AFLP Evaluation of Genomic Variation and Population Structure in *Andropogon gerardii*

R.C. Lynch[1 *] and A.M. Martin[1]

Abstract:
AFLP profiles were generated for *Andropogon gerardii* (big bluestem) samples taken from individuals displaying a wide range of phenotypes, including plant heights from 10.5 inches to 75 inches, across five sites in Boulder County. A Bayesian clustering analysis of the allelic frequencies determined that the samples did not form a genotypic population structure based on either plant stature or sample collection site. Furthermore, pair wise genetic distances between all samples revealed a mean of 6.3% difference, indicating the degree of local genetic diversity in big bluestem.

[1] Department of Ecology and Evolutionary Biology, University of Colorado, Boulder CO 80309

[*] rlynch@colorado.edu
[1]

Introduction:

Prairie grasslands on both the local and global scale are highly endangered, yet Boulder County currently has the unique opportunity to preserve and restore various versions of these unique ecosystems. Ranging from the Atlantic Ocean to the Rocky Mountains, from southern Canada to northern Mexico, the tallgrass *Andropogon gerardii* (big bluestem) is a major constituent of various prairie plant communities.[1] Continued detailing of its ecological, reproductive, and evolutionary characteristics will contribute useful information to the scientific community, and inform effective conservation and restoration efforts. Due to the severity of habitat reduction and fragmentation, a key question to ask is: are there distinct genetic sub-populations within the taxon *Andropogon gerardii*, and if so, do the populations form any spatial structure?

In light of the cost of propagating native seed stocks, all prairie grassland restoration strategies should be carefully evaluated before large scale implementation. Knowing where to initially harvest seeds from, and understanding the environmental habitat limitations of seeds will greatly enhance the cost and time effectiveness of seeding efforts across the variety of big bluestem habitats in Boulder County. Big Bluestem is the dominant vegetative component in two conservation targets, identified in the City of Boulder-Open Space and Mountain Parks (COB OSMP) Grassland Ecosystem Management Plan, as the Xeric Tallgrass Prairie and Mesic Bluestem Prairie. This project seeks to both fill in the unique evolutionary story behind local big bluestem grass populations, and by extension, helps answer whether one seed stock grown from one of these communities can be used for restoration in both types of environments.[2] Posed at a basic level, we are asking, how closely related are the populations growing in xeric and in mesic systems? Are there forces of natural selection driving genetic differentiation and

or reproductive isolation between populations? Or are they simply one variety that has the phenotypic plasticity to thrive across a range of habitats?

Work carried out on Boulder County populations of big bluestem has already described an interesting variation in chromosome number that affects seed productivity and plant vigor. Although the 60 and 90 chromosome cytotype versions do not seem to assort specifically with either xeric or mesic environmental conditions, this does not mean that adaptive changes are lacking in populations associated with these conditions.[3,4] Previous studies of grass genetics have revealed differentiation that can occur in both small scales, down to several meters, and across larger geographic distances that encompasses different climatic conditions.[5,6] Big bluestem can be found in quite a wide variety of locations, and has at least a few intriguing phenotypic traits with noticeable diversity, here in Boulder County. Big bluestem was found growing between 10.5 and 75 inches tall, with a spectrum of leaf/stem colors ranging from bright green to deep purple, and notably, having anything between completely smooth to fairly pubescent leaf surfaces. In particular, leaf pubescence can, at least in some cases, be an adaptation to xerophytic conditions.[7] Although thought of as a plains species, big bluestem was observed just below tree line, on both sides of the continental divide living in open, tree free areas at well over 9,000 ft. Given that most BCPOS lands lay on the varied terrain of the transition between plains to high mountains, there is perhaps all the more opportunity for micro-niche specialization. On the other hand, being a wind pollinator, one would perhaps expect strong local gene flow-and thus a relatively homogeneous population on the local scale which this study is addressing. Sympatric speciation events occurring within a geographic region where the initial single species has no physical gene flow

barrier, is not unheard of though—with the evolutionary radiation of the striking diversity of cichlid fish species in African lakes being a famous example. At the genetic level there is no way to guess how different various types of big bluestem are, without actually making some genetic comparisons from population samples.

