Overview The ability of organisms to adapt to new and changing conditions is vital in the face 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 new mutations in genomic regions 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. Through four objectives I aim to answer these questions using both real and simulated data to create and test theoretical predictions in the system. I will first estimate the distribution of fitness effects of new mutations in teosinte, maize's progenitor, then use this information to parameterize and simulate the demographic and selective history of maize during its domestication. This creates expectations for how the genetic architecture of various traits, some of which are hypothesized to be under selection during domestication, may change and can be compared to real data to test if predictions match reality. These results can be validated using a synthetic *Zea* line what contains both teosinte and maize in the same background. Further, to investigate a broader range of demographic and selective pressures, I will simulate the history of various landraces of maize that expanded across the Americas to assess further impacts of these forces on genetic architecture of important traits. This research will be conducted at the University of California, Davis, under the supervision of Dr. Jeffrey Ross-Ibarra and co-supervised 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. 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. Traditional QTL and GWAS approaches suffer from a lack of resolution required for understanding genetic architecture. This research proposes to resolve this issue and provide deeper insights into the genetic architecture of adaptation. Maize/teosinte serve as a perfect system for this study as it has the advantage of a known demographic and selective history (domestication) and both genomic and phenotypic data for populations were (maize) and were not (teosinte) subject to these selective and demographic forces. Furthermore, this work can inform future research on this important crop and breeding of this and other species for adapting to future climate change as well as advancing 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 will be made publicly available at all stages through online repositories of simulation and analysis code as well by public archiving the data. Additionally, I will make published results available on public pre-print servers, or when possible in open-access 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, teaching skills on computational and genomics analyses and develop workshops on population and quantitative genetics and computational tools that I will organize to teach at the Langebio research center in Irapuato, Mexico. I will also have the opportunity to assist Dr. Ross-Ibarra in developing modules for his undergratuate genetics class as well as visit local high schools to introduce these ideas to students.

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 ecologically or economically important phenotypic traits have a complex, quantitative genetic basis, and considerable heterogeneity exists within and among species in this genetic architecture — the number of causal loci, their effect sizes, and allele frequencies (Orr, 2001; Slate, 2005). In order to understand the maintenance of variation or adaptive potential of these traits, we must understand how their genetic architecture changes 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.

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.

Current approaches for studying the genetic architecture of traits are limiting. Quantitative trail locus (QTL) mapping approaches cannot achieve the resolution required to separate causal loci. Genome-wide-association analyses, while improving resolution, suffer from multiple testing difficulties that may mean they fail to detect loci of small effect. Many studies in plants suffer from relatively small population sizes, and are thus unable to asses the potential importance of rare mutations of large effect (cf Thornton et al., 2013). Alleles missed in most QTL or GWAS studies may nonetheless be of central importance to adaption (Rockman, 2011; Mackay et al., 2009), and novel approaches are thus needed to understand how evolutionary processes affect the architecture of quantitative traits.

is this true? if it's a matter of rare things and small pops can anything fix this? not just an inability of the methods in general?

Thornton et al. don't get into sample size and power. A better citation would be doi:10.1371/journal.pgen.1004179

The current study aims to investigate how these factors shape the architecture of quantitative traits in maize. Thanks to its economic importance and history as a model genetic organism, extensive genomic (Wright et al., 2005; Hufford et al., 2012) and phenotypic (Wallace et al., 2014; Weber et al., 2009) data exist for both maize and its wild ancestor, making maize an ideal system to in which to study processes affecting genetic architecture.

I propose four 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 I. builds a simulation model to match the genetic architecture of phenotypically important traits in teosinte. Objective II. 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 III. validates this comparison using a synthetic line of maize and teosinte combined onto the same genetic background. Then Objective four 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

I. Model the genetic architecture of phenotypes in teosinte

Objective one 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 Management Plan), I 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 (Clark et al., 2005) and recombination rates (Rodgers-Melnick et al., 2014), 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 (Beaumont et al., 2002): 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 `https://github.com/molpopgen/fwdpy`) 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 (Weber et al., 2009) and on-going (see Data Management Plan) analysis of 16 phenotypic traits in a natural population of teosinte. The ABC approach will 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 these results add to important bodies of work such as understanding the genetic basis of complex diseases, among others (Eyre-Walker and Keightley, 2007).

