gene flow between teosinte and maize populations may have continued in some cases. Selection pressures in different lowland and highland habitats may also interact with these demographic events.

The objective of this project is to simulate the impact of multiple demographic events and selective pressures either individually or in combination. This will show how much variation is possible in terms of the effects on genetic architecture underlying traits for adaptation. The core set of simulations in this project will be designed to match the known landraces in maize, but will include more broadly a designed experiment to compare the presence or absence of particular evolutionary forces. Though there is currently not sufficient genomic and phenotypic data for all landrace populations of maize, it is likely that this will exist in the future, leaving an excellent future comparison to be made in terms of how well these simulation results match real populations across the species range. The goal of this project is to inform our understanding of the relative importance or insignificance of the tested demographic cases and how sensitive genetic architecture is to these. There is a debate currently in the field of evolutionary biology and human demographic history about the impact of events such as range expansions on the genome and the frequency of deleterious alleles (Henn et al., 2015a; Sudmant et al., 2015), necessitating further understanding of how such factors might interact under various selective environments. Furthermore, deleterious alleles likely play a large role in many adaptive phenotypes: crop plants have undergone dramatic demographic shifts, usually involving a domestication bottleneck followed by expansion as cultivation spread, and some authors even argue that selection on domestication traits has inadvertently increased the frequency of alleles deleterious for other phenotypes (Gunther and Schmid, 2010). Consistent with this, it has

recently been shown that genes associated with a number of quantitative traits in maize are enriched for deleterious alleles compared to randomly chosen genes (Mezmouk and Ross-Ibarra, 2014). Such information is crucial for understanding variation in phenotype, designing breeding strategies, utilizing diversity from wild relatives, or even engineering new traits using biotechnology.

Similar to objective two, I will simulate regions of the genome within individuals known to underly important phenotypic traits. These individuals will occupy populations that will be subjected to combinations of demographic and selective pressures including some of the following. To recapitulate the various landraces of maize, I will include cases of a second bottleneck after the domestication bottleneck, populations undergoing little or significant additional range expansion (Central American versus South American), stronger or weaker selection on flowering time and phenological traits (warmer lowland versus colder highland adapted populations), and cases with or without gene flow from sympatricpopulations of teosinte. I will do additional simulations to cover a wider range of parameters covering these cases, i.e. more extremes of a longer or larger range expansion, stronger and weaker bottlenecks, and higher or lower levels of gene flow among populations. This will allow assessment of how important the details of demography are in determining the genetic architecture of local adaptation to different conditions.

is that accurate, there is teosinte around them at least?

feel like I kind of repeated myself at the end here, maybe move the methods in earlier?

Data

This project will make use of both published data as well as data generated or currently being generated as part of a related Plant Genome project (Biology of Rare Alleles IOS-1238014) of which sponsoring scientist Dr. Ross-Ibarra is a Co-PI. As part of this project, Dr. Ross-Ibarra's group is currently in the process of sequencing 70 teosinte and 55 maize genomes to high depth. These genome sequences should be finished by the end of 2015 and will be made publicly available early 2016 via the group website (`www.panzea.org`). The individuals used for genome sequencing are also the parents of two large mapping populations of ≈ 5000 progeny. Both populations have been genotyped and phenotyped for a number of traits including seed yield, flowering times, and plant height. These data are currently available via collaborators. Publication of results from analyses of these data must await publication of collaborators' GWAS analyses, but but as these currently underway and comparisons of simulations in Objectives I and II would not begin until early 2017 at the soonest we do not foresee this being an issue.

jri todo: add mention of zea synthetic

C. Training Objectives

This fellowship will provide me with an ideal opportunity to learn the skills needed to enhance my ability to conduct cutting edge research in the fields of genomics and computational biology, both areas in which I expect to continue my future research and which are greatly expanding in evolutionary biology. I will gain many skills related to genomic data analysis through these projects, learning bioinformatics and analysis skills for large datasets. I have limited experience working with genomic data from my dissertation, thus making this a vital step in my career. Genomic technology and data are growing at an incredibly fast pace, and working directly with such data will teach me the most up to date, accurate, and efficient approaches. I will also improve my computational biology skills through the proposed simulations, learning a new and useful programming language, Python, that I can apply throughout this research and my future research in evolutionary biology.

D. Career Development & Future Research

My career goal is to develop an innovative research program in evolutionary biology, studying population genetics and the processes that impact genetic diversity. I believe such research is key

for the future, not only for the field of evolutionary biology, but also in applied scenarios such as understanding prevalence of genetic diseases in humans, adaptation of species to climate change, or strategies for improving agricultural products for a growing world population. My dissertation research has approached some of these questions in a more theoretical and less applied mindset. The work I will conduct during this fellowship would have more direct potential for application in the field of maize agriculture. For me, this is necessary and vital experience for my career development as I decide between pursuing a more applied research program, potentially in industry or government research scientist positions, or in pursing a career as an academic researcher at a university.

