

# Identifying genomic regions under selection during the domestication of noug (*Guizotia abyssinica*)

## Abstract

Certain traits are selected for during the course of domestication. Evidence of positive selection manifests itself as reduced variation between particular loci in the domesticate relative to its wild progenitor (Volis 2008). I propose to look for evidence of such selective sweeps in the Ethiopian composite crop *Guizotia abyssinica* relative to its progenitor *Guizotia scabra ssp. schimperii* using a genome scan in combination with a genetic map to look for regions of reduced diversity in the domesticate.

## Introduction

Domestication is often characterized by strong artificial selection. Both “bottom up” approaches that rely on *a priori* selection of traits and “top down” methods that rely on population genetic signatures of selection can be used to identify the genes responsible for domestication (Wright and Gaut 2005). Until recently, the expense of genome-wide scans has been unjustifiable for non-model or commercially unimportant species. However with the advent of next-generation technology and the reduction in cost that accompanies it, we now have the opportunity to investigate the genomic basis of domestication in many species.

Evidence of positive selection manifests itself as reduced variation around selection loci, as well as strong divergence in the domesticate relative to its wild progenitor (Volis 2008). I propose to look for evidence of such selective sweeps in the Ethiopian composite crop *noug (Guizotia abyssinica)* relative to its progenitor *Guizotia scabra ssp. schimperii* using a genome scan in combination with a genetic map to look for regions of reduced diversity and increased divergence in the domesticate. Specifically, I will ask the questions: 1) What genes were selected during domestication and improvement in noug? 2) What proportion of the genome do they represent?

Noug is an annual crop that is primarily grown in Ethiopia, Eritrea and southern India as an oilseed but is also exported to western markets as a birdseed (Getinet & Sharma 1996). This crop is a member of the *Compositae* (Sunflower) family and is sympatric to its progenitor, *Guizotia scabra ssp. schimperii* across much of its range in Ethiopia (Hiremath and Murthy 1987).

Noug is a unique crop because it does not display the strong signs of artificial selection and looks similar to its progenitor (Dempewolf et al 2008). It does not display the typical domestication syndrome traits of apical dominance, non-shattering seed and self-compatibility (Funk et al 2009). Therefore it would be

advantageous to take a “top-down” approach which does not rely on the *a priori* identification of a phenotype of interest. Using this approach, I will look for evidence of artificial selection on this species and attempt to identify target candidate genes.

Noug is an under-developed crop that shows little evidence of genetic differentiation between domesticated accessions (Dempewolf 2011). Expanding our knowledge of the domestication history of noug and attempting to understand what traits were selected for during its domestication may help us understand what traits were important to ancient farmers and this may help us to improve the crop today.

## **Experimental Design**

In this experiment, I will attempt to find evidence of selective sweeps occurring during the domestication of noug by comparing allelic diversity of shared loci between the crop species and its progenitor. A selective sweep centers around a functional variant that actually causes a change in phenotype. Genetic hitch-hiking occurs when neutral alleles close to the selected variant increase in frequency due to linkage (Maynard Smith & Haigh 1974). A recent selective sweep may show a wide swath of reduced variation in neutral alleles close to the selected variant. As time goes on, recombination may slowly separate the functional allele from its hitch-hiking neutral alleles (Maynard Smith & Haigh 1974). Therefore, an ancient selective sweep may show little or no area of reduced variation and may be more difficult to detect (Renaut et al. 2011). However, because the domestication of noug has occurred recently (Boardman 1999), it is likely that this method will identify domestication genes, if they exist in noug.

In order to adequately capture the diversity of each species, I propose to sample a small number of individuals from many populations throughout the range of *G. scabra ssp. schimperi* as well as many accessions of domesticated Noug throughout Ethiopia, Eritrea, and, if possible, southern India. As my study is concerned with selective sweeps that occurred during domestication rather than during the possible creation of informal landraces, it is more important to sample the breadth of the species, rather than depth of fewer populations. This data could also allow me to examine the population structure of noug in more detail, although a previous study of this species (Dempewolf 2011) has found little evidence of population structure within noug.