II. Model quantitative genetics of maize domestication

The model developed in Objective I. will enable simulation of quantitative traits under stabilizing selection. Here I will modify this model to incorporate both demographic change and directional 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, I will extend the simulation framework developed in Objective I. to include estimates of population size change during the domestication bottleneck and subsequent expansion (Wright et al., 2005). Using parameters estimated above, I will simulate the same quantitative 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. I will compare these results to GWAS in both published (Wallace et al., 2014) and our own on-going analyses of maize data (see Data Management Plan).

Differences between simulations and observed data are informative for the history of selection on phenotypes, and I will again use fwdpy and an 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 more dramatically 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 (Chevin and Hospital, 2008). Traits in maize such as kernel weight and kernel row number, among other ear-related traits (Figure 1), are expected to have been selected for crop improvement, 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, and in cases where these predictions are not matched may indicate genetic correlations with selected traits that could be further investigated by study of the G matrix in maize.

wonder if this might be a case where it's too much jargon for reviewers without defining the G matrix? could just be vague and say "...could be further investigated."

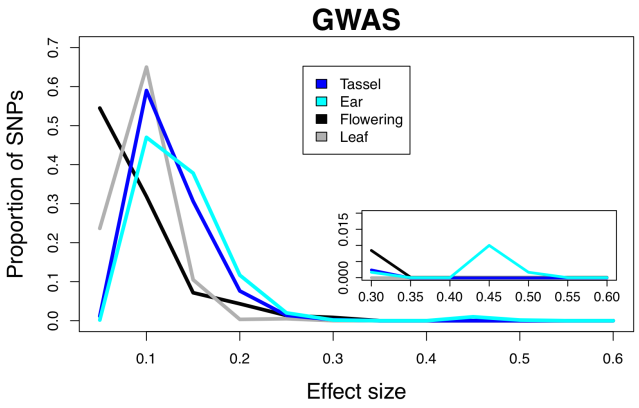

Figure 1: Effect sizes from GWAS analyses in maize, grouped by trait category. Inset shows the largest effects. Ear traits, showing the largest effect sizes, were likely selected during maize domestication while the others were not. Figure from Brown et al. 2011.

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 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 (e.g. Caplins et al., 2014), 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 (Hancock et al., 2011; Henn et al., 2015a,b; Lohmueller, 2014; Simons et al., 2014). 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.

III. Validate model predictions in synthetic mapping populations

Models developed in Objective I. and II. lead to direct predictions of the genetic architecture of quantitative traits and how they should differ between maize and teosinte. To validate the predictions of these models, I will take advantage of two mapping populations containing mixtures of teosinte and maize alleles: a set of near isogenic lines of teosinte alleles introgressed into a maize background (Liu et al., In Press), and a synthetic population resulting from the cross of 11 teosinte and 26 maize parents (see Data Management Plan). Genotype and phenotype data are available for both populations, and of necessary I will add additional genotyping to augment marker density. I will perform a GWAS on these samples for the traits already used to calibrate our models. Identification of GWAS hits in regions of the genome coming from maize and teosinte will allow us to compare the effect size distribution of causal variants in a common environment and genetic background. With information on the correlation between phenotype and fitness, our model and simulations should be able to predict the difference in genetic architecture for new traits. I will thus perform a similar analysis for traits not included in our calibration to test the predictive ability of our model.

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

After domestication, maize spread across Central and South America and adapted to a variety of environments across a wide elevational range. These populations have experienced further demographic and selective pressures in addition to the initial domestication bottleneck. For example, South American populations are inferred to have experienced a second severe bottleneck (Takuno et al., 2015), and more generally 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). The effects of range expansion on deleterious alleles remains controversial (Henn et al., 2015a; Sudmant et al., 2015), and even less is known about their impact on quantitative traits. Changing selection pressures from both different human populations and new environments likely also interact with these demographic events.

Following Objective II., I will simulate quantitative traits in populations undergoing further selection and demographic change. This will require new simulations to incorporate range expansion or other demographies as well as multiple shifts in selection. Though current data likely does not provide the resolution to estimate these parameters for all maize landrace populations, qualitative predictions about changes in genetic architecture can be compared to GWAS results from publicly available genotype and phenotype data of nearly 5000 landraces from across the Americas (Hearne et al., 2015). This will allow assessment of how complex demography and changing selection regimes affect the genetic architecture of local adaptation to different conditions.

what about ditching this section and just citing the data management section?

yes, all over this idea!

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 efficient, and rewarding for 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.

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,

Jeff, feel free to make that sound better

Likewise Kevin, feel free to modify

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.

should probably add something about Farm and Kevin's cluster

you can copy-paste what I wrote in the sponsoring doc

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

M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162:2025–2035, Dec. 2002.

