Title:

Assessing the remarkable morphological diversity and transcriptomic basis of leaf shape in *Ipomoea batatas* (sweetpotato)

Authors:

Sonal Gupta[1], David M. Rosenthal[2], John R. Stinchcombe[3], Regina S. Baucom[1]

Addresses

1. Ecology and Evolutionary Biology Department, University of Michigan, Ann Arbor, MI

2. Department of Environmental and Plant Biology, Ohio University, Athens, Ohio.

3. Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

Abstract:

Leaf shape is a spectacularly diverse trait that influences various aspects of plant physiology, and is even correlated with crop yield and quality in multiple species. However, a broad genotype-to-phenotype analysis of leaf shape, and subsequent identification of genes controlling leaf shape traits is limited to only a handful of species. Here, we perform an initial characterization of leaf shape variation in *Ipomoea batatas*, the sweetpotato, at multiple scales of analysis. We examine leaf shape variation across many diverse, greenhouse-grown accessions using a variety of morphometric analyses, and complement this with a transcriptomic survey to identify gene expression changes associated with two commonly studied leaf shape traits--circularity and aspect ratio. Additionally, we examine the potential that leaf shape in this species is genetically controlled and/or influenced by the environment by performing a field study with replicate clones of 74 accessions grown in two geographically separate common gardens. We comprehensively describe the remarkable morphological diversity in leaf shape in sweetpotato, and identify 147 differentially regulated genes associated with circularity and aspect ratio, providing an initial set of hypotheses regarding the genetic basis of leaf shape in this species. Interestingly, this work shows important differences in the factors controlling leaf shape between sweetpotato and other species: we found little evidence for a role of homeobox domain genes in the production of leaf shape traits, as identified in *C. hirsuta*, *Capsella*, *Arabidopsis* and tomato, and likewise find that leaf shape traits are primarily under genetic control and not strongly influenced by the environment or by gene by environment interactions.

Keywords:

Leaf shape, sweetpotato, digital morphometrics, EFD, transcriptomics, genetic architecture, GxE

INTRODUCTION

After the initial stages of plant domestication, where a wild species is modified through artificial selection to generate a species suitable for human use (Meyer & Purugganan, 2013), crop germplasm is further enhanced during a phase typically referred to as 'crop improvement' (Simmonds, 1976). Crop improvement, also known as crop diversification, involves the deliberate breeding of traits that increase yield and/or crop health in a range of environments as well as the breeding of traits that fit human preferences (Meyer & Purugganan, 2013). Crop improvement allows for the optimization of traits that lead to high productivity, and as such, knowledge of the genetics underlying these traits, along with an understanding of trait-climate interactions is crucial for the production of superior cultivars. Because crop improvement includes modifications that increase yield, abiotic and biotic stress tolerances, and improved nutritional quality, among other alterations, there are a wide variety of traits that can be involved in crop improvement (Meyer & Purugganan, 2013).

Leaf shape is one such potential crop improvement trait (Mathan et al., 2016). Leaf shape varies spectacularly across genera, species, and even varies among genetic lineages of the same species (Klein et al., 2017). Leaves can vary in their degree of lobing, length-to-width ratio, venation patterning, prominence of tips and petiolar sinus, or any combinations of the above. Leaf shape influences vital physiological properties, such as thermoregulation (Gurevitch, 1988), water relations (Wyatt & Antonovics, 1981), and defense against fungal enemies and herbivores (Brown et al., 1991). More specifically, leaves that are lobed reduce heat and water stress due to thinner boundary layers (Schuepp, 1993); multiple studies have found that shade leaves are less dissected than sun leaves (Vogel 1968, Jones 1995, Anderson 1991).

Importantly, leaf shape is correlated with crop yield traits. In tomato (Chitwood *et al.* 2013) and cotton (Andres *et al.*, 2017), an increase in leaf complexity is positively correlated with expression of genes involved in photosynthesis. Studies have also found an association between leaf shape and both photosynthetic efficiency and fruit sugar content (Chitwood et al., 2013; Vuolo et al., 2016). Thus, leaf shape, by influencing photosynthetic rate, could increase crop yields or influence crop quality (Long et al., 2006; Zhu et al., 2010). Although leaf shape attributes can be modified through selective breeding, identifying the genes underlying leaf shape phenotypes has the potential to more readily allow for the creation of superior cultivars through transgenic approaches. Identifying leaf shape genes can also address important questions about leaf development. However, to date, genes associated with leaf shape have been identified in only in a few species (*Arabidopsis*, tomato, barley, cotton; Kim et al., 2002; Kimura et al., 2008; Vlad et al., 2014; Jost et al., 2016; Andres et al., 2017).