To assess regions of reduced diversity within the domesticate, I will use genotyping-by-sequencing (Elshire et al 2011). GBS is a method of targeted sequencing that reduces the complexity of the genome while retaining a ‘genome-wide scan’ function. A methylation sensitive restriction enzyme is used to cut genomic DNA (extracted from somatic tissue samples) in order to target gene-rich

regions. Next, libraries of the cut fragments are prepared and sequenced on the Illumina HiSeq. SNPs are identified by aligning the cut fragments of multiple samples with each other and the reference transcriptome.

To determine the genetic distance between loci, I will create genetic maps of noug and its progenitor by crossing a domesticate and a wild individual. I will then genotype each parent and 192 F1 offspring using GBS and then create linkage maps for each parent (e.g. Kriegner et al 1999).

I will use a maximum likelihood statistical method to assign genotypes to each sample as done in Hohenlohe et al. 2010. This method removes potential SNPs that cannot be confidently assigned due to too few reads and estimates sequencing error on a per site basis. I will analyze the divergence between the progenitor and the domesticate by calculating the observed heterozygosity as the proportion of the diploid genotypes in the samples that are heterozygotes, as well as the expected heterozygosity to obtain an  $F_{ST}$  value at each locus (Hohenlohe et al. 2010). I will then identify outlier loci, which will have higher levels of differentiation than the rest of the genome.

I will align the GBS data to a consensus transcriptome assembly of noug and *G. scabra ssp. schimperii* (Rieseberg lab, unpublished) to separate fragments that do not align to the reference transcriptome. Then I will assemble the remaining reads *de novo* to capture SNPs outside of gene regions or in genes not expressed in the reference transcriptome.

I will also compare the nucleotide diversity between noug and *G. scabra ssp. schimperii* at each verified SNP and use a sliding window to find regions within noug that have a reduction in diversity relative to the rest of the genome and the progenitor. During domestication, situations such as founder effects, migration and genetic drift may have caused the domesticate to experience a global reduction in nucleotide diversity relative to its progenitor and looking at relative reductions in diversity is an attempt to remove this effect (Chapman et al 2008).

Once regions of relatively reduced diversity in noug are identified, it would be interesting to identify the potential function of candidate SNPs which may be functional variants. As I am aligning reads to the reference transcriptome, some of the SNPs discovered will be expressed as mRNA and may actually work to change the coding region of a protein. To identify candidate SNPs that change the coding region of a protein, I could use an open-reading-frame finder program to search for potential ORFs in the region of reduced variation. If a SNP is located within a potential open reading frame, I could determine whether the various alleles are non-synonymous changes. I could also use BLASTx to match ORFs previously found in regions of low diversity and high divergence to homologous coding sequence in *Arabidopsis thaliana*. Annotations from high ranking *A. thaliana* genes may give an idea of the function of the candidate genes in noug.

## Possible Outcomes

Even though noug does not display a strong domestication syndrome (Dempewolf 2011), the species has likely been under conscious or unconscious artificial selection during its domestication. As the SNPs discovered with GBS will be scattered throughout the transcriptome and their number will be very large, it is very likely that I will find areas of lower diversity in the domesticate relative to the progenitor if selective sweeps occurred during domestication. A large region of decreased nucleotide diversity may indicate a newer selective sweep while a small region of decreased diversity may be indicative of an older selective sweep (Maynard Smith & Haigh 1974).

In a similar survey done in sunflower, the authors found evidence of selection for genes that affect flowering time and amino acid composition. They also found an area under selection that harbored QTL for seed size (Chapman et al 2008). As cultivated noug has larger seeds compared to wild *G. scabra ssp. schimperii* (Getinet & Sharma 1996), I may find evidence of selection on genes that control seed size. Ethiopian farmers recognize three types of noug that vary in maturation time (Petros et al 2007), so I may find evidence of selection for genes involved in regulating flowering time. Since noug is grown for oil production (Getinet & Sharma 1996), I may find evidence of selection on genes relating to oil production.