The skills I will develop during this fellowship, as described in section C, will benefit my career and put me on the cutting edge for analyses of the newest genomic data and the most recent computational approaches for biological simulations. Interacting with Dr. Ross-Ibarra, as well as other researchers at UC Davis, and with Dr. Kevin Thornton at UC Irvine, will be both intellectually stimulating and rewarding experiences that will help me accomplish my career goals. Drs. Ross-Ibarra and Thornton are both at the forefront of a popular movement for open science, making all stages of the research process transparent to any interested parties, and providing products such as data and code immediately and publicly. This is a work ethic I strongly agree with and hope to contribute to as an independent researcher. Our work together will better equip me with the tools and experience that make open science easy, efficient, and profitablefor all. I believe that this will equip me as a competitive, knowledgeable, and independent researcher able to conduct interesting and useful research throughout my future research program on topics of local adaptation, demographic history, population structure and genetic architecture of important traits. Furthermore, Dr. Ross-Ibarra has an excellent track record of helping his post-doctoral fellows secure promising positions for their future careers, including 3 assistant professorships at universities, 2 research scientist positions in the seed industry, one at an NGO, and one in the USDA.

unfortunate word choice. rewarding?

E. Sponsoring Scientists and Host Institution

The University of California Davis (UCD) is the ideal place to conduct the proposed research. UCD has a world-renowned program in evolutionary biology and faculty in population genetics who are at the top of the field. Jeff Ross-Ibarra is an expert on teosinte, maize, its domestication, and the associated population genetics and genomics of the system. Kevin Thornton is an accomplished quantitative geneticist and computational biologist at UC Irvine, who will also contribute greatly to this research. They will both serve as effective and capable mentors for my post-doctoral research. In particular, Jeff has been studying the maize/teosinte system for many years with a great network of collaborators providing vast resources of data. His work has contributed largely to our knowledge of this system, and more generally on domestication and adaptation as evolutionary processes. Kevin is also the developer and maintainer of fwdpy, the python package proposed for completing the simulations. He will thus serve as a great resource in terms of knowing the exact capabilities of the simulation method and any assumptions of its model that must be taken into account. Furthermore, the Department of Ecology and Evolution, the Department of Plant Biology, and the Department of Plant Sciences at UCD have many exceptional faculty doing research relevant to my interests, providing many research groups to interact with on a daily basis for potential collaborations or feedback on this research. For example, I look forward to interacting with scientists interested in population genetics, such as Graham Coop, and in adaptation, such as Johanna Schmitt. UCD has the necessary computing resources for our proposed work, and as described, vast sources of knowledge and experience on the topics I plan to investigate, ensuring the success of this work. I am excited to join and contribute to UCD's active and vibrant scientific community.

Jeff, feel free to make that sound better

Likewise Kevin, feel free to modify

should probably add something about Farm and Kevin's cluster

F. Milestones & Timeline

Year 1: Estimate DFE in teosinte; design model & perform ABC (Obj. I)
Year 2: Model domestication & selection, validate with empirical data (Obj. II, III)
Year 3: Model local adaptation, population expansion (Objective IV).

G. Broader Impacts

The proposed research will have wide-ranging impacts for both the public and the scientific community. I will ensure that my results are available to the public at all stages of these projects by maintaining code and scripts online at my GitHub account, which will allow other researchers to access analysis methods or data cleaning tools as well as simulation details and parameters which can provide a building block from which further research can be conducted. I will present new findings at international conferences and submit publications to open-access pre-print servers. I will also be able to broadcast my work more widely to the public through a strong online presence I maintain on Twitter and The Molecular Ecologist, a blog I have contributed to in the past. The impacts of this research on corn as a crop species useful for both food and fuel resources is also of an innate broad impact, as understanding the genetics underlying adaptation will ensure a viable future for such crop species in the future. Lastly, this research will contribute greatly to my own career development, improving my knowledge on genomics and working in an economically important crop species. I will be able to learn these vital tools as well as to teach them to undergraduate students in the lab and into the future as the field of genomics continues to grow.

big thing missing here is outreach etc. blogs/twitter a good start, but perhaps offer a workshop, produce other online content? here they are looking for how your work will impact others. examples in the summary doc are a good start too!

genetic architecture underlying traits affects how easy/hard and quick/slow local adaptation can occur. important for breeding, conservation, predicting response to climate change, etc. useful info for crops/domesticated species (and also things like disease in humans?)

i think pointing out this is important for humans too is useful in the proj. significance section

References Cited

P. J. Brown, N. Upadyayula, G. S. Mahone, F. Tian, P. J. Bradbury, S. Myles, J. B. Holland, S. Flint-Garcia, M. D. McMullen, E. S. Buckler, and T. R. Rocheford. Distinct genetic architectures for male and female inflorescence traits of maize. *PLoS Genetics*, 7:e1002382, Nov. 2011.

J.-M. Chia, C. Song, P. J. Bradbury, D. Costich, N. de Leon, R. J. E. John Doebley, B. Gaut, L. Geller, J. C. Glaubitz, M. Gore, K. E. Guill, J. Holland, M. B. Hufford, J. Lai, M. Li, X. Liu, Y. Lu, R. McCombie, R. Nelson, J. Poland, B. M. Prasanna, T. R. Tanja PyhŁjŁrvi, R. S. Sekhon, Q. Sun, M. I. Tenaillon, F. Tian, J. Wang, X. Xu, Z. Zhang, S. M. Kaeppler, J. Ross-Ibarra, M. D. McMullen, E. S. Buckler, G. Zhang, Y. Xu, and D. Ware. Maize HapMap2 identifies extant variation from a genome in flux. *Nature Genetics*, 44:803–807, June 2012.

