1 Summary

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.

Intro

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?)

selection interacts with demography. $N_e s > 1$ for selection to win. as demography changes, $N_e s$ changes. so demography is known to directly affects the strength of selection, but what is less studied is how this may restructure the genome in terms of how evolution proceeds for adaptation to new or changing conditions.

Different demographic histories such as bottlenecks, repeated founder effects during expansion, or rapid population growth can have effects that interact with selection. Small population sizes can lead to purging (recessives become homozygous and removed in smaller pops), but also increase in deleterious alleles (surfing phenomenon due to random genetic drift). mutations that affect a trait related to fitness will then be impacted by demography x selection interaction. in annual plants, nearly ALL traits related to fitness we thus predict that genetic architecture (number and size of mutations) should be impacted by demographic change. this is not understood well in any system. controversial in humans (lohmueller vs. pritchard etc.). in plants, summaries are on gross overall N_e (small vs. big) and ignore recent demographic change.

Proposed Research

The maize/teosinte system, has a well described demographic history, supported by archaeological and other human records of its use and domestication in the Americas.

Maize was domesticated ≈9000 years ago in SW Mexico which bottlenecked populations to a small size, but was subsequently followed by a large population expansion more recently. This demographic event can be detected genetically, as shown by Beissinger in prep (github repo), where populations were bottlenecked to 5% of their previous size followed by spread across central and southern America and into highland and lowland environments. Populations of maize have since recovered to much larger effective population sizes than even before the domestication bottleneck (humongous growth to at least 300K but maybe as much as 1E9, citation).

Being a crop species that is incredibly useful to humans (lots of citations and examples of usage), there is also a wealth of knowledge on quantitative traits. will need to explain rare alleles pops and experiment some (PGRP15 grant)

This system thus serves as a great resource to compare the effects of this history on the genome in maize versus its extant, wild ancestor teosinte, particularly to investigate how the genetic architecture

of these important traits may have evolved differently as a result of this combination of demography and selection.

1.1 Obj 1 - Estimate the Distribution of Fitness Effects (DFE)

note to self, update the data description

using teosinte genomes

- from polymorphism/divergence data. use HapMap2 or current teosinte genomes (I think I'd vote for latter) `http://goo.gl/CLmsmX` and `http://www.genetics.org/content/177/4/2251.short`) can either use estimated demographic model or use noncoding sites to normalize SFS

Methods: estimate the DFE using Eyre-Walker's DoFE `http://www.lifesci.susx.ac.uk/home/Adam_Eyre-Walker/Website/Software.html`

this part prob not in proposal:

can validate by comparing to GERP distribution partitioning variance components, e.g. `http://www.ncbi.nlm.nih.gov/pubmed/25439723` GREML software

useful because the distribution of mutation effect sizes is not generally known, and is especially difficult for small effect mutations. objective 1 will inform perhaps what the DFE may look like in any organism with a history similar to teosinte? and just in general add to the body of literature on genetic architecture, mutation effects

1.2 Obj 2 - Simulate scenarios of different traits.

from objective 1 we now know the DFE of teosinte we already know the demographic history of maize since its split from teosinte (citations)

use the DFE results to parameterize a model that will simulate the evolution of maize and its genetic architecture through time during and since its domestication we can simulate maize that expanded into S America separately since it has a different demography and then compare any differences the regions may show in the end simulate traits w/ varying correlation with fitness new mutation effects on fitness determined by DFE, effect on trait by correlation between trait and fitness

Methods: fwdpy (python version of fwdpop, cite `http://www.genetics.org/content/198/1/157.abstract`)

evaluate: how many loci contribute to important traits? how strong are these effects? how do details of demography impact outcome? test against theory e.g `http://arxiv.org/abs/1312.3028`

standing questions: does the DFE significantly change in a meaningful way or a certain direction? mean value the same but narrower or wider distribution? skewed more one way or the other? might expect this to be a different answer for maize in S America vs Central since S America has had a second bottleneck, so more founder effects and more chance for drift

(could also do some broader simulated examples just to stand alone and see if other various outcomes may occur - just don't plan on comparing these to any real data)

yes, once we can show we can recapitulate real data, i think this is useful

1.3 Obj 3 - compare simulation results to modern maize genomes, known to have undergone the same demographies simulated in objective 2

compare to GWAS for maize/teo. do we recapitulate observations? if not, why? are there differences between central and southern American pops?

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

Methods: same DoFE approach in C American maize pops for direct comparison in S American pops, can do comparison on subset, e.g. GERP, which we would have from Obj 1 if we compare to other approaches for the sort of validation of the DFE (definitely worth doing if b/c of 2nd bottleneck more deleterious stuff rose in frequency and was then eliminated)

Big picture thoughts

it is thought, and shown in some human pops, that demog. history such as expansion leads to an increase in delet alleles, and of larger effects - b/c of continued inferred expansions and bottlenecks. is there any evidence of this in maize?

Quantitative phenotypes such as yield, plant height, and flowering time are of critical importance to agriculture. Deleterious alleles likely play a large role in many of these 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 hypothesis, my lab has recently shown that genes associated with a number of quantitative traits in maize are enriched for deleterious alleles compared to randomly chosen genes (cite mezmouk2014). However, while we know that demography impacts the frequency of individual deleterious variants, we have a poor understanding of the interaction of demography and selection on phenotypic variation. In particular, we know little about how these two forces interact to determine the genetic architecture – the number of genes and their effect – of a trait. Such information is crucial for understanding variation in phenotype, designing breeding strategies, utilizing diversity from wild relatives, or even engineering new traits using biotechnology.

see ideas and text in "service_award.tex" that I uploaded too.