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.

R. Bukowski, X. Guo, Y. Lu, C. Zou, B. He, Z. Rong, B. Wang, D. Xu, B. Yang, C. Xie, L. Fan, S. Gao, X. Xu, G. Zhang, Y. Li, Y. Jiao, J. Doebley, J. Ross-Ibarra, V. Buffalo, E. S. Buckler, Y. Xu, J. Lai, D. Ware, and Q. Sun. Construction of the third generation zea mays haplotype map. *bioRxiv*, 2015. doi: 10.1101/026963.

S. A. Caplins, K. J. Gilbert, C. Ciotir, J. Roland, S. F. Matter, and N. Keyghobadi. Genetic spatial structure in a butterfly metapopulation correlates better with past than present demographic structure. *Molecular Ecology*, 281(1796):20141798–20141798, Oct. 2014.

L. M. Chevin and F. Hospital. Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, 180(3):1645–1660, Nov. 2008.

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.

R. M. Clark, S. Tavare, and J. Doebley. Estimating a nucleotide substitution rate for maize from polymorphism at a major domestication locus. *Molecular Biology and Evolution*, 22:2304–2312, Aug. 2005.

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.

A. Eyre-Walker and P. D. Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618, Aug. 2007.

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.

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.

S. Hearne, C. Chen, E. Buckler, and S. Mitchell. Unimputed gbs derived snps for maize landrace accessions represented in the seed-maize gwas panel. `http://hdl.handle.net/11529/10034`, 2015. Accessed: 2015-02-16.

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, 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.

Z. Liu, J. Cook, S. Melia-Hancock, K. Guill, C. Bottoms, A. Garcia, O. Ott, R. Nelson, J. Recker, P. Balint-Kurti, S. Larsson, N. Lepak, E. Buckler, L. Trimble, W. Tracy, M. D. McMullen, and S. A. Flint-Garcia. Expanding maize genetic resources with pre-domestication alleles: maize- teosinte introgression populations. *Plant Genetics*, In Press.

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.

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.

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.

E. Rodgers-Melnick, P. J. Bradbury, R. J. Elshire, J. C. Glaubitz, C. B. Acharya, S. E. Mitchell, C. Li, Y. Li, and E. S. Buckler. Recombination in diverse maize is stable, predictable, and associated with genetic load. *PNAS*, 112:3823–3828, Feb. 2014.

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.

S. Takuno, P. Ralph, K. Swarts, R. J. Elshire, J. C. Glaubitz, E. S. Buckler, M. B. Hufford, and J. Ross-Ibarra. Independent Molecular Basis of Convergent Highland Adaptation in Maize. *Genetics*, 200:1297–1312, Aug. 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.

Biographical Sketch — Kimberly Julie Gilbert

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, New Phytologist, 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. Unless additional genotyping is needed in Objective III., this project will only generate simulated datasets. These real-world maize and teosinte datasets are used within objectives I., II., and III. of the proposal, while simulated data is produced within Objectives I.,II., and IV..

Sequence, Genotype, and Phenotype 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 sequences should be complete 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 $\sim$5000 progeny. Both populations have been genotyped and phenotyped for a number of traits including seed yield, flowering times, and plant height, and would be used for comparisons in Objectives I. and II.. Finally, the same project has developed, genotyped, and phenotyped a synthetic population of 11 teosinte and 26 maize parents, resulting in a population that is roughly 12% teosinte genes, 40% B73 (the reference genome line), and 2% from 25 other inbred maize lines. This latter population will be used in Objective III. to compare model predictions to maize and teosinte regions segregating in a common population. All of these data are currently available from collaborators. Publication of results from analyses of these data must await publication of collaborators' GWAS analyses, but as these are currently underway and comparisons with empirical data would likely 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 teosinte (Weber et al., 2009) and maize (Wallace et al., 2014) as well as $\sim$1500 maize genomes through the HapMap 3 project (Bukowski et al., 2015) and 4,000 landraces each with 1 million SNPs and several phenotypes (Hearne et al., 2015).

Simulated Data Outputs from simulations will consist of genotype data and allele effect sizes from 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. Additionally, links to the exact code used (hosted on GitHub) and random number seeds will be included in publications, so that the exact simulated data can be replicated in the future if necessary.

General Organization

All obtained and generated datasets will be stored locally and remotely backed up on the 300 terabyte array for the Ayala School of Biological Sciences. In the rare event of simultaneous failure of these storage facilities, maize and teosinte datasets can be re-obtained from original 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.