Sweetpotato, *Ipomoea batatas*, is one of the most widely cultivated staple root crops worldwide, as it is grown on at least eight million hectares in 114 countries (Khoury et al., 2015). Domesticated at least 5000 years ago in Central or South America (Huaman, 1987), this species produces the highest amount of edible energy per hectare (de Vries et al., 1967) and is used both as a root crop and as forage for livestock (Murugan et al., 2012). It is of high importance in developing countries such as the sub-Saharan Africa for its high levels of vitamin A and C as well as calcium and iron (Kays and Kays, 1998). Sweetpotato exhibits high morphological variation across its ~6000 documented varieties (Huaman, 1987), and this includes a remarkable array of leaf shapes (Huaman, 1987). For example, leaves of this species can vary from being

highly dissected and lobed, with varying degree of lobing--ranging from 3 to 7 lobes--to being the more typical heart-shape (or chordate) form. Only a limited number of studies have examined the variation in leaf shape in this species (Huaman, 1987, Hue et al., 2012), and these studies focused primarily on the degree of lobing and leaf length and width respectively. The extensive range of leaf phenotypes in *I. batatas*, along with its association with carbon assimilation, warrants an investigation of the potential that leaf shape can be a useful crop improvement trait in this species.

Here, and as a first step towards our overarching goal of assessing leaf shape as potential crop improvement trait, we present a broad analysis of leaf shape variation in sweetpotato. For a trait to be a crop improvement trait, there should be evidence of phenotypic variation among cultivated varieties, this phenotypic variation should be genetically controlled, and there should be evidence of a functional relationship between aspects of the trait and crop yield or quality. In this work, we examine the first two conditions. We first show the remarkable diversity of leaf shape variation in *I. batatas* by accessing common morphometric measures--circularity, aspect ratio, solidity--and by performing an Elliptical Fourier Descriptor (EFD) analysis, which provides a broader perspective of variation due to traits typically difficult to define and measure. We then survey gene expression differences associated with leaf shape types to identify potential genes associated with leaf shape diversity. Finally, we determine the extent to which leaf shape traits are genetically controlled and/or are influenced by the environment using a common garden study of 74 sweetpotato accessions in two variable common gardens. Specifically, we asked: (1) Do leaf shape traits vary between cultivars and if so, what aspect of shape influences this variation the most? (2) what are the gene transcription differences associated with two well-known leaf shape traits--circularity and aspect ratio? and (3) what is the estimated heritability of leaf shape traits, and does GxE play a significant role in altering these traits? Our results help identify potential crop improvement leaf shape traits in sweetpotato, which could then be focused on in future investigations for the identification of functional relationship between the trait and yield.

METHODS

Leaf shape variation in *I. batatas*

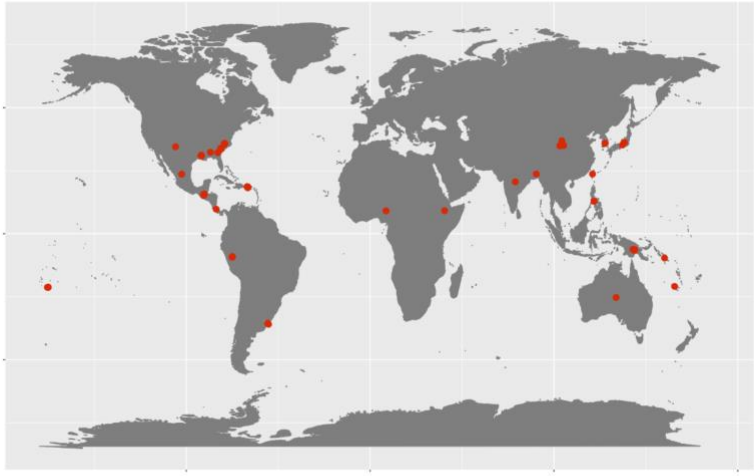

Figure 1. Geographic diversity of the 74 chosen sweetpotato, *Ipomoea batatas,* accessions. Red dots represent the origin of the chosen samples.