D. R. Easterling. Climate Extremes: Observations, Modeling, and Impacts. *Science*, 289(5487): 2068–2074, Sept. 2000.

A. Eyre-Walker. Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *PNAS*, 107:1751–1756, 2010.

W. Fu, R. M. Gittelman, M. J. Bamshad, and J. M. Akey. Characteristics of Neutral and Deleterious Protein-Coding Variation among Individuals and Populations. *The American Journal of Human Genetics*, 95(4):421–436, Oct. 2014.

G. Gibson. Rare and common variants: twenty arguments. *Nature Reviews Genetics*, 13:135–145, Feb. 2012.

K. J. Gilbert and M. C. Whitlock. Evaluating methods for estimating local effective population size with and without migration. *Evolution*, 69(8):2154–2166, July 2015.

S. Gravel, B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, F. Yu, R. A. Gibbs, The 1000 Genomes Project, and C. D. Bustamante. Demographic history and rare allele sharing among human populations. *PNAS*, 108(29):11983–11988, July 2011.

T. Gunther and K. J. Schmid. Deleterious amino acid polymorphisms in Arabidopsis thaliana and rice. *Theor. Appl. Genet.*, 121:157–168, Mar. 2010.

A. M. Hancock, D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin, R. Sukernik, G. Utermann, J. K. Pritchard, G. Coop, and A. Di Rienzo. Adaptations to Climate-Mediated Selective Pressures in Humans. *PLoS Genetics*, 7(4):e1001375, Apr. 2011.

B. M. Henn, L. R. Botigue, C. D. Bustamante, A. G. Clark, and S. Gravel. Estimating the mutation load in human genomes. *Nature Reviews Genetics*, 16(6):333–343, May 2015a.

B. M. Henn, L. R. Botigue, S. Peischl, I. Dupanloup, M. Lipatov, B. K. Maples, A. R. Martin, S. Musharoff, H. Cann, M. Snyder, L. Excoffier, J. Kidd, and C. D. Bustamante. Distance from Sub-Saharan Africa Predicts Mutational Load in Diverse Human Genomes. Technical report, May 2015b.

M. B. Hufford, P. Bilinski, T. Pyhäjärvi, and J. Ross-Ibarra. Teosinte as a model system for population and ecological genomics. *Trends in Genetics*, 28(12):606–615, Dec. 2012.

P. D. Keightley and A. Eyre-Walker. Joint Inference of the Distribution of Fitness Effects of Deleterious Mutations and Population Demography Based on Nucleotide Polymorphism Frequencies. *Genetics*, 177(4):2251–2261, Dec. 2007.

P. D. Keightley and W. G. Hill. Quantitative genetic variability maintained by mutation-stabilizing selection balance in finite populations. *Genetical Research*, 52:33–43, 1988.

S. Klopfstein, M. Currat, and L. Excoffier. The Fate of Mutations Surfing on the Wave of a Range Expansion. *Molecular Biology and Evolution*, 23(3):482–490, Dec. 2005.

K. E. Lohmueller. The Impact of Population Demography and Selection on the Genetic Architecture of Complex Traits. *PLoS Genetics*, 10(5):e1004379, May 2014.

T. F. C. Mackay, E. A. Stone, and J. F. Ayroles. The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, 10:565–577, Aug. 2009.

Y. Matsuoka, Y. Vigouroux, M. M. Goodman, J. Sanchez G, E. Buckler, and J. Doebley. A single domestication for maize shown by multilocus microsatellite genotyping. *PNAS*, 99:6080–6084, Apr. 2002.

S. Mezmouk and J. Ross-Ibarra. The Pattern and Distribution of Deleterious Mutations in Maize. *G3*, 4:163–171, Jan. 2014.

H. A. Orr. The genetics of species differences. *Trends in Ecology and Evolution*, 16:343–350, Oct. 2001.

H. A. Orr. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2):119–127, Feb. 2005.

D. R. Piperno, A. J. Ranere, I. Holst, J. Iriarte, and R. Dickau. Starch grain and phytolith evidence for early ninth millennium B.P. maize from the Central Balsas River Valley, Mexico. *PNAS*, 106: 5019–5024, Jan. 2009.

S. Renaut and L. H. Rieseberg. The Accumulation of Deleterious Mutations as a Consequence of Domestication and Improvement in Sunflowers and Other Compositae Crops. *Molecular Biology and Evolution*, 32(9):2273–2283, Aug. 2015.

M. V. Rockman. The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution*, 66(1):1–17, Nov. 2011.

Y. B. Simons, M. C. Turchin, J. K. Pritchard, and G. Sella. The deleterious mutation load is insensitive to recent population history. *Nature Genetics*, 46(3):220–224, Feb. 2014.

