# Project Summary

**Overview** How does the genetic architecture of quantitative traits change as a result of demography and selection during the maize domestication bottleneck and the further bottleneck that some maize populations underwent from Central America to South America?

If we compare maize and teosinte (which are locally adapted in their respective pops), do we see evidence of:

- many genes of small effect or few genes of large effect underlying important traits?
- do these differences match our predictions based on their differing demographic histories?

Broadly relevant because these results could inform on maintaining diversity in crops for the future in ways that may have much larger long term impacts in the maintenance of diversity as well as fitness/adaptation in the face of climate change and future adaptation to changing environmental conditions.

**Intellectual Merit**

**Broader Impacts**

# Project Description

## A. Introduction

A critical goal of evolutionary biology today is to understand how organisms are able 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 (cite climate change papers). Understanding the process and the outcomes of adaptation can also further inform crucial goals such as improving crop yields.

The demographic history of populations or species plays a major role in the strength of selection on traits and organisms (cite), and thus has the ability to leave lasting changes or signatures in the genome (cite). A large body of work aims to infer past demographic histories, such as bottlenecks, repeated founder effects during expansion, or rapid population growth, of populations based on signatures in the genome (cite). However, inferring demographic history is a difficult task, as there are many possible histories that may have occurred, and they have the potential to leave similar, or even identical, signals in some species (cite). Knowing such demographic history, however, is key, as it has a direct relationship with population size. For example, population bottlenecks greatly decrease the effective population size ($N_e$) which can increase the effect of random genetic drift. The effect size, or strength of selection ($s$), for a given locus, can also impact the efficiency of selection across the genome (selection effective when $N_e s > 1$). Thus, as demography changes, so does $N_e s$ and therefore so does the occurrence of different beneficial, deleterious, or neutral loci across the genome.

e.g. stationary pop growth looks same as range expansion unless you look at the right parameters

The complexity of demographic history is increased manyfold when considering the various architectures that species' genomes comprise. Many ecological traits important for local adaptation have a complex, quantitative genetic basis, and much heterogeneity has been found among these traits both within and among species (cite Orr 2001, Slate 2005). Whether this genetic architecture consists of many loci of small effect or few of large effect can therefore play a role in the impact of demography and selection, as well as mutation, on the genome. How the genome may be restructured as a result of various demographic histories and those effects on selection is not well studied. The genetic architecture underlying traits can affect how easily or quickly local adaptation may occur, or how long and difficult that process may be (cite Yeaman?). Such knowledge can therefore greatly contribute to improve breeding and conservation efforts as well as for predicting responses to environmental changes.

My proposed research aims to investigate these important factors (demographic history, selection, and genetic architecture) and their effects and interactions in terms of shaping the genome. I will address these questions using the maize/teosinte system. Maize (*Zea mays*, ssp. *mays*) is the domesticated species commonly known as corn, an annual plant. It is known to have been domesticated from its ancestor, the wild teosinte, approximately 9,000 years ago in southwestern Mexico (cite Matsuoka et al 2002). In particular, subspecies *Zea mays parviglumis* is the direct progenitor of maize. While two other subspecies of wild teosinte, *mexicana* and *huehuetenangensis*, exist and are found at higher elevations on the Mexican Central Plateau and in western Guatemala, respectively (cite Hufford 2012). Maize and teosinte are extensively studied and serve as ideal models for this research. Demographic history of the species is well documented: supported by both archaeological and other human records of its use and domestication in the Americas (cite). Extensive genetic and genomic data is also available, allowing the known demographic history of the species to be assessed in terms of genomic signatures of domestication (cite?), as well as specifically parameterized in terms of important values such as effective population sizes (cite Beissinger in prep). This provides a unique opportunity to compare these closely related subspecies which

have undergone various different demographic processes contributing to their present-day genetic make-up.

In this research project 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 the genetic architecture of traits changes as a result of demography and selection**. Objective one uses existing genomic data in teosinte to characterize the genetic architecture of phenotypically important traits. Using the results of this study for parameterizing key genomic characteristics, objective two then simulates the demographic history of teosinte's domestication into modern-day maize. Objective three then characterizes the genetic architecture of the same traits using genomic data from modern maize and compares to theoretical expectations of the genetic architecture of these traits based on the results produced from objective two's simulations. This research will further our understanding of demography and selection's impact on the genome and its architecture, as well as our ability to detect and predict changes in genetic architecture after certain demographic events.