Vegetative slips for 74 publicly available accessions of sweet potato were ordered from USDA (https://npgsweb.ars-grin.gov/gringlobal/search.aspx?). These accessions represent a world-wide sample; the location of origin of 74 accessions is shown in Figure 1. The slips were grown at the UM Matthaei Botanical Garden under standardized growth conditions (16 hrs light/8 hrs night cycle) for approximately six months, at which time we sampled 4-6 leaves

of 59 accessions and scanned them for leaf shape analysis. These images were then used to extract leaf shape traits values using the custom macros of ImageJ (Abràmoff et al., 2004). Briefly, the leaves were converted to binary images and the outlines from these binary images were used to measure circularity, aspect ratio and solidity.

We elected to focus on these classic shape descriptors because they each capture a distinct aspect of leaf shape (Li et al., 2018). Circularity, measured as $4\pi\frac{area}{perimeter^2}$, is the ratio of area to true perimeter of the leaf (excluding holes) and is thus influenced by serrations and lobing and elongated shapes; higher serrations/lobing leads to higher perimeter and thus lowers circularity. Aspect ratio, in comparison, is measured as the ratio of the major axis to the minor axis of the best fitted ellipse, and is influenced by the length and width of the leaf. Lastly, solidity measured as $\frac{area}{convex\ hull}$, is sensitive to leaves with deep lobes, or with a distinct petiole and can be used to distinguish leaves lacking such structures. Solidity, unlike circularity, is not very sensitive to serrations and minor lobings, since the convex hull is remains largely unaffected.

For a more global analysis of leaf shape via Elliptical Fourier Descriptor (EFDs), we used the program SHAPE (Iwata and Ukai, 2002). EFDs capture the variation in shape represented by the outline, which is difficult to categorize via the classical shape descriptors. “ChainCoder” was first used to extract contours of leaf as chain code which was then used as input for “Chc2Nef”to calculate normalized EFDs. Normalization was based on the longest radius which allows for manual orientation with respect to the proximal-distal axis of the leaf. EFDs resulting from only the first 20 harmonics of Fourier coefficients were used for PCA using “PrinComp”. For each PC, coefficients of EFD were calculated at -2 and +2 SD (standard deviation) and the respective contour shapes were reconstructed from the coefficients by inverse Fourier transformation.

RNA-Seq library construction and sequencing

We sequenced and analyzed the transcriptomes of 19 individuals of *I. batatas* to examine the transcriptomic basis of leaf shape variation in this species and obtain an initial set of candidate genes that influence leaf shape. We selected greenhouse-grown accessions displaying very high and low lobedness, and likewise chose accessions exhibiting high/low aspect ratio. Eleven individuals were used to assess variation in circularity (six entire (566613, 585054, 585108, 592993, 595868, 612701), five lobed (573318, 585061, 585067, 599387, 634379) and 9 individuals were chosen to assess variation in aspect ratio (five high AR (208886, 344123, 566617, 566638, 585061), four low AR (318848, 531122, 564154, 599392)); one accession (585061) was common for circularity and AR. Individual accessions were randomly arrayed on greenhouse benches, and variation in gene expression due to environment was reduced by both treating all individuals to the same standardized growth conditions and by moving pots around benches randomly throughout the growing period.

We used three to five of the smallest leaves (<1 cm in length) from multiple branches of each individual accession for RNA extractions, and combined replicate leaves per individual to increase the coverage of the transcriptome. We sampled all individuals on the same day and within 1 hour of one another to reduce variation due to developmental stage and/or time of collection. We froze samples in liquid nitrogen prior to preserving them at -80° for further processing. We performed RNA extraction using Qiagen RNeasy Plant mini kit with the optional DNase digestion step, and constructed libraries using the TruSeq Stranded mRNA Sample

Preparation protocol (LS protocol). After barcoding, we bulked all libraries and performed one lane of Illumina HiSeq2500 sequencing.