Methods:

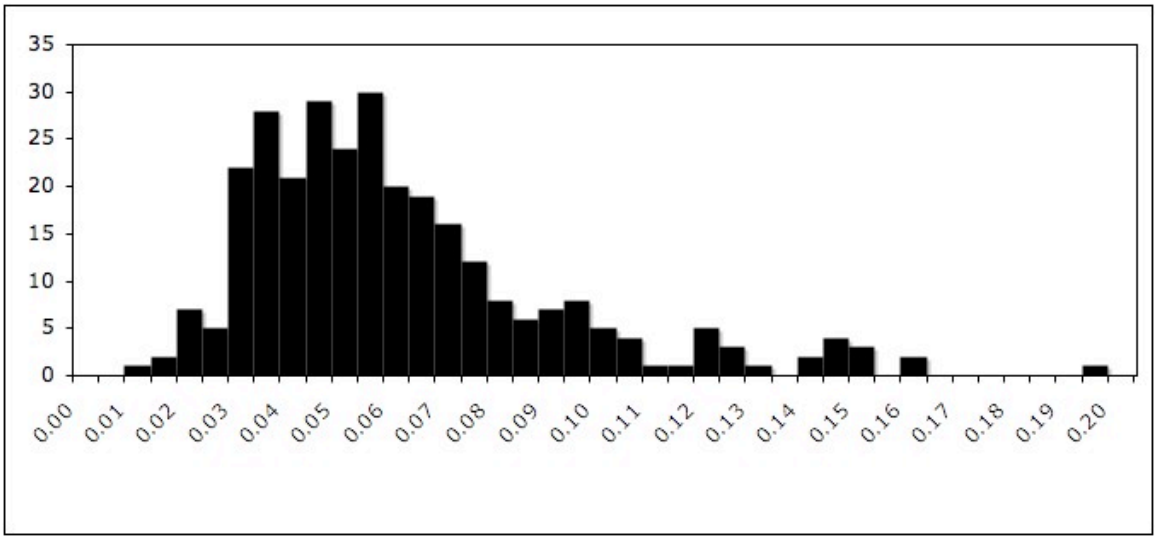
Sampling: Individual Big Bluestem plants were sampled from five distinct BCPOS and OSMP properties. The overall sampling scheme was intended to measure both between *and* within population differences. Rabbit Mountain (BCPOS) and Jewel Mountain (OSMP) were our sites with the most distance separating them. Jewel Mountain also represents the oldest soil surface in the Front Range of Colorado, and harbors a unique mixed community of plains and montane plant species.[8] The Burke property (OSMP) was clearly the wettest soil, and had dense patches of very tall plants. The Mayhoffer property (BCPOS) appeared the driest, and Betasso preserve (BCPOS) was the highest elevation.

Analysis: DNA was extracted from the leaf tissue using a bead beating, phenol/ethanol precipitation protocol from the MoBio Power Plant kit. DNA extractions were tested and standardized using a Nanodrop Photospectrometer as preparation for polymerase chain reaction based amplified fragment length polymorphism analysis (AFLP-PCR).[9,10] Initial runs tested selective primer sets, and informed our choice of the most informative AFLP pattern, with the least chance for error. After the second selective PCR amplification, samples with different fluorescently labeled primers were pooled and sent to the Nevada Genomics Center for fragment size separation on their ABI Prism 3730 capillary polyacrylimide gel analyzer. GeneMapper 4.1 software was

used to call fragment presence or absence automatically, but manual evaluation was used where necessary. Settings were tuned to help call the most number of fragments in hopes of improving downstream population resolution.[11] Primer sets that generate up to 808 clean fragments (markers), were then each used on 47 samples in order to explore overall population diversity.

To quantify the overall local genetic diversity, the Phylip v 3.6 package was used to calculate all pair wise AFLP profile distances using a modified Nei and Li formula and Jukes-Cantor model of evolution.[12] The resulting distribution demonstrates an overall local heterozygosity, and that larger difference outliers may hint at some interesting larger scale gene flow dynamics, which cannot be fully understood under the current sampling scheme.

Figure 1



Histogram of corrected pair wise distances between all 47 AFLP profiles containing 808 markers. Mean=0.063, SD=0.03, Minimum=0.011, Maximum=0.20

The Bayesian clustering program, Structure v 2.3, was next used to test series of population scenarios. The term population (or species) is a very difficult term to define in a way that actually reflects biological reality, and subjective measures of phenotypic differentiation are often used to rank relatedness between individuals. Unfortunately, traditional systems of taxonomy and classification very often misrepresent true evolutionary relationships. For example, Aristotle clustered organisms by the number of legs used—a seemingly natural division at the time, but one that mistakenly groups humans with chickens to the exclusion of mice and cat. Or more recently, scientific attempts at classification of North American leopard frogs based on detailed morphological traits, failed to identify many cryptic species that superficially appeared the same, but had actually evolved many unique adaptations to their respective environments.[13] Using the genotypic profiles generated by the AFLP runs, the Structure model characterizes each putative population (K) by specific allelic (marker) frequencies and assumes Hardy-Weinberg equilibrium and complete loci linkage equilibrium.[14] In this way we can more accurately estimate genetic relatedness than can be done using potentially misleading morphological traits. Ln probabilities for (X|K) for values for K (categories or number of assumed populations) of 1-5 were then calculated and repeated five times each. Averages for each K-value were calculated and tested for significance of variance (ANOVA). A significant difference was found between the different K-values ($P=5.28x10^{-13}$), and K=4 was found to have the highest probability. However, we must proceed with caution when interpreting these results in a biologically meaningful way.