J. Slate. Quantitative trait locus mapping in natural populations: progress, caveats, and future directions. *Molecular Ecology*, 14:363–379, Oct. 2005.

N. Stoletzki and A. Eyre-Walker. Estimation of the neutrality index. *Molecular Biology and Evolution*, 28(1):63–70, Jan. 2011.

P. H. Sudmant, S. Mallick, B. J. Nelson, F. Hormozdiari, N. Krumm, J. Huddleston, B. P. Coe, C. Baker, S. Nordenfelt, M. Bamshad, L. B. Jorde, O. L. Posukh, H. Sahakyan, W. S. Watkins, L. Yepiskoposyan, M. S. Abdullah, C. M. Bravi, C. Capelli, T. Hervig, J. T. S. Wee, C. Tyler-Smith, G. van Driem, I. G. Romero, A. R. Jha, S. Karachanak-Yankova, D. Toncheva, D. Comas, B. Henn, T. Kivisild, A. Ruiz-Linares, A. Sajantila, E. Metspalu, J. Parik, R. Villems, E. B. Starikovskaya, G. Ayodo, C. M. Beall, A. D. Rienzo, M. F. Hammer, R. Khusainova, E. Khusnutdinova, W. Klitz, C. Winkler, D. Labuda, M. Metspalu, S. A. Tishkoff, S. Dryomov, R. Sukernik, N. Patterson, D. Reich, and E. E. Eichler. Global diversity, population stratification, and selection of human copy-number variation. *Science*, 349(6523), Sept. 2015.

K. R. Thornton. A C++ Template Library for Efficient Forward-Time Population Genetic Simulation of Large Populations. *Genetics*, 198:157–166, Sept. 2014.

K. R. Thornton, A. J. Foran, and A. D. Long. Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. *PLoS Genetics*, 9:e1003258, Feb. 2013.

J. G. Wallace, P. J. Bradbury, N. Zhang, Y. Gibon, M. Stitt, and E. S. Buckler. Association mapping across numerous traits reveals patterns of functional variation in maize. *PLoS Genetics*, 10: e1004845, Dec. 2014.

A. L. Weber, Q. Zhao, M. D. McMullen, and J. F. Doebley. Using association mapping in teosinte to investigate the function of maize selection-candidate genes. *PLoS One*, 4:e8227, Dec. 2009.

S. I. Wright, I. V. Bi, S. G. Schroeder, M. Yamasaki, J. F. Doebley, M. D. McMullen, and B. S. Gaut. The effects of artificial selection on the maize genome. *Science*, 308:1310–1314, May 2005.

S. Yeaman. Local adaptation by alleles of small effect. *The American Naturalist*, 186:S74–S89, Oct. 2015.

S. Yeaman and M. C. Whitlock. The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7):1897–1911, Mar. 2011.

Biographical Sketch — Kimberly Julie Gilbert

maybe still want to change around which pubs go in 5 most important?

A. Professional Preparation

Institution, Location	Major	Degree	Year
University of Virginia, USA	Biology	B.Sc.	2010
University of British Columbia, Canada	Zoology	Ph.D.	2016 (expected)

B. Publications

Five Publications Most Closely Related to the Proposed Project

1. **Gilbert KJ**, MC Whitlock (2015) Evaluating methods for estimating local effective population size with and without migration. *Evolution*, 68(8), 2154-2166.
2. **Gilbert KJ**, MC Whitlock (2015) Q_{ST}-F_{ST} comparisons with unbalanced half-sib designs. *Molecular Ecology Resources*, 15(2), 262-267.
3. Caplins SA, **KJ Gilbert**, C Ciotir, J Roland, SF Matter, N Keyghobadi (2014) Landscape structure and the genetic effects of a population collapse. *Proceedings of the Royal Society B*. 281: 20141798; doi: 10.1098/rspb.2014.1798
4. Keller SR, **KJ Gilbert**, PD Fields, DR Taylor (2012) Bayesian inference of a complex invasion history revealed by nuclear and chloroplast genetic diversity in the colonizing plant, *Silene latifolia*. *Molecular Ecology*, 21(19), 4721-4734.
5. Whitlock MC, **KJ Gilbert** (2012) Q_{ST} in a hierarchically structured population. *Molecular Ecology Resources*, 12(3), 481–483.

Four Other Significant Publications

1. Santiso X, L Lopez, **KJ Gilbert**, R Barreiro, MC Whitlock, R Retuerto (2015) Patterns of genetic variation within and among populations in *Arbutus unedo* and its relation with selection and evolvability. *Perspectives in Plant Ecology, Evolution and Systematics*, 17(3), 185-192.
2. Vines TH, RL Andrew, DG Bock, MT Franklin, **KJ Gilbert**, NC Kane, EJ Kleynhans, J-S Moore, BT Moyers, S Renaut, DJ Rennison, T Veen, S Yeaman (2013) Mandated archiving greatly improves access to research data. *FASEB Journal*, 27(4), 1304-1308.
3. **Gilbert KJ**, RL Andrew, DG Bock, MT Franklin, NC Kane, J-S Moore, BT Moyers, S Renaut, DJ Rennison, T Veen, TH Vines (2012) Recommendations for utilizing and reporting population genetic analyses: The reproducibility of genetic clustering using the program STRUCTURE. *Molecular Ecology*, 21(20), 4925-4930.
4. Vines TH, AYK Albert, RL Andrew, F Débarre, DG Bock, MT Franklin, **KJ Gilbert**, J-S Moore, S Renaut, DJ Rennison (2014) The availability of research data declines rapidly with age. *Current Biology*, 24, 94-97.