RNA-Seq data processing and transcriptome analysis

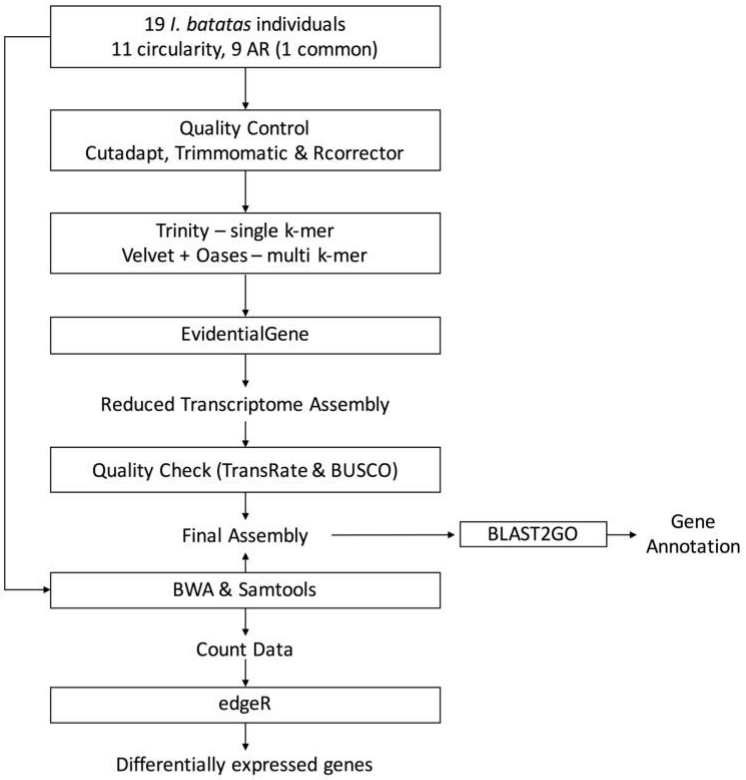

Figure 2. Methodology for RNA-Seq data processing for differential gene expression

An overview of RNA-Seq data processing and transcriptome analysis is given in Figure 2. Briefly, we performed quality control for the obtained raw reads to trim the adaptors, discard low quality reads and eliminate poor-quality bases. We used cutadapt (Martin, 2011) to remove the adapters, and trimmomatic (Bolger et al., 2014) (-phred33 SLIDINGWINDOW:4:20 LEADING:30 TRAILING:30 MINLEN:25) to clean the reads based on length and quality score. Further, we performed error correction of the RNA-Seq data using rcorrecter (Song and Florea, 2015) to retain only high quality data.

Next, we used filtered reads separately from one entire and one lobed individual, randomly chosen, for *de novo* transcriptome assembly, which served as a reference transcriptome for differential analysis. To get a comprehensive assembly, we used both a single k-mer approach, using Trinity (Grabherr et al., 2011), with k=25, and multi k-mer approach, using Velvet/Oases (Zerbino and Birney, 2008; Schulz et al., 2012) with k-mer ranging from 23-41 and 93-99 with a step-size of 2. Next, we used the EvidentialGene tr2aacds pipeline (http://arthropods.eugenes.org/EvidentialGene/trassembly.html) to merge the all the assemblies to remove redundancy and to get biologically most useful set of transcripts.

We then evaluated the obtained set of primary transcripts using TransRate (Smith et al., 2016) and BUSCO - Benchmarking Universal Single-Copy Orthologs (Simão et al., 2015), which reports basic summary statistics (like n50, % reads mapped, etc.) and checks for the completeness of the transcriptome respectively. For annotation of this *de novo* assembled transcriptome, we blasted the transcripts against the NR database with an e-value threshold of 10^{-6} and other default parameters and used only the top 20 hits for annotation. Additionally, we identified conserved protein domains by searching through the InterPro collection of databases. We used the results from these to functionally annotate using BLAST2GO v4.1.9 (Conesa et al., 2005) by identification of Gene Ontology (GO) Slim terms and KEGG pathways.

Differential gene expression--We mapped reads from all 19 individuals to the assembled transcriptome using BWA-MEM (Li, 2013) and estimated the read counts of only the uniquely mapping samtools (samtools view -F 0x904 -f 0x40 -b ; samtools idxstats; Li et al., 2009). We then used these uniquely mapped read counts to filter out the lowly expressed genes using the Bioconductor package edgeR version 3.18.1 (Robinson et al., 2010) such that transcripts were retained only if they were expressed at a counts-per-million above 0.5 in at least two samples. We then normalized the libraries in edgeR (using the trimmed mean of *M*-values method) followed by differential gene expression analysis between lobed vs entire individuals and high aspect ratio vs low aspect ratio individuals using classic pairwise comparison of edgeR version 3.18.1. We extracted the significance of differentially expressed transcripts (DETs) with FDR <= 0.05.

