1 Background

Plant domestication and agriculture rank among the most important innovations in human civilization, leading to rapid increases in population and economic specialization. From the standpoint of the plants, domestication by humans similarly ensured larger populations than might have otherwise occurred. Since both humans and plants benefit, the process of domestication can be framed as a form of bio-cultural coevolution [1–3]. In this proposal, I focus on the plant side of this framework. Through domestication, early farmers made plants more favorable to humans and cultivation. For example, domesticated plants tend to have larger fruits, less toxic compounds, less branching, and reduced seed dormancy [4, 5]. These and other traits, collectively called the Domestication Syndrome, allow farmers to grow, harvest, and consume fruits, seeds, and vegetative parts. This syndrome can become so extreme, such as in maize, that the plant becomes entirely dependent on human cultivation for survival and dispersal [6]. As early farmers moved seeds across greater distances, domesticates adapted to new environments. For example, humans brought maize northward from Mexico into the SW United States approximately 4000 years ago where maize had to adapt to drier conditions [7] and modify flowering time to adjust to long summer days [8]. Researchers initially suspected only a handful of genes controlled the domestication syndrome of maize [9]. Although large effect alleles contribute to variation in some traits (ex. *tb1*, apical dominance [10] and *tga1*, fruitcase loss [11]), many traits are highly polygenic (ex. flowering time [12, 13] and plant height [14])[4, 15, 16]. It is perhaps not surprising, then, that more than a thousand genes in maize show signatures of selection during domestication compared to wild relatives [17].

Answers about how plant and human coevolution unfolded over the last ten thousand years remain incomplete. Archaeological data have identified where and when domestication has occurred for many plant species [18–21], but how domesticates moved either with or through human communities remains murky in most cases. Instead, crop dispersal has often been modeled without considering human activity explicitly (e.g. [22]). Recent studies of contemporary communities have demonstrated farmer selection, maintenance, and dispersal through seed exchanges and community networks shape patterns of modern maize diversity [23, 24] and other crops like fonio [25], sometimes with greater effect than environmental variables [23]. For nearly all plants, though, it is an open question how early farmers' movement and trade shaped both dispersal and genetic trajectories. While a subset of obvious morphological differences between wild relatives and domesticates have been studied [26], few traits have been extensively studied in both. This hinders identification of traits separating wild from domesticated forms, and thus identifying the suite of traits which interested early farmers. Finally, different lines of evidence offer conflicting views on the speed of domestication. While archaeological and genetic evidence suggests domestication was a protracted process lasting thousands of years [8, 19–21, 27, 28], genetic and experimental studies show that key genes or traits could have changed rapidly over a few hundred years [29–31]. Reconciling these differences remains an area of active research [1, 3, 18].

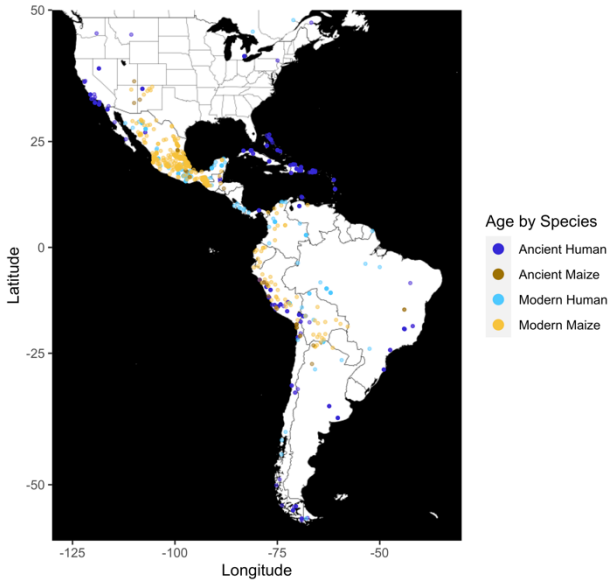

Figure 1: *Distribution of both ancient and modern human and maize samples. Not shown are 39 human samples from northern North America (34 ancient, 5 modern).*

The domestication and dispersal of maize across the Americas is an excellent example of and system for developing coevolutionary models. Extensive genomic resources exist for both humans and maize with widespread sampling of modern populations and ancient samples (Figure 1). Because of the

cultural and economic importance of maize, the research community has invested heavily in extensive phenotyping experiments across a variety of genotypes and environments. Extensive population sampling, high-quality reference genomes and supporting -omics datasets, and sustained, well-documented, coordinated phenotyping efforts across diverse genotypes and environments of maize and humans creates the opportune dataset. Maize is thus a prime candidate to test human-plant coevolution models.