C. Select Conference Presentations (chosen from 11 presentations)

2015 Validating SNP loci underlying local adaptation in lodgepole pine; KJ Gilbert, S Yeaman, KE Lotterhos, KA Hodgins, H Suren, JA Holliday, S Nadeau, SN Aitken, MC Whitlock *Poster, 15th ESEB Congress, Lausanne, Switzerland*

2014 Evaluating methods for estimating effective population size in the presence of migration; KJ Gilbert & MC Whitlock *Oral presentation, Evolution, Raleigh, USA*

2012 Range expansion and adaptation across heterogeneous environments; KJ Gilbert & MC Whitlock *Poster, Evo-WIBO Conference (Evolutionary Biology in the Pacific Northwest), Port Townsend, USA*

2011 Inferred invasion history of *Silene latifolia* into North America utilizing population genetic data and approximate Bayesian computation; KJ Gilbert, SR Keller, PD Fields, DR Taylor *Poster, 13th Congress of the European Society for Evolutionary Biology, Tuebingen, Germany*

D. Grants and Awards

Cordula and Gunter Paetzold Fellowship *UBC* $18,000CAD 2015 – 2016

Declined; Zoology Graduate Fellowship *UBC* $11,000CAD	2015 – 2016
Ann and William Messenger Graduate Fellowship *UBC* $700CAD	2015
Zoology Graduate Fellowship *UBC* $11,000CAD	2014 – 2015
Frieda Granot Graduate Scholarship in Interdisciplinary Research *UBC* $200CAD	2013 – 2014
Theodore E Arnold Fellowship *UBC* $7,750CAD	2013 – 2014
Patrick David Campbell Graduate Fellowship *UBC* $8,050CAD	2013 – 2014
Declined; Zoology Graduate Fellowship *UBC* $10,000CAD	2013 – 2014
BRITE Fellowship *UBC* $10,500CAD *per annum*	2011 – 2013

E. Synergistic Activities

1. **Working Groups:** Participated in the NESCent Reproducible Science Hackathon (2014), a 21-member working group aimed at developing a curriculum and workflow for teaching reproducible science to researchers of any background. Participated in the SimBank NESCent Catalysis Meeting (2014) which was a 25-member working group to create a collection of openly available simulation results to facilitate testing of statistical population genetic and phylogeographic methods.
2. **Teaching:** Teaching assistant for Fundamentals of Evolutionary Biology (Fall 2012-Spring 2013) where I taught three sections per term of 45 students each and lead discussion-based tutorials. Teaching assistant for Fundamentals of Biostatistics (Fall 2013, 2014, 2015). Taught two sections of 70 students total in 2013, and in 2014 and 2015 served as the lab coordinator for 254 and 276 students enrolled in the course, respectively, while teaching one section of 36 and 35 students respectively.
3. **Service:** Served as a Graduate Student Council Member for the American Society of Naturalists (2013-2016, chair 2015-2016). Served as the graduate student representative on the 2014 evolutionary biology CRC2 job search for the Department of Zoology, University of British Columbia. Organize the Biodiversity Research Centre's weekly evolution discussion group (2014-2016), for students, post-docs, and faculty from the departments of Zoology, Botany, Forestry, and Fisheries to discuss current papers in evolutionary biology. Reviewer for *Molecular Ecology Resources, Ecology and Evolution, Tree Genetics & Genomes*.
4. **Outreach:** Volunteer mist-netting and bird banding with local Vancouver non-profit organization Wild Research (2013-2015) where I participated in winter, spring migration, and fall migration bird monitoring at Iona Island Bird Observatory, taught volunteers proper bird handling, aging, data collection, and mist net extraction techniques, and assisted in educating public visitors to the station about the species conservation and monitoring, and the general tasks of running a banding station.

F. Collaborators (Total: 25)

U Alberta Jens Roland; *U Basel* Peter Fields; *U British Columbia* Dan Bock, Diana Rennison; *U Calgary* Sam Yeaman; *UC Davis* Serena Caplins; *U Cincinnati* Stephen Matter; *CIRB Paris* Florence Débarre; *Colorado State* Brook Moyers; *U Colorado* Nolan Kane; *U Coruna* Rodolfo Barreiro, Lúa López; *Kwantlen Polytechnic U* Michelle Franklin; *U Laval* Jean-Sébastien Moore; *Mol. Ecol. Managing Editor* Timothy Vines; *U Montreal* Sébastien Renaut; *U New England (Australia)* Rose Andrew; *UT Austin* Thor Veen; *Trent U* Claudia Ciotir; *U Santiago de Compostela* Rubén Retuerto Franco, Xabier Santiso; *U Vermont* Stephen Keller; *U Virginia* Douglas Taylor; *Western U* Nusha Keyghobadi; *Women's Health Research Institute* Arianne Albert