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 during 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.

Sponsoring Scientist Statement

I. Brief description of the research projects in the host research group(s), including a statement of current and pending research support

Kevin, please add

Research in the Ross-Ibarra lab focuses on the evolutionary genetics of maize and teosinte, including studies of local adaptation, genome evolution, and experimental evolution. The proposed project is related to the goals of the funded "Biology of Rare Alleles" grant. That grant seeks to understand the impacts of rare alleles and predict phenotypes from genotyping data. While this proposal will ideally utilize data produced by that grant, its focus on the mechanisms of evolutionary change and the proposed modeling framework are outside of the scope of the "Biology of Rare Alleles" grant.

is this credible enough, or do I need to say more?

The Thornton lab focuses on fundamental questions relating evolutionary processes to the interpretation of population- and quantitative- genetic data. The proposed project is related to the grant "Population genetics of polygenic adaptation". That grant seeks to describe how patterns of genetic variation are impacted by adaptation to a phenotypic "optimum shift" and to estimate the relevant model parameters. Although there is some overlap between some of the aims of the current proposal and my R01, the approach proposed here is different and complementary.

Current Support

Title: Adaptive gene flow from teosinte to highland maize in central Mexico **Agency:** UC MEXUS (Ross-Ibarra, Co-PI) **Amount:** $24,897 **Dates:** 07/15 12/16
kjgI think something is weird here unless it is meant to be formatted this way? Should the " " be replaces with an en-dash? **Calendar Person-months:** 0.24
Title: Functional genomics of maize centromeres **Agency:** NSF-PGRP (Ross-Ibarra, Co-PI) **Amount:** $754,409 **Dates:** 06/10 5/16**Calendar Person-months:** 2.4
Title: US Mexico planning visit and workshop to assess the genomic basis of local adaptation in maize **Agency:** NSF-CNIC (Ross-Ibarra, Co-PI) **Amount:** $36,450 **Dates:** 09/14 12/16 **Calendar Person-months:** 0.5
Title: Biology of rare alleles in maize and its wild relatives **Agency:** NSF-PGRP (Ross-Ibarra, Co-PI) **Amount:** $3,221,212 **Dates:** 05/15 4/18 **Calendar Person-months:** 2
Title: Population genomics of polygenic adaptation. **Agency:** NIH/NIGMS **Amount:** $1,000,000 (Thornton, PI) **Dates:** 09/2015 - 08/2019 **Calendar Person-months:** 3

Pending Support

Title: Evolutionary genomics of maize **Agency:** HHMI (Ross-Ibarra, PI) **Amount:** NA (decided by agency) **Dates:** 07/16 6/21**Calendar Person-months:** 1
Title: Research PGR: The genetics of highland adaptation in maize **Agency:** NSF PGRP (Ross-Ibarra, PI) **Amount:** $4,531,773 **Dates:** 05/16 4/21 **Calendar Person-months:** 2.4
Title: Research PGR: The evolutionary role of hybridization and introgression in the genus Zea **Agency:** NSF-DEB (Ross-Ibarra, Co-PI) **Amount:** $415,775 **Dates:** 07/16 6/19 **Calendar Person-months:** 0.6

II. Description of how the research and training plan for the applicant would fit into and complement ongoing research of the sponsor(s) as well as an indication of the personnel with whom the Fellow would work.

The proposed research nicely extends current work in the Ross-Ibarra lab on the impacts of demography and selection on allele frequencies to their effects on complex phenotypes. By investigating how genetic architecture may change, this proposal also shows potential to be extremely informative of work underway on complementation and the basis of hybrid vigor. Additionally, the work will complement existing research in the Thornton lab on polygenic adaptation. The Thornton lab is primarily concerned with the analysis of human and *Drosophila* data sets. The proposed research will add an important perspective from a major model plant system.

Kim will be encouraged to interact with other postdoctoral fellows and graduate students in both labs. These interactions provide opportunities to work outside the focus of the current proposal, to develop new collaborations, and to coauthor additional publications. We hope and expect that some of these interactions may evolve into formal collaborations on projects not involving either sponsor.

III. Explanation of how the sponsor(s) will determine what mentoring the applicant needs in research, teaching, and career development skills and how these would be translated into a specific plan that fosters the development of the applicant's future independent research career.

Research We have planned a rigorous guided reading discussion to ensure Kim gets quickly up to speed on the necessary quantitative genetics theory as well as the relevant literature on quantitative traits in maize, and have discussed with Kim online and formal coursework options for learning python programming. We will both meet with Kim weekly to ensure progress on the research; Jeff also has an open-door policy encouraging interaction with lab members outside of formal meeting times.