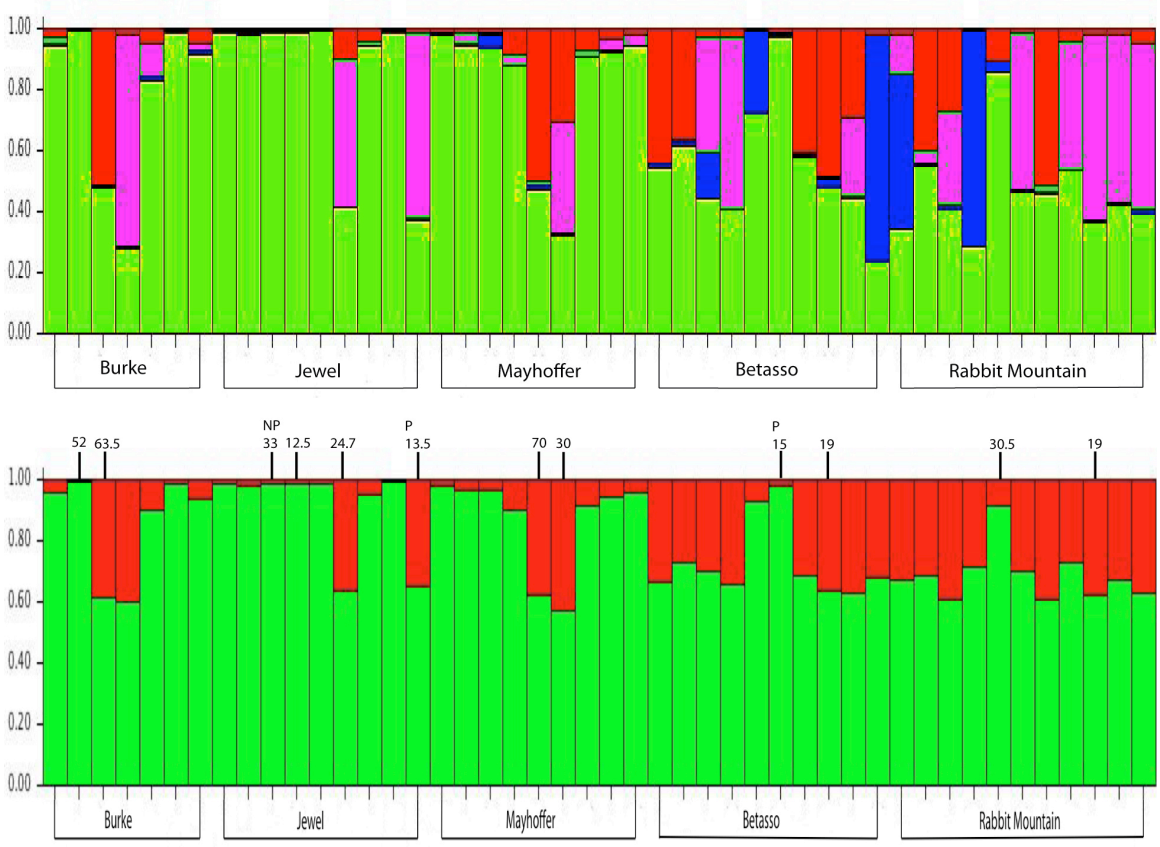


Figure 2

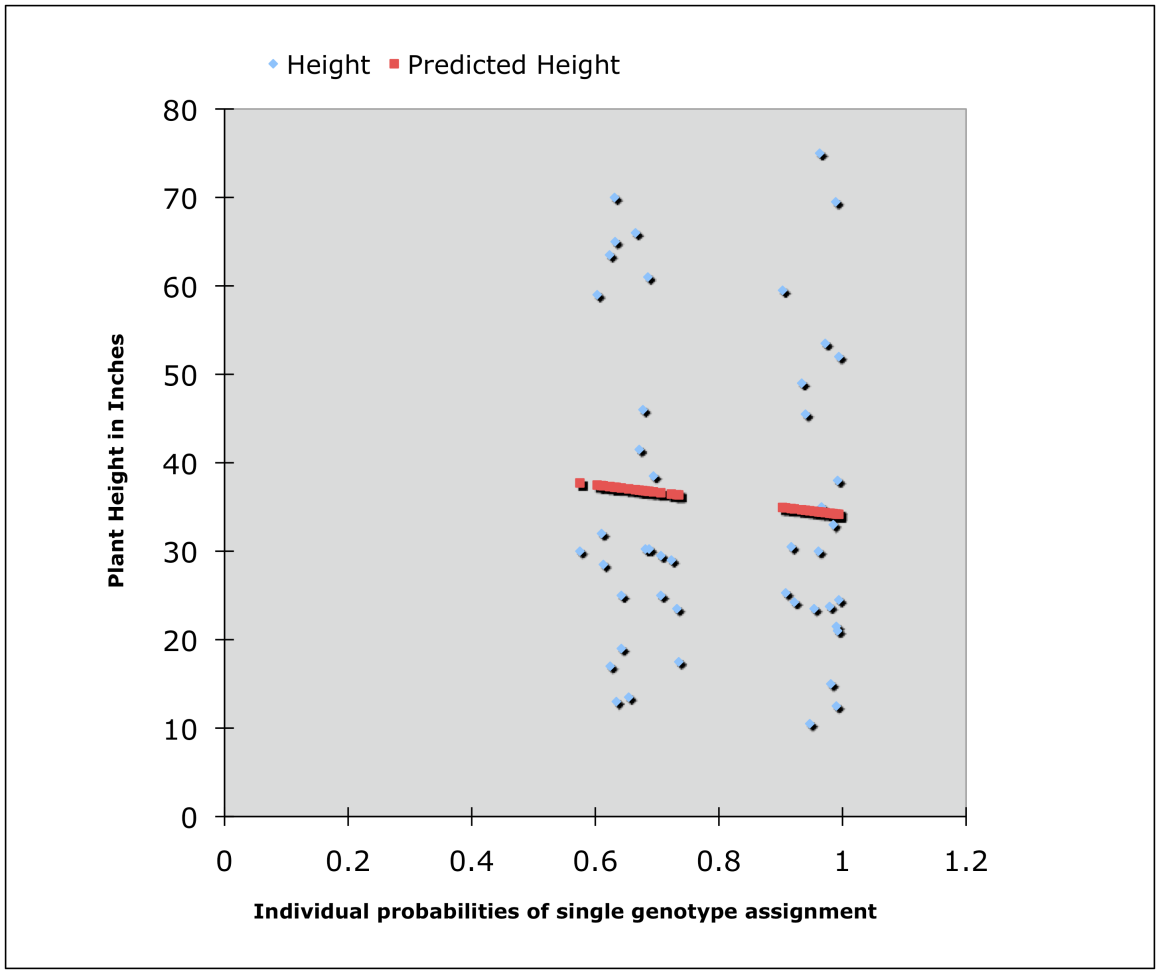
Comparison of K=2 on the bottom (testing the hypothetical distribution of putative mesic and xeric populations) vs. the most probable K=4 (top) population structure for big bluestem samples with a range of plant heights from 10.5 inches to 75 inches from both mesic and xeric sites. Vertical bars represent individual plants, ordered by collection location, while each color stands for an attempted genotype assignment. The Burke individuals were overall the largest (mean=60.57 inches vs all others mean=31.32), and the site soil probably the wettest. Select plant heights are shown across the top of the K=2 set in inches, while clear cases of pubescence (P) or non-pubescence (NP) are also indicated.

Discussion:

When looking at how Structure attempted to assign actual genotypes per sample there are several key points to keep in mind. First, there is no clear delineation of genotypes between any of the sample sites. If K really equaled two or four, and one of the population genotypes encodes for larger plants that only grew in the mesic sites, then you would expect there would all one color of bars for the Burke samples and all other

colored bars for the all the other smaller stature samples. This is not the case at all, each site has a mix of both all green genotype, and the mixed genotypes (two or more colors per bar). There appears to be no association between genotype assignment and plant height or leaf pubescence, and a formal regression analysis between genotype assignment probability as a predictor of plant height only reinforces this (Fig. 3). It is also important to keep in mind that the Structure model is not well suited to data in which allele frequencies vary gradually across geographic distances. With our sampling scheme limited to Boulder County, it is not yet possible to determine if such a isolation by distance trend exists, and across what geographic scale it acts.

Figure 3



Linear regression analysis of genotype assignment (K=2 scenerio) vs. plant heights. For the F-test, P=0.625, so there is obviously no significant prediction of plant height based on the Structure genotype assignment.