Here I propose inferring population genetic models of coevolution using human domestication and dispersal of maize in the Americas. First, I will develop a novel approach to estimate a combined admixture graph of humans and maize. With these graphs, I can formally test coevolution models describing how maize moved with and between human communities. Second, I will use these graphs to identify the key traits and timing of selection during maize evolution. These estimates will inform our understanding of the temporal sequence of phenotypic selection on plant traits by early domesticators during initial domestication and later range expansion. Finally, and importantly, I will apply these methods to other domesticates to test the generality of patterns found in maize. This research fits the call for competitive area 3: Plant Genome Postdoctoral Research Fellowships by harnessing the power of high-throughput sequencing and phenotyping efforts with new analytical methods to answer basic questions about plant domestication and crop-human interactions.

2 Research Plan

Aim 1: Reconstruct human influence on maize dispersal using joint admixture graphs

Admixture graphs are models representing gene flow and isolation between populations over time, by extending population phylogenies to include reticulation [32, 33]. Allele frequencies characterize a population (node), and shared genetic drift connects each population (edges) [32, 34]. Use of ancient DNA samples of known age help calibrate admixture models, as shown in [33]. Because migration events are allowed between any node, the number of possible graphs increases exponentially with additional populations [35]. Thus, it becomes impractical to test every possible graph given a large number of populations. By using known histories [8, 36–38], the number of possible graphs becomes manageable to test. Typical use of admixture graphs have focused on a single species or set of closely related species (for examples, see [39–42]). Researchers have compared the topography of two species' admixture graphs [43], but admixture graphs were created separately and then *posthoc* comparisons made. Here, I propose a novel approach — using the human admixture graph as the backbone to then estimate a maize admixture graph (Figure 2).

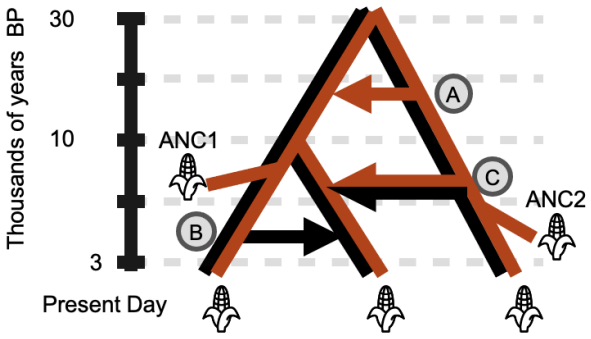

Figure 2: *Cartoon representation of a joint admixture graph with the distant past at the top. The human admixture graph is in black, and maize is in rusty red. Arrows between branches represent admixture events, labeled as A, B, and C. A and B correspond to admixture events in a single species, maize and humans respectively. C shows a shared admixture event with a dual-colored arrow. Ancient samples are placed as separate branches to the most closely related lineages and time point, illustrated by ANC1 and ANC2.*

Expected Outcome: I will develop a method for estimating admixture graphs using two species, which I expect to be broadly applicable across systems other than maize.

Subaim 1.1: Estimate human admixture graph: I will estimate the admixture graph of humans in the Americas using modern and ancient samples. The genetic data come from published sequences that cover the geographic range of the Americas and ages ranging from recent to 12,712 years before present (Figure 1). This admixture graph will be informed by previous analyses of population genetic histories of indigenous groups [36, 37].

Subaim 1.2: Formally test human effects on maize dispersal: Given the dependency of maize on human cultivation, I can constrain the topology of the maize admixture graph to that of the human admixture graph estimated in subaim 1.1 and take into account known maize population histories [8, 38]. I will allow maize to form new edges, corresponding to admixture events in maize that are not shared in humans due to economic activity such as trade. The maize genetic data also spans the geographic range of the Americas and includes extant traditional varieties and ancient samples (Figure 1). These data come from previously published work and unpublished data of ~ 300 traditional varieties from the Ross-Ibarra lab. By estimating this joint admixture graph, I can formally test human effects on maize dispersal and admixture.