Field experiment

We performed a field experiment to examine leaf shape traits in the field and to determine whether environment and GxE interactions influenced these traits. We generated replicate individuals for the field experiment by planting replicate 5 cm cuttings of the stem of each accession in 4-inch pots which were randomly positioned on a mist bench in the Matthaei Botanical Gardens. During the first week of June, we planted three to seven replicates of each of the 74 accessions in two common gardens in the field--one located at the Matthaei Botanical Gardens in Ann Arbor, MI, and the other at the Ohio University Student Farm, West State Street Research site in Athens, OH. Replicates were planted in either three (MI) or seven (OH) blocks in a completely randomized block design with 14-inch spacing between individuals. Blocks were kept relatively weed free but were otherwise allowed to grow undisturbed. We sampled 2-5 leaves from each individual the first week of October, prior to the first frost, and scanned them for ImageJ (Abràmoff et al., 2004) analysis of leaf shape traits – circularity, aspect ratio and solidity.

Data analysis--We first examined the potential for variation in leaf shape due to environmental differences (i.e. variation due to being grown in MI or OH) by performing an ANOVA. To normalize the leaf shape traits, we used the function TransformTukey from rcompanion version 2.0.0 (Mangiafico, 2018). TransformTukey is a power transformation based on Tukey's ladder of Powers, which loops through multiple powers and selects the one that normalizes the data most. These normalized leaf shape traits were then used as dependent variables and accession, garden, block effects and an interaction term of accession and garden as independent variables in the following model:

(Trait ~ Accession + garden + block + Accession:garden).

The term accession represents the genetic component, garden represents variation due to environment (plasticity), Accession:garden represents the GxE component and the block effect captures microenvironmental variation (and was nested within each garden). To quantify the effects of each of these variables on leaf shape, we calculated eta squared ($\eta 2$) as a measure of the magnitude of effect size using the Bioconductor package lsr version 0.5 (Navarro, 2013). Eta squared for an effect is measured as SS_{effect}/SS_{total}, where SS_{effect} is the sum of the squares of the effect of interest and SS_{total} is the total sum of the squares of all the effects including interactions. In other words, it is a measure of the proportion of variance in the dependent variable associated with independent variable and is one of the most commonly reported estimate of effect size for

ANOVA (Levine and Hullett, 2002; Ialongo, 2016). Further, we calculated broad sense heritabilities of leaf shape traits to determine the extent to which leaf shape traits are genetically controlled in each environment. Broad sense heritability was calculated using linear mixed modeling with the Bioconductor package sommer version 3.4 (Covarrubias-Pazaran, 2016) based on the phenotypic data collected from the two fields separately. The model used was

Trait~1, random=~Accession + block + Accession:block, rcov= ~units

Variance components thus obtained were plugged in the equation below to calculate the broad-sense heritability (H^2) using the formula:

$$H^2 = \frac{V_g}{V_g + V_e + V_{gxe} + V_r}$$

where, V_g is the genotype variance, V_e is the environmental variance due to the blocks, V_{gxe} is the variance associated with V_{gxe}(accession:block) and V_r is the residual variance.

RESULTS

Leaf shape variation among accessions

Although sweetpotato exhibits vast morphological variation, studies quantifying this variation are still lacking. We sought to globally quantify leaf shape variation in a wide range of samples (Figure 1) by first utilizing the classic shape descriptors--circularity, aspect ratio and solidity--and by further performing an Elliptical Fourier Descriptor (EFD) analysis on leaf outlines.

We found wide variation in the leaf traits across 59 *I. batatas* accessions (Table 1). Among the three traits examined, circularity is the most variable with PCV (Phenotypic Coefficient of Variation) of 22.61% while aspect ratio is least variable with a narrow distribution and PCV of 4.76%.

Trait	**Range**	**Mean**	**SD**	**PCV (%)**
Circularity	0.09-0.71	0.50	0.12	22.61
Aspect Ratio	1.03-1.26	1.10	0.05	4.76
Solidity	0.44-0.95	0.84	0.10	11.85

Table 1: Leaf shape trait values across the 59 chosen sweetpotato accessions. SD represents standard deviation while PCV represents phenotypic coefficient of variation.