Graduate Advisor (Total: 1) *University of British Columbia* Michael C. Whitlock

Biographical Sketch — Jeffrey Ross-Ibarra

1 Professional Preparation

Institution	Area	Degree / Training	Dates
University of California Riverside	Botany	BA, MS	1998, 2000
University of Georgia	Genetics	PhD	2006
University of California Irvine	Genetics	Postdoctoral Research	2008

2 Professional Appointments

Position	Institution	Dates
Associate Professor	University of California Davis	2012-present
Assistant Professor	University of California Davis	2009-2012
Profesor de Asignatura	Universidad Nacional Autónoma de México	2001

3 Products

Most Relevant to the Proposed Research

- Mezmouk S, **Ross-Ibarra J** (2014) The pattern and distribution of deleterious mutations in maize. (2014) G3 4:163-171
- Hufford MB, Xun X, van Heerwaarden J, Pyhäjärvi T, Chia J-M, Cartwright RA, Elshire RJ, Glaubitz JC, Guill KE, Kaeppler S, Lai J, Morrell PL, Shannon LM, Song C, Spinger NM, Swanson-Wagner RA, Tiffin P, Wang J, Zhang G, Doebley J, McMullen MD, Ware D, Buckler ES, Yang S, **Ross-Ibarra J** (2012) Comparative population genomics of maize domestication and improvement. NATURE GENETICS 44:808-811
- Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, **Ross-Ibarra J**, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the Nested Association Mapping and Inbred Association panels. PLANT PHYSIOLOGY 158: 824-834
- van Heerwaarden J, Doebley J, Briggs WH, Glaubitz JC, Goodman MM, Sánchez González JJ, **Ross-Ibarra J** (2011) Genetic signals of origin, spread and introgression in a large sample of maize landraces. PNAS 108: 1088-1092
- **Ross-Ibarra J**, Tenaillon M, Gaut BS (2009) Historical divergence and gene flow in the genus Zea. GENETICS 181: 1399-1413.

Additional Products

- Gerke JP, Edwards JW, Guill KE, **Ross-Ibarra J**, McMullen MD (2015) The genomic impacts of drift and selection for hybrid performance in maize. GENETICS *In Press*
- Takuno S, Ralph P, Swarts K, Elshire RJ, Glaubitz JC, Buckler ES, Hufford MB, and **Ross-Ibarra J** (2015) Independent molecular basis of convergent highland adaptation in maize. GENETICS 200:1297-1312
- Wills DM, Whipple C, Takuno S, Kursel LE, Shannon LM, **Ross-Ibarra J**, Doebley JF (2013) From many, one: genetic control of prolificacy during maize domestication. PLOS GENETICS 9(6): e1003604.
- Studer A, Zhao Q, **Ross-Ibarra J**, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. NATURE GENETICS 43:1160-1163.

- Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, **Ross-Ibarra J**, Ware DH, Buckler ES (2009) A first-generation haplotype map of maize. SCIENCE 326: 1115-1117.

4 Synergistic Activities

- Faculty Development Award in recognition of university service, 2015
- Editor, G3, PeerJ, Axios Reviews
- DuPont Young Professor 2012-2014 and faculty advisor DuPont Pioneer graduate student symposium in plant breeding 2012-present
- Functional Genetics of Maize Centromeres US-Mexico exchange program, 2011-present
- Presidential Early Career Award for Scientists and Engineers 2009

5 Collaborators and Other Affiliations

Collaborators and Co-editors (Total: 56)

Cornell U Peter Bradbury, Jeffrey Glaubitz, Susan McCouch, Qi Sun, Feng Tian, Sharon Mitchell; *USDA-ARS* Edward Buckler, Sarah Hake, James Holland, Sherry Flint-Garcia, Mike McMullen, Doreen Ware, Jode Edwards; *U Southern California* Peter Ralph; *UC Davis* Alan Bennet, Daniel Runcie, Ed Taylor, Graham Coop, Keith Bradnam, Ian Korf, David Neale, Amélie Gaudin; *UC Irvine* Kevin Thornton; *Carnegie Institute* Davide Sosso; *Stanford* Wolf Frommer; *LANGEBIO* Ruairidh Sawers; *U Georgia* Kelly Dawe; *Arizona State* Reed Cartwright; *U Missourri* James Birchler, Katherine Guill, David Wills; *Beijing Genomics Institute* Song Chi, Xun Xu; *U Wisconsin* John Doebley, Jiming Jiang, Shawn Kaeppler; *Syngenta* William Briggs; *Monsanto* Lisa Kanizay; *Dupont Pioneer* Andy Baumgarten, Justin Gerke, Oscar Smith, Tabare Abadie; *U Minnesota* Roman Briskine, Peter Morrell, Chad Myers, Nathan Springer, Peter Tiffin; *MIT* Mary Gehring; *NC State* Major Goodman; *INRA* Clementine Vitte, Maud Tenaillon; *Brigham Young* Clinton Whipple; *Danforth Center* Anthony Studer; *Universidad de Guadalajara* Jesus Sànchez Gonzàlez; *Iowa State* Carolyn Lawrence; *U Hawaii* Gernot Presting; *UC Riverside* Mitchell Provance