Aim 2: Describing the temporal dynamics of polygenic selection in maize

Despite the stark phenotypic differences between maize and teosinte [26], the genetic signal of adaptation is hard to identify as selection can drive small shifts across many loci producing drastic changes in quantitative traits [44]. It is also usually unclear when these shifts occurred that have resulted in the distinct suite of phenotypes observed today. However, these difficulties can potentially be overcome by calculating polygenic scores from GWAS for phenotypes of interest for modern and ancient samples. These scores can then be used to help identify subtle shifts in allele frequency that drove large changes in polygenic score and phenotype [45, 46]. Furthermore, estimated polygenic scores for ancient samples leads to predicted trait values for traits that are not preserved (*e.g.*, plant height, flowering time; but see [47]). As the age of ancient samples is known through radiocarbon dating, their sequences can act as proxy for allele frequencies in ancestral populations allowing us to identify the emergence of domestication traits, even if they're not direct ancestors of modern samples.

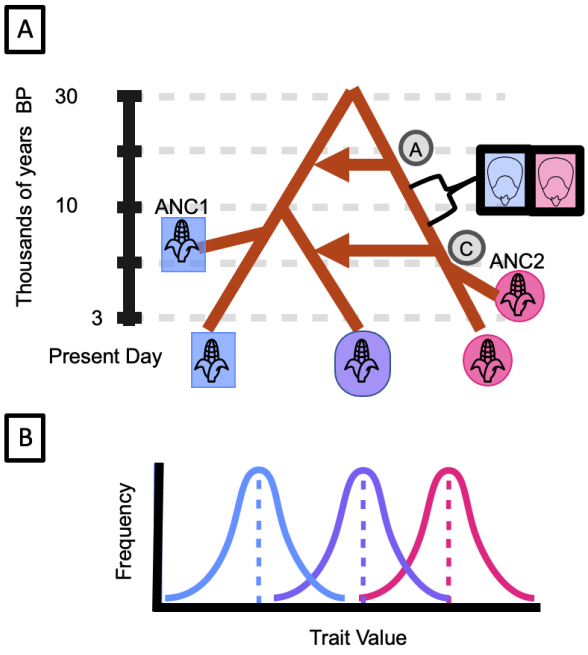

Figure 3: *Cartoon representation identifying polygenic selection on the maize admixture graph from Figure 2. Panel A shows the maize admixture graph with admixture events A and C. The bracket corresponds to a period of polygenic selection on a trait— which could be temperature tolerance, germination rate, kernel color, etc.— here visually referenced as kernels with a blue background becoming red. Maize lineages experiencing the selection have the selected trait value indicated by a red circle, the unselected trait values represented by a blue rectangle. Because of admixture (event C), the admixed population has an intermediate trait value indicated by a purple rounded rectangle. Panel B shows each of the corresponding lineages (blue, purple, or red) differ in their trait value distribution. Kernel image made by Poeticus from the Noun Project.*

A recently developed method combines GWAS results and admixture graphs to identify in which lineages and at what time polygenic selection for a trait has occurred [48] (Figure 3). I propose to apply this approach using the admixture graph from Aim 1 and association mapping results for more than 200 phenotypes from diverse maize mapping populations (Table 1) to test several explicit hypotheses.

First, I can identify the traits that were important during the initial phase of domestication or were later improved [18], and those traits that have continued evolving outside of artificial selection [49] by linking detection of selection to specific segments of the admixture graph [48]. Identifying traits that show evidence of selection early will lend insight into the initial domestication process, which may have been sporadic as communities

vacillated between foraging and cultivation [20, 21]. Further I can begin to discern the timing of selection across traits by evaluating trends across branches of the admixture graph. If traits were under greater initial selective pressure that later eased, I should see more estimates of selection in earlier branches instead of a more constant distribution across all branches.

Second, I can identify branch specific selection that may have been important for local adaptation and range expansion. For example, maize lineages brought into the highlands of Mexico (and later South America and the SW United States) had to adapt to a colder, drier environment and show distinct local adaptation through stalk anthocyanin accumulation and increased macrohairs [49, 50]. Maize also had to adapt to the long day length in higher latitudes through changes in flowering time [8, 12, 51]. Comparing these well-studied examples with my results will validate my approach for detecting signatures of local adaptation to new environments. But even for these examples, not all of the traits, or even the timing [52] of adaptation are known, and even less is known for most other instances of adaptation in maize.

Finally, I can begin to identify changes in maize associated with specific human cultures. Studies demonstrate current farmers and communities have different goals and selection criteria for their maize cultivars [23, 24]. Early cultures likewise had diverse priorities. I can begin to identify selection due to cultural preferences by tracking selection signatures in maize associated with a human community. For example, *su1* alters the ratio of starch to sugar in the endosperm which affects food proprieties important for culinary preferences and was under selection during maize expansion into the SW United States [7].