Figure 3 shows the phenotypic diversity with respect to two leaf traits, circularity and aspect ratio. Of our 59 accessions, 10 exhibit low circularity (defined as circularity < 0.50). PI 599387, for example, exhibited leaves that are very deeply lobed and thus has a low circularity (0.09) value. In contrast, PI 566647 has no serrations or lobing (entire margins) and thus exhibits

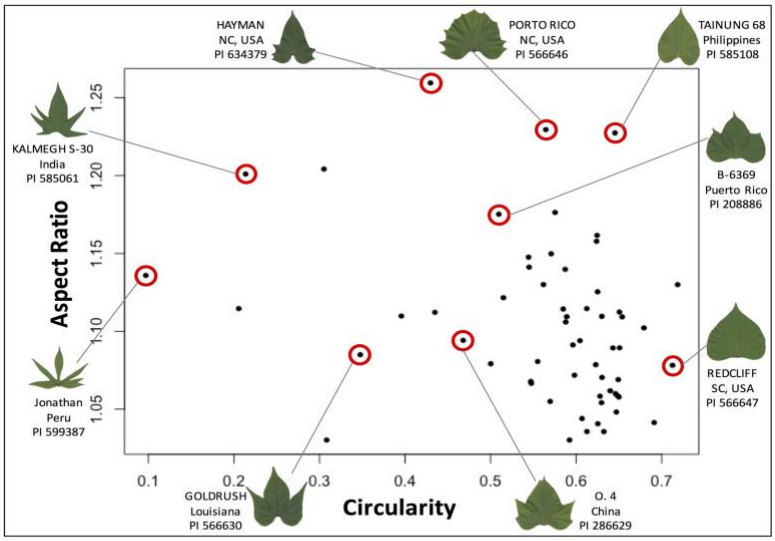

Figure 3. Leaf shape variation in a sample of accessions of sweetpotato, *Ipomoea batatas*, highlighting exceptionally high morphological variation.

high circularity (0.71; Figure 2b). Additionally, we found 22 of the 59 accessions to exhibit high aspect ratio (defined as aspect ratio >= 1.11). For example, PI 531134 (AR = 1.03) has almost equal values of major and minor axis and thus a low aspect ratio value. In contrast, the leaves of PI 208886 (AR = 1.268) are much wider, i.e., a larger major to minor axis, and thus has high aspect ratio value. Most often this increase in AR in sweetpotato manifests itself with increase leaf width (eg. PI 566646, PI 208886) relative to length (eg. PI 634379). Further, although solidity values range from 0.44-0.95, only 5 accessions had solidity values less than 0.7, potentially overestimating the PCV to 11.85%. The lack of low solidity values indicates that only a few accessions have deeply lobed leaves (eg. PI 599387, Solidity = 0.44), in contrast to accessions with slightly lobed leaved (eg. PI 566630, Solidity = 0.76).

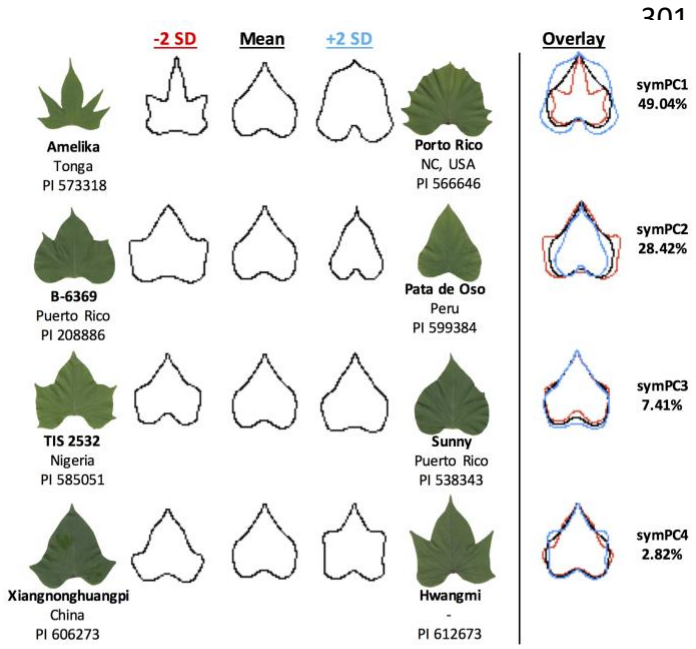

Figure 4: EFDs of symmetrical shape variation. Contours represent eigenleaves resulting from PCA on symmetrical shape (symPC) on EFDs. Shown are the first four PCs with the percent variation explained by each; 87.79% of total variation is explained. -2SD (red) and + 2SD (blue) represent two units of standard deviation from the mean along the PC. Representative leaves of accessions with extreme PC values are shown.