## Future Directions

The genetic architecture of domestication in noug could be explored more thoroughly to determine the distribution of loci under selection, and compare the distribution to that of noug's closest relative, sunflower. As noug does not display traits associated with strong domestication syndrome, it would be interesting to compare the architecture of selection with other crops that display similar or stronger domestication syndromes.

## References

1. Boardman, S. J. 1999. The agricultural foundation of the Aksumite empire, Ethiopia. *Plants and People in Africa: Recent Archaeobotanical Evidence*. (ed M. van der Veen), pp. 137-147. Plenum Publishing Corp, New York.
2. Chapman, M. A., Pashley, C. H., Wenzler, J., Hvala, J., Tang, S., Knapp, S. J., Burke, J. M. 2008. A Genomic Scan for Selection Reveals Candidates for Genes Involved in the Evolution of Cultivated Sunflower (*Helianthus annuus*). *The Plant Cell* 20: 2931 - 2945
3. Dempewolf, H., Rieseberg, L. H. & Cronk, Q. C. 2008. Crop domestication the *Compositae*: a family-wide trait assessment. *Genetic Resources and Crop Evolution* 55: 1141 - 1157

4. Dempewolf, H. *Patterns of domestication in the Compositae and beyond*. Diss. University of British Columbia, Vancouver, 2011. Print.
5. Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., Mitchell, S. E. 2011. A Robust, Single Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6(5): e19379
6. Funk, V. A., Susanna, A., Stuessy, T. F. & R.J., B. (Eds.). 2009. *Systematics, evolution and biogeography of Compositae*. International Association for Plant Taxonomy, University of Vienna, Vienna.
7. Getinet, A. & Sharma, S. 1996. Niger, *Guizotia abyssinica* (L. f.) Cass. *Promoting the conservation and use of underutilized and neglected crops*. p. 59. International Plant Genetic Resources Institute Rome, Rome.
8. Hiremath, S. C., Murthy, H. N. 1987. Domestication of niger (*Guizotia abyssinica*). *Euphytica* 37: 225 - 228
9. Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., Cresko, W. A. 2010. Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD tags. *PLoS Genetics* 6(2): e1000862
10. Kriegner, A., Cervantes, J. C., Burg, K., Mwanga, R. O., Zhang, D. P. 1999. A Genetic Linkage Map of Sweetpotato (*Ipomoea batatas* (L.) Lam.) Based on AFLP Markers. *CIP Program Report 1999-2000*: 303 - 313
11. Maynard Smith, J., Haigh, J. 1974. The hitch-hiking effect of a favorable gene. *Genetical Research*, 23: 23 - 25
12. Petros, Y., Merker, A., Zeleke, H. 2007. Analysis of Genetic Diversity of *Guizotia abyssinica* from Ethiopia using Inter Simple Sequence Repeat Markers. *Hereditas*, 144: 18 - 24
13. Renaut, S., Nolte, A. W., Rogers, S., Derome, N., Bernatchez, L. B. 2011. SNP signatures of selection on standing genetic variation and their association with adaptive phenotypes along gradients of ecological speciation in lake whitefish species pairs (*Coregonus* spp.). *Molecular Ecology*, 20: 545 - 559
14. Volis, S. 2008. Chapter X: Detection of Signatures of Positive Selection in Naturally Occurring Genetic Variation. In *Population Genetics Research Progress* First Edition; Koven, V. T., EDs.; Nova Science Publishers, Inc. Hauppauge NY pp. 279 - 310
15. Wright, S. I., Gaut, B. S. 2004. Molecular Population Genetics and the Search for Adaptive Evolution in Plants. *Molecular Biology and Evolution*, 22(3): 506 - 519