Expected Outcome: I will identify when and in which lineages polygenic selection has acted upon a variety of maize traits and relate these patterns to domestication, range expansion, and human activity.

Aim 3: Generalizing maize domestication patterns to other species

Early indigenous farmers domesticated many other plant species alongside maize. While maize ranks among the best systems to develop methods and test hypotheses of domestication, more can be learned from including additional domesticates. For example, amaranth (*Amaranthus caudatus*) may have at one point been as culturally important as maize [53], but has not been completely domesticated [53–55]. By applying the methods developed in Aim 1 and 2 to additional domesticated species of the Americas, I can evaluate if patterns identified in maize are general and have parallels in other contemporaneous plants and animals, and the extent to which they are species-specific.

For this aim, I will use published sequences of domesticated and wild accessions of amaranth, sunflower, tomato, common bean, and quinoa and related QTLs (Table 1). While there are no or limited ancient DNA samples available for these species, the methods described in Aim 1 and 2 should remain applicable. Additional genomic resources have been forthcoming, and continuing trait mapping efforts will add more crops to this list (ex. recent release of potato genomes [56]).

Expected Outcome: I will apply the novel methods developed in Aim 1 and 2 to additional domesticates to uncover shared patterns during domestication.

Table 1: *Mapping populations and diversity panel resources for domesticated crops from the Americas. To be included, the population need high density genetic markers (≥ 5K SNPs) and associated phenotypes. GWAS results tabulates the number of traits measured across publications for a given panel. A given trait may be measured in multiple studies.*

Domesticate	Panel Name	Panel Size	GWAS Results
Maize	NAM [57]	26 founder lines	143 [12,14,58-71]
	282 Diversity Panel [57]	282 accessions	100 [60,62,67,69,70,72-78]
	Ames Association Panel [79,80]	2500 accessions	17 [14,16,81-83]
	Wisconsin Diversity Panel [84]	627 accessions	31 [85-90]
	527 Diversity Panel [91]	527 accessions	19 [92-94]
Tomato	F2 *S. pennellii* Introgression Lines [99,100]	50-76 lines	19 [99-101]
	Diverse Accessions	295-628 accessions	8 [102,103]

	Graph Pangenome [104]	838 genomes	19,535 eQTLs, 970 metabolite traits [104,105]
Sunflower	Diverse Accessions	543 accessions, 1401 plants	110 [106-107]
	SAM [108]	271 accessions	24 [109-111]
	F1 Mapping Population (*H. annuus*) [112]	480 hybrids	1 [112]
	F2 Mapping Population (*H. argophyllus*) [113]	400 individuals	10 [113]
Common Bean	Andean Diversity Panel [114]	237 accessions	30 [114-116]
	Middle American Diversity Panel [117]	280 accessions	7 [117,118]
	Diverse Accessions	118 accessions	5 [119]
	$F_{2:7}$ RIL Population [120]	85 RILs	11 [120]
Quinoa	Diverse Accessions	106-310 accessions	24 [121-123]
Amaranth	Diverse Accessions	121-1000 accessions	12 [53,95-98]

3 Training Plan

This post-doctoral fellowship is an opportunity to formally train in statistical population genetics and learn project management skills for large, interdisciplinary collaborations. My fascination with domestication as an evolutionary process led me to a Ph.D. in comparative genomics where I used bioinformatics to understand short- and long-term outcomes of maize hybridization. I have the skills to manage large datasets and extract useful information, but I have limited training in theoretical and statistical population genetics. Since population genetics bridges large-scale bioinformatics and biological meaning, it is a crucial framework to address long-standing biological questions in the field of domestication. The training opportunity to work with Dr. Ross-Ibarra and Dr. Coop, who are leaders in the field, is an exciting prospect. Furthermore, this proposal will allow me to expand my bioinformatic and genomic experience to humans (Aims 1, 3) and other domesticates (Aim 3) as well as the interactions between species.