Graduate Advisors and Postdoctoral Sponsors (Total: 3)

UC Riverside Norman Ellstrand; *U Georgia* James Hamrick; *UC Irvine* Brandon Gaut

Thesis Advisor and Postgraduate Sponsor (Total: 14)

Postdoctoral: *Iowa State* Matthew Hufford; *Graduate U Advanced Studies* Shohei Takuno; *U Oulu* Tanja Pyhäjärvi, *KWS* Sofiane Mezmouk; *Wageningen* Joost van Heerwaarden; *USDA* Tim Beissinger; *UC Davis* Kate Crosby, Sayuri Tsukahara, Simon Renny-Byfield, Jinliang Yang **Graduate:** Dianne Velasco, Paul Bilinski, Anna O'Brien, Michelle Stitzer

Biographical Sketch — Kevin Richard Thornton

6 Professional Preparation

Institution	Area	Degree / Training	Dates
University of Puget Sound	Botany	BA	1997, 2000
University of Chicago	Genetics	PhD	2003
Cornell University	Genetics	Postdoctoral Research	2007

7 Professional Appointments

Position	Institution	Dates
Associate Professor	University of California Irvine	2012-present
Assistant Professor	University of California Irvine	2007-2012

8 Products

Most Relevant to the Proposed Research

- **Thornton**, K. R. (2014) A C++ template library for efficient forward-time population genetic simulation of large populations. Genetics 98:157-166 PMID: 24950894
- **Thornton, K**. (2003) libsequence, a C++ class library for evolutionary genetic analysis. Bioinformatics 19(17): 2325-2327 PMID 14630667
- **Thornton, K. R.**, A. J. Foran, and A. D. Long (2013) Properties and modeling of GWAS when complex disease risk is due to non-complementing, deleterious mutations in genes of large effect. PLoS Genetics 9: e1003258. PMID 23437004

Additional Products

- Cridland, J. M., **K. R. Thornton** and A. D. Long (2015) Gene expression variation in Drosophila melanogaster due to rare transposable element insertion alleles of large effect. Genetics 199: 85-93.
- Baldwin-Brown, J., A. D. Long, and **K. R. Thornton** (2014) The Power to Detect Quantitative Trait Loci Using Resequenced, Experimentally Evolved Populations of Diploid, Sexual Organisms. Molecular Biology and Evolution 31: 1040-1055. PMID 24441104
- Open-source software: `http://molpopgen.github.io/fwdpp/`
- Open-source software: `http://molpopgen.github.io/libsequence/`

9 Synergistic Activities

- Open-source software: `http://molpopgen.github.io/fwdpy/` This software is unpublished, and will be a key resource for this proposal.
- Editor, G3

10 Collaborators and Other Affiliations

Collaborators and Co-editors (Total: 7)

- *Cornell University* Andrew G. Clark
- *North Carolina State University* Trudy Mackay
- *Princeton University* Peter Andolfatto
- *Rochester University* Daniel Garrigan, Daven C. Presgraves
- *UC Irvine* Anthony (Tony) Long
- *University of Kansas* Stuart MacDonald

Graduate Advisors and Postdoctoral Sponsors (Total: 2)

University of Chicago Manyuan Long
Cornell University Andrew G. Clark

Thesis Advisor and Postgraduate Sponsor (Total: 3)

Postdoctoral: Rebekah R. Rogers **PhD Thesis Advisor:** Julie M. Cridland, Jaleal S. Sanjak

Data Management Plan

Data Types

This proposal will make use of existing maize and teosinte datasets from multiple sources as well as data currently being generated and only generate simulated datasets. These real-world maize and teosinte datasets are used within objectives one and two of the proposal, while simulated data is produced within objective two, but mainly in objective three.

Sequence, Genotype, and Phenotype Data part of a related Plant Genome project (Biology of Rare Alleles IOS-1238014) of which sponsoring scientist Dr. Ross-Ibarra is a Co-PI. As part of this project, Dr. Ross-Ibarras group is currently in the process of sequencing 70 teosinte and 55 maize genomes to high depth. These genome sequences should be finished by the end of 2015 and will be made publicly available early 2016 via the group website (www.panzea.org). The individuals used for genome sequencing are also the parents of two large mapping populations of 5000 progeny. Both populations have been genotyped and phenotyped for a number of traits including seed yield, flowering times, and plant height. These data are currently available via collaborators. Publication of results from analyses of these data must await publication of collaborators GWAS analyses, but but as these currently underway and comparisons of simulations in Objectives I and II would not begin until early 2017 at the soonest we do not foresee this being an issue. In the unlikely event of a problem in the generation of that data, existing data is also available publicly for ~1500 maize genomes through the HapMap 3 and 4 projectsand 4,000 landraces each with 1 million SNPs and several phenotypes is also publicly available through the SeeDs project (cite). The synthetic *Zea* line used for validation in objective II is available from

cite Panzea?