# B. Research Objectives, Methods & Significance

## Objective I: Investigate the genetic architecture of important traits in teosinte

This first objective aims to estimate the distribution of fitness effects (DFE) in teosinte (*Zea mays parviglumis*), the wild progenitor of domestic maize. The DFE describes the fitness effects of various mutations that are possible within the genome. Mutations can broadly be classified as beneficial, deleterious, or neutral, but in actuality, mutation effect sizes occur on a continuum of strongly deleterious to strongly beneficial, with any value in between. Characterizing the DFE of teosinte uncovers whether the genetic architecture of important quantitative phenotypic traits in teosinte are underlain by many loci of small effect, few loci of large effect, or any combination therein.

Estimating the DFE in teosinte will contribute not only to the purposes of this study, but inform a larger body of work aimed at understanding how common different genetic architectures are for traits important in adaptation. The DFE is difficult to estimate and not broadly understood in evolutionary biology for any given organism or in terms of how much it may vary across organisms. This difficulty arises particularly for cases where there are many loci of small effects: such small effect sizes are difficult to detect on an individual basis. This has implications for studies attempting to identify loci important in adaptation struggle when each locus out of many may only contribute a small amount to the trait (cite Outi Savolainen's work in Scots pine). 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 genotype-environment associations or genome-wise association studies (GWAS).

With the vast genomic resources available in maize, an accurate DFE should be estimable. Using 70 individuals sequenced from a population of teosinte as well as 2500 individuals with genotyping by sequencing (GBS) data, the distribution of fitness effects can be estimated using the DoFE software [cite Fay, Wycoff and Wu (2001), Smith and Eyre-Walker (2002), Bierne and Eyre-Walker (2004) and Eyre-Walker and Keightley (2009), Stoletzki and Eyre-Walker (2011)] which runs a variety of analyses to estimate the components of the distribution of fitness effects. We can apply this methods to sequence data from regions of the genome known to underly important phenotypic traits. Some such quantitative phenotypes include yield, plant height, and flowering time, which are of critical importance to agriculture (cite).

I might have this data description wrong Jeff? I had in my notes also 4 additional pops in the GWAS study? I think I probably need to add a little more detail on the datasets too

Additionally, it will be possible to validate some of the performance of DoFE on the portion of loci that are inferred to be deleterious. Additional approaches such as GERP or Provean scores (cite) aim to classify the degree of effect size for deleterious mutations, as well as approaches that partition variance components of quantitative traits.

## Objective II: Simulate the demographic history of maize domestication from teosinte

This portion of the proposed project aims to use the DFE results of Objective 1 to parameterize a model that will simulate the evolution of maize and its genetic architecture through time during and since its domestication. Maize originated approximately 9,000 years ago in southern Mexico (cite Matsuoka et al 2002, others?) during a single domestication event of *ssp. parviglumis*. Archaeological records also confirm this dating and single location (cite) as well as the subsequent spread and growth in population size of domestic landraces across the Americas into both lowland and highland environments [cite Wilkes, H. G. (1967) Teosinte: The Closest Relative of Maize (Harvard Univ., Cambridge, MA)]. South American landraces of maize also underwent a second bottleneck event during their expansion (cite). Using genetic data, the precise demographic parameters of this history have been estimated (cite Beissinger et al in prep). This provides information on the ancestral effective population size of maize ($N_a \approx$ 120,000), the size to which the population was bottlenecked during domestication (5% of this $N_a$), the subsequent size to which populations rapidly expanded (3 times as large as $N_a$), and lastly on the genome-wide mutation rate ($3.8 \times 10^{-}8$ cite Clark et al 2005 MBE 22, 2304 – 2312.)

but maybe as much as 1E9, citation

The information described above can be used to parameterize simulations of maize genomes undergoing this same demographic history. The DFE of teosinte established from objective 1 provides the necessary parameters for the number of loci and their various effect sizes to be modeled as contributing to each phenotypic trait of interest. Population sizes can be matched to those known during the domestication process, enforcing the desired demographic history, and mutations can be induced at the inferred rates to create simulations as accurate to reality as possible. The forward simulation program fwdpy (a Python implementation of fwdpp, cite) provides the framework for performing these simulations of complex demographies with natural selection. This approach allows explicit modeling of the genomic architecture desired: distribution of effect sizes across the desired number of loci, including deleterious mutations and their effects on the phenotype. Recombination parameters, as well as setting population sizes, and even sampling from simulated populations is possible.

have/need info on recombination/linkage?

I will simulate the various landraces of maize which each have a different demographic history during their spread across Central and South America into both lowland and highland environments.I can then compare the different genetic architectures that evolve from these various conditions. From the simulation results, I will evaluate how many loci are found to contribute to phenotypically important traits (on average and how variable this number is), how strong the mutational effects are at each locus and if and how this relates to the number of loci contributing. This will allow assessment of how important the details of demography are in determining the genetic architecture of local adaptation to different (lowland vs. highland) or similar (lowland Mexico vs. lowland S. America) conditions.