By analyzing allele frequencies, we have clearly shown that much larger stature big bluestem plants from mesic environments are not genotypically differentiated from smaller stature big bluestem populations from xeric environments. Although AFLP fragments are not technically alleles in the traditional use of the term, they represent a

sample of both coding and non-coding genomic regions, some of both of which would be expected to change through a speciation event. A more likely explanation for the wide range in plant sizes is the well established effects of differential water and nitrogen availability.[15,16,17] Big bluestem being a perennial, plant age could certainly play a role in observed phenotypic differentiation. This is not to say that every plant analyzed is even close to being the same genetically, only that no set of alleles could be identified as predicting plant growth or location. One possible scenario that could result in the mixed genotypes, is that Boulder County receives pollen and or seeds from the truly structured subpopulations of big bluestem, which lie outside of our sampling range. All of these genetic varieties mix randomly in Boulder and appear to thrive in all conditions across which we sampled. Only by expanding the geographic area and range of environmental conditions from which we sample big bluestem, would we start to see this possible true isolation and differentiation of sub-populations. Of particular interest is the pubescence that appears on the leaf surfaces of some plants locally. Does this trait originate from genes from higher elevation or more Northern populations? Is there a temperature or moisture threshold beyond which non-pubescent plants are not able to survive? How far and in what pattern does pollen and seed effectively travel across the broader big bluestem range?

In any case, the big bluestem of Boulder County that were sampled in this study appears to be a mixing pot of genetic diversity, and due to wind pollination, do not suffer from gene pool limitations which can adversely effect fragmented populations of other organisms. That these plants can appear so different in the field is actually not surprising whatsoever. Just consider how a bonsai tree can, through careful alteration of

environmental variables, grow to but a fraction of its relative's stature. Yet if the seeds or cuttings of said bonsai were to be planted back into a more favorable habitat, they would be again able gain many many times the size of the bonsai. Some plants, it seems, have a large capacity for phenotypic plasticity.

When considering restoration strategies, it would seem natural to attempt seeding efforts with an equal diversity of genes that are already found in BCPOS properties. By using the largest starting pool of wild collected seeds possible, regardless of collection site, you can maximally feel confident in the long-term survivability of propagated individuals.

---

[1] City of Boulder Open Space and Mountain Parks Grassland Ecosystem Management Plan, Public Review Draft, October 2008

[2] Gustafson DJ et al, Raising Local Seeds in Prairie Restoration, NATIVE PLANTS, 2005

[3] Keeler KH, Impact of Intraspecific Polyploidy in Andropogon gerardii Populations, Am. Midl. Nat. 152:63–74, 2004

[4] Norrmann GA et al, Evolutionary Implication of Meiotic Chromosome Behavior, Reproductive Biology, and Hybridization in 6x and 9x Cytotypes of Andropogon Gerardii, American Journal of Botany 84(2): 201–207, 1997

[5] Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey: structure, evolution, and application in breeding. *Theoretical and Applied Genetics*, **77**, 421–455

[6] Hamrick JL, Godt MJW (1997) Effects of life history traits on genetic diversity in plant species. In: *Plant Life Histories* (eds Silvertown J, Franco M, Harper JL), pp. 102–118. Cambridge University Press

[7] Warming E, (1909) Ecology of Plants: An Introduction to the Study of Plant Communities

[8] Report of Findings Five-Year Study of the Ecology of Bluestem-dominated Grasslands of the Rocky Flats Area 1996-2001 Jefferson and Boulder Counties, CO

[9] Vos P et al, AFLP: A New Technique for DNA Fingerprinting, Nucleic Acids Research, 1995, Vol. 23, No. 21 4407-4414

[10] Gustafson DJ et al, Random amplified polymorphic DNA variation among remnant big bluestem populations from Arkansas' Grand Prairie, Molecular Ecology 8, 1693–1701, 1999

[11] Holland BR, Clarke AC, Meudth HM (2008) Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. Syst Biol 57:347–366

[12] Felsenstein J, (2005) PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle*

[13] Hillis, DM, Wilcox TP, (2005) Phylogeny of the New World true frogs (*Rana*)

[14] Pritchard JK, et al, (2000a). Inference of population structure using multilocus genotype data. Genetics, 155:945–959

[15] Hsiao TC, (1973) Plant responses to water stress. Ann. Rev. Pl. Physiol. 24. 519-70.

[16] Etherington JR, Rutter AJ, (1964) *Journal of Ecology*, Vol. 52, No. 3, pp. 677-689

[17] Bradshaw AD, et al, (1964) *Journal of Ecology*, Vol. 52, No. 3, pp. 665-67