Science is collaborative. As a future lead researcher, I will need to forge connections with other research groups and lead collaborative projects to develop a robust research program. I have experience in collaborative research within and across institutions, but not as the lead researcher. This opportunity to be co-advised is an ideal training ground for developing and managing broad research collaborations. The environment at UC Davis is highly collaborative and a hub for population genetics. Beyond the immediate network of my co-advised lab groups, I can connect with anthropologists and Native American Studies scholars who study Mesoamerican foodways. Connecting with this large network of scholars focused on population genomics, evolution, and indigenous agriculture and food systems will be important for my development as an interdisciplinary scholar. Through this opportunity, I will gain the skills necessary to plan, manage, and deliver as lead on a collaborative project and connect with an array of researchers in my specific and related fields.

4 Career Development

I have taken a broad approach in my Ph.D. training to understand the extent of biological insight that can be gained from the increasingly large datasets generated from high-throughput sequencing. The formal training this fellowship offers in population genetics will be key to drawing such biological insight from high-throughput sequencing datasets. And while I plan to build a career in maize genetics and evolution, branching out into other systems will broaden my perspective, strengthen my skills in all dimensions of research, and permit insight into general principles of crop-human coevolution.

As a future professor, I also value teaching and curriculum development experience. I have sought out professional development opportunities specifically geared towards understanding pedagogy and

gaining experience guest lecturing. While these experiences have been illuminating, I have yet to develop *de novo* curriculum or teaching materials. Through this fellowship, I will have the opportunity to translate research components into curriculum either as a section of a larger course or as a freshman seminar on domestication. This important experience will enable my growth as an educator.

5 Choice of Sponsoring Scientists and Host Institution

Dr. Jeffrey Ross-Ibarra and Dr. Graham Coop are leading experts in population genetics and genomics, have worked with ancient DNA, and have collaborated previously. Dr. Ross-Ibarra specializes in maize genetics and has done foundational work in using demographic modeling and genomics to better understand maize domestication. Dr. Coop specializes in human genetics and has made significant contributions to understanding polygenic adaptation, estimating admixed ancestry, and identifying the timing of selective events. Their land grant university, University of California Davis, is well-known as a hub of collaboration and population genetics expertise in the community. I believe this will be an encouraging environment for developing interdisciplinary research collaborations and professional networks.

6 Timetable

In year 1, I will accomplish Aim 1 and begin work on Aim 2. I will also build a collaborative network of other population geneticists, evolutionary biologists, and scholars of Native American studies both within and outside of my co-advised lab groups. I will connect with the experts on campus, such as anthropologist Dr. Brenna Henn who works on human population genetics, scholars from the renowned Department of Native American Studies, and researchers of Mesoamerican foods. Further, I will attend summits hosted on campus such as the Seeds and Cultures Summit which focuses on Mesoamerican foodways. In year 2, I will complete Aim 2 and start work on Aim 3. I will present the findings of Aim 1 and the preliminary results of Aim 2 at conferences as a poster or short talk. At this point, I will be writing the main paper for the maize results of Aim 1 and 2. I will also partner with my network established in year 1 to develop curricula from research findings and build a framework for a freshman seminar. At this time, I will begin testing these materials as a guest lecturer. In year 3, I will finish work on Aim 3 and prepare the main paper for publication. I will present on the full set of results at conferences as short talks. I also plan to deploy the finished teaching materials as an online resource and use them as part of either established biology courses or as a special topics discussion course. At this point I will be applying for academic jobs.

7 Broader Impacts

The methods developed through this research are intended to be general, with applicability to all domesticated animal and plant species that have the enabling genomic, agronomic, and archaeological resources. Methods and their associated code will be made freely available so other researchers can use them in their own systems. In addition to the dissemination of the research findings through traditional methods such as conference presentations and publications, I will develop curricula for undergraduate education. The goal of the curricula will be to engage student's curiosity about population genetics and showcase the possible questions that can be answered with new high-throughput sequencing, ancient samples, trait mapping, and new analytical methods. The curricula will contain the methods and findings of the proposed research aimed at biology majors. Because of the sensitive nature in developing teaching materials involving historically excluded groups, I will seek collaboration from scholars specializing in Native American studies to better contextualize plant domestication as a relationship between humans and plants both historically and today. All teaching materials will be made available online and hosted on community resource platforms such as the Teaching Tools in Plant Biology hosted by *The Plant Cell* and the American Society of Plant Biologists. I also plan to use the materials to guest lecture in the Introduction to Evolution, Crop Evolution, and/or Foods in Mesoamerica courses at UC Davis. Finally, I will seek out opportunities to mentor and support graduate and undergraduate students, as I have done throughout my Ph.D. program, both formally (6 undergraduate students, 3 graduate student organizations) and informally.