I think this could be interesting, even if we can't compare them all in the end to real data since we only have it for some - unless it might be in the scope of this project to try and generate some of that data

Objective 2 will inform whether the DFE changes under certain demographies, and if so whether there is any meaningful direction in this change. We may hypothesize changes to a wider or narrower or more skewed distribution of fitness effects depending on the demography at hand. Population bottlenecks are a common demographic occurrence during the geographic spread of populations (cite) which will be included in our simulations. Such bottlenecks reduce population sizes greatly, and this reduction 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, cite Klopfstein et al 2006). 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). The second bottleneck of S. American populations may thus result in a difference between landraces in terms of genetic architecture of adaptive traits.

we have no GWAS data for S.Amer. pops, but do have genomes and GERP. we could get freq. etc. of del. mutations from sims and compare to GERP

does this make any sense/is it a valid comparison?

goes in intellectual merit section?

These hypothesized changes in the DFE are not well understood in any system and are con-

troversial in humans (cite lohmueller vs. pritchard etc.). 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 (cite gunther2010). 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 (cite mezmouk2014). Such information is crucial for understanding variation in phenotype, designing breeding strategies, utilizing diversity from wild relatives, or even engineering new traits using biotechnology.

### Objective III: Estimate and compare the genetic architecture underlying traits in maize post-domestication from simulations and genomic data

With the DFE of teosinte estimated in Objective 1 and the simulated evolution of this DFE from teosinte into domesticated maize in Objective 2, we now are primed to compare the predictions of Objective 2 with real world data from modern maize. Similar to Objective 1's approach, Objective 3 will estimate the DFE of modern maize (*Zea mays mays*). The expectation is that if the estimated demographic model and DFE are reasonable, the genetic architecture of simulated phenotypes should closely mimic that of real data found here. If these results are not recapitulated in the maize data, then this indicates that there are complexities in the demographic history and/or our genomic model parameters that are not well enough understood currently. In this case, we can explore the sensitivity of genetic architecture to changes to the demography or the DFE. Understanding this sensitivity will then lead to improved estimation of these important parameters for the future.

We will recapitulate the methods described in Objective 1 using genomic data from 55 sequenced individuals from a maize population as well as 2,500 individuals from the GWAS datasets.

I definitely just need to chat to you more about the data Jeff, realizing I didn't take good enough notes on things

Need some concluding remarks and things. Currently at about 5 pages of the 6 allowed, and still have to fill in C and G, so keep length in mind!!

## C. Training Objectives

Hmmm, seems somewhat related to D, need to differentiate and write more about training here, may lead to changes in D.

## D. Career Development & Future Research

My career goal is to become an independent, academic researcher who is able to push the boundaries of population genetics and evolutionary biology. As I have in my dissertation research, I aim to study population genetic and genomic processes both empirically from real work data and theoretically through simulations. I believe one of the strongest ways to advance our knowledge is through such comparisons of situations where every parameter, current and historical, is known, as well as the evolutionary outcome (simulations) to natural situations where processes or effects that are poorly understood or still unknown to us can diverge from theoretical expectations and provide the basis for further study and investigation into these processes.

The skills I will develop during this fellowship 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. First, I will gain many skills related to genomic data analysis throughout the course of this research. I have minimal exposure and experience directly working with such data from my dissertation, thus making htis a vital opportunity. 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 skills of computational biology through the proposed simulations and be able to learn a new and useful programming language used widely in evolutionary biology, Python. 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 profitable for all. Interacting with Drs. Ross-Ibarra and Thornton, as well as other researchers at UC Davis, on a regular basis will be both intellectually stimulating and rewarding experiences that will help me accomplish my career goals.

think of better word?

## 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 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 *XX*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 in Objective 2. 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

don't forget to fill in

## F. Milestones & Timeline

| | |
|---|---|
| Year 1 | put more specifics here Objective 1 |
| Year 2 | Objective 2 |
| Year 3 | Objective 3 |

want to do the DFE of maize and teosinte at the same time? no reason we can't right, and then do the simulations. Though if it takes time, teosinte should be prioritized so the sims can be parameterized and started while the maize DFE is estimated

## 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, blog posts I can contribute to The Molecular Ecologist, a blog I have contributed to in the past. The wide-ranging 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.

# References Cited

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.

# 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

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

*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, 15$^{th}$ 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, 13$^{th}$ 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 generate genotype and full-genome sequence data, phenotype data, analytical code, germplasm, and publications.

## Data Archiving, Plan for Sharing, Public Access Policy

**Genotype and Sequence Data** EDIT THIS WHOLE DOCUMENT

# 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.