Writing We work with lab members to develop an initial outline of their papers, but lab members are expected to prepare a complete first draft of manuscripts and figures. This responsibility is reflected in the fact that postdocs and students are not only first but also co-corresponding author on their papers. We then work with lab members to revise and edit their manuscript in a process that often goes through numerous rounds of revision. Manuscripts are shared with other lab members and colleagues to encourage practice providing (and receiving) constructive criticism, and final versions of our papers are posted as preprints online upon submission to take advantage of community feedback that can be incorporated into the published paper. One clear goal will be first authorship on submitted papers, with the expectation of approximately one first author paper per year of duration of the postdoc. We will also encourage lab members to participate in other writing activities, including blog posts on our journal club papers, formal reviews for journals, and grant proposals. The Ross-Ibarra lab has a documented history of successful funding with postdoctoral scholars as Co-PIs, and these efforts provide valuable training (and even initial funding) for the scholars' future academic careers.

Teaching Postdoctoral fellows in the Ross-Ibarra Lab are encouraged to participate in formal teaching opportunities including development of lectures for a graduate course on ecological genomics or a large undergraduate genetics lecture course. Jeff works with the postdoctoral fellow to develop the lecture and provides feedback on teaching style and and content afterwards. Kim will also be encouraged to practice presenting her work to broad audiences both in joint lab meetings and formal seminars at UC Davis and UC Irvine, as well as at local (e.g. Bay Area Population Genomics) and international (SMBE, PAG) meetings. Both Jeff and Kevin provide guidance and feedback on such presentations.

Mentoring Another important aspect of training will be experience mentoring graduate students and undergraduates. Previous efforts to encourage such supervision in our labs have been very successful, with postdoc-mentored students presenting conference posters on their research or earning authorship on papers. Supervisory experience has proven helpful for postdocs applying for jobs, especially in industry.

Career Development During weekly meetings we will focus not only on current progress on the project but also on progress towards Kim's broader career goals. We will encourage Kim to take advantage of formal infrastructure for professional development in place at UC Davis, including training in responsible conduct of research, grantsmanship, mentoring, career development, authorship of journal papers, and teaching. We will encourage attendance at conferences to present results and build relationships with other leaders in the field. The Ross-Ibarra lab has a proven record of success in placing postdoctoral scholars in careers industry, government, nongovernmental research organizations, and academic positions, and we will continue to encourage postdocs to explore a range of career opportunities.

IV. Description of the role the sponsors will play in the proposed research and training and the other resources that will be available to the Fellow to complete his or her training plan during the fellowship.

Kevin will primarily help Kim to develop the necessary models and code, and Jeff will primarily help in analysis of simulation and empirical data and guidance in the interpretation of the results. However, Kim will meet weekly with both advisors to facilitate research interactions among the sponsors and ensure that everyone is in agreement on progress and continued plans. We will additionally formally address long terms goals and progress in separate meetings every 6 months. Though Kim will be based primarily in the Ross-Ibarra lab at UC Davis, we expect the development of the simulation framework will involve extended visits to the Thornton lab at UC Irvine.

UC Davis is regularly ranked as one of the top three centers of research in both plant and evolutionary biology. Kim will have access to a broad set of discussion groups and seminars, as well as to the experience and guidance of world-class set of scientists that can help her with all aspects of the development of her research, engagement in diversity issues, and career path.

Both labs have access to extensive computational resources. The Ross-Ibarra lab is a contributing partner in the college computer cluster, giving the lab dedicated access to 192 processors, with the opportunity for use of nearly 800 additional CPU as resources allow. Recent additions to the cluster have provided it with additional CPU as well as six new shared high-memory (512Gb RAM) nodes, one of which is dedicated to the Ross-Ibarra lab. Dr. Ross-Ibarra is a faculty member of the UC Davis Genome Center, a large facility that includes bioinformatics and genotyping cores as well as access to additional computational facilities. At UCI, the Thornton lab has 192 processors divided amongst 3 64-core machines. Campus-wide, the lab has access to an additional $\sim$3,000 processors in a mix of high- and low- priority queues. With respect to this project, we have effectively unlimited storage in the form of a 300 terabyte array for the Ayala School of Biological Sciences.

V. Description of the limitations, if any, be placed on the Fellow regarding the research following the fellowship.

Kim would be free to continue and expand on the proposed work either in maize or other systems and to continue development of fwdpy simulation code.