Cite accepted Plant Gen paper from Sherry

Simulated Data Outputs from simulations will consist of sequence data of genomic regions representing important phenotypic traits. This data will be organized to describe how users can interpret the data in terms of what phenotype each region represents and what file corresponds to each demographic and selective scenario simulated. All of this information will be contained in Readme files associated with each directory of simulated datasets.

General Organization

All obtained and generated datasets will be stored locally and remotely on UC Davis's online Google Drive. In the rare event of simultaneous failure of these storage facilities, maize and teosinte datasets can be re-obtained from original authors and simulated datasets can be re-generated from the same random seed to produce identical results. In addition to data, analysis and simulation scripts will be stored locally and remotely on my GitHub account. I will additionally use GitHub to version control all scripts so that previous versions are not deleted nor is confusion of the proper published version a problem. Using version control and GitHub are already protocols I follow currently and will provide no future difficulty in maintaining this workflow. Readme files can be generated for each set of analysis or simulation scripts that serve as a sort of metadata for reproducing the analyses within and instructions for future users to understand how each script is run.

better way to describe this?

Data Archiving, Plan for Sharing, Public Access Policy

The existing maize and teosinte datasets are all either currently publicly available or will be made publicly available before the completion of this research.

Simulated data will be publicly archived concurrently with publication of the research results. These datasets will be archived on Dryad Digital Repository, unless otherwise directed to another appropriate, publicly accessible data archive by the publication journal. Detailed metadata will be provided to ensure reproducibility of the studies. The scripts used to generate these data files and conduct analyses will be made available with detailed instructions on my GitHub account. I will post links to all data sources on my academic website, which is a standard I already maintain.

Since all data is either simulated or comes from existing plant datasets, there will be no privacy concerns for sharing data.

Dissertation Summary - Kimberly J. Gilbert

A major obstacle in evolutionary biology is the difficulty of population genetic inference in the face of confounding factors, such as demographic history. My dissertation work has focused on several topics related to this broad area of research:

1. Evaluating the ability of statistical genetic methods to estimate effective population sizes in the face of migration (Gilbert and Whitlock, 2015)
2. Assessing the factors related to local adaptation at range edges during species expansion
3. Validating SNP loci under selection for adaptation to climate in lodgepole pine (*Pinus contorta*)

Effective population size, N_e, is a fundamental parameter in population genetics, evolutionary biology, and conservation biology, yet its estimation can be fraught with difficulties. Several methods to estimate Ne from genetic data have been developed that take advantage of various approaches for inferring N_e. The ability of these methods to accurately estimate N_e, however, has not been comprehensively examined. This part of my dissertation work employed seven of the most cited methods for estimating N_e from genetic data (Colony2, CoNe, Estim, MLNe, ONeSAMP, TMVP, and NeEstimator including LDNe) across simulated datasets with populations experiencing migration or no migration. The simulated population demographies were an isolated population with no immigration, an island model metapopulation with a sink population receiving immigrants, and an isolation by distance stepping stone model of populations. We found considerable variance in performance of these methods, both within and across demographic scenarios, with some methods performing very poorly. The most accurate estimates of N_e can be obtained by using LDNe, MLNe, or TMVP; however each of these approaches is outperformed by another in a differing demographic scenario. Knowledge of the approximate demography of population as well as the availability of temporal data largely improves N_e estimates.

Species range edges have boundaries that cannot always be explained ecologically or geographically, which leaves the question of what evolutionary forces may prevent populations at range edges from adapting and expanding the species range further. A large body of theoretical work has investigated many evolutionary parameters' effects on local adaptation in edge populations, but one area lacking in research is that of the interaction of the landscape with the ability to locally adapt. This study investigates how more realistic, heterogeneous environmental gradients (compared to the linear gradients that previous studies investigate) may interact with dispersal distance and the effect size of mutations. I have simulated a range of parameter combinations that show a strong relation of mutation effect size on the ability to spread across the landscape. As environmental heterogeneity increases, migration load (reduction in fitness due to dispersal away from an area previously adapted to) increases, and local adaptation becomes more difficult, especially in smaller populations at the range edge, slowing the speed of expansion across the landscape.

A history of range expansion can confound many inferences that population genetics aims to understand. Identifying the loci that underlie traits contributing to local adaptation is one such inference that is a major goal in evolutionary biology today. The lodgepole pine (*Pinus contorta*) is a major timber tree in the Pacific Northwest which has a history of expansion post-glaciation, and either one or putatively a second glacial refugia from which this expansion occurre1.9d. Climate change is spurring foresters to plant trees for future harvest that will be best adapted to future climates for optimal yield, hence identifying loci underlying adaptation to climate change is a key goal. I am conducting a validation study of SNP loci identified through GWAS, genotype-environment association, and F_{ST} outlier tests to assess how often these methods may produce false positives as a result of population structure and spatial autocorrelation of genetic clines due to range expansion with gradients in environmental variables (i.e. temperature and precipitation). I have sampled a provenance trial (common garden study) in British Columbia to compare performance of populations from a range of native temperatures (MAT -3.7°C - 11°C) planted across test sites of varying temperature (MAT -1.4°C - 5°C) from which I will be able to test if predicted alleles do indeed show increased performance in mature, natural-grown trees.