To get a more global estimation of leaf shape variation, not represented by the classical shape descriptors, we performed an EFD analysis on leaf outlines (Figure 4). EFD analysis provides a unique opportunity to describe leaf shape qualities that are traditionally difficult to quantify (Chitwood et al., 2014). In total, we processed 303 leaves from 59 accessions to identify leaf shape qualities that explain symmetrical shape variation in sweetpotato. Low symPC1 values describe leaves with deep lobing and prominent tip (PI 573318) whereas high symPC1 values explain non-lobed leaves with flattened leaf tips and enclosed petiolar sinus (PI 566646). Low symPC2 and low sym PC3 values describe leaves with two lobes closer to the tip of the leaf, whereas low symPC5 values describes leaves with two lobes closer to the petioles. In

contrast, high symPC2 and high symPC3 values describe non-lobed leaves; PC3 additionally describes variance associated with shape and prominence of petiolar sinus. The four PC components together explain 87.79% of total variance relating to symmetrical leaf shape variance in sweetpotato.

Sequencing and *de novo* assembly of *I. batatas* transcriptome

We observed extensive phenotypic variation in leaf shape in sweetpotato as demonstrated above using classical shape descriptors and EFDs. Next, to identify the gene expression changes associated with leaf shape in sweetpotato, we elected to focus on circularity and aspect ratio, which are leaf shape traits that have been shown to be associated with crop quality in other species. For our analyses of the transcriptome, Illumina HiSeq2500 returned a total of 266 million (125bp) paired-end sequence reads; on average, each individual had 14 million (M) reads. After cleaning and filtering this was reduced to 233 million reads. We assembled transcriptomes based on multi k-mers for one lobed and one entire type, randomly chosen, with Trinity (Grabherr et al., 2011) and Velvet/Oases (Zerbino and Birney, 2008; Schulz et al., 2012). We then pooled these assemblies together and removed redundant transcripts using the EvidentialGene pipeline http://arthropods.eugenes.org/EvidentialGene/trassembly.html) which returned a total of 34,450 high confidence transcripts. While the minimum length of the transcript was 107 bases, only a fraction of the transcripts (13.65%) were <300 in length. Additionally, 76.53% of the reads from the chosen lobed and entire type mapped back to the assembled transcriptome. The sequence statistics of the reduced transcriptome assembly is presented in Table 2.

Number of transcripts	Min Len (base)	Max Len (base)	Number of bases	Mean Len	ORF percent	n50	% reads mapped
34,540	107	16,428	35,930,680	1,035	79.95%	1,614	76.53%

Table 2: Sequence statistics of the reference transcriptome obtained from EvidentialGene pipeline.

To assess the completeness of the assembly, gene content from single-copy orthologs were checked against the transcriptome through BUSCO with eukaryotic plant dataset (Simão et al., 2015). These genes have evolved under single-copy control and thus a high percent of duplicates may indicate erroneous assembly. The results indicate that the *de novo* transcriptome assembly is of high quality with 91.32% (1315/1440) complete genes found (single copy genes~87%) of which a mere 4.51% were duplicates. Additionally, only 6.32% of genes were missing from the assembled transcriptome. Thus, our sequencing and assembling strategy produced a relatively complete transcriptome. Using blastx, 24,565 transcripts were annotated by the functional description of their top 20 hits. Plant transcriptomes, without genome information, have been annotated with varying percentages, ranging from 13.03% (Wang et al., 2010a) to 83.8% (Ness at al., 2011). Thus, our functional annotation of 71.1%, provides sufficient information for further analysis.

Identification and functional annotation of differentially expressed transcripts (DETs)

We found 47 gene transcripts to be differentially expressed in the comparison of lobed vs circular-leaf individuals (Figure 5a) and 100 gene transcripts to be differentially expressed when contrasting individuals with low vs high aspect ratio (Figure 5b). Most of the DETs had high differences in expressions; only 3 and 10 DETs had fold-change of less than 2 for circularity and AR respectively. On an average, we found that 11 million unique paired-end reads (range 7.66M - 14.23M) mapped back from all 19 individuals to the reference transcriptome (net mapping efficiency of 89.65% with the paired-end high-quality reads). This indicates that we had sufficient read depth (>10M) to continue with our differential expression analysis as has been shown by Wang et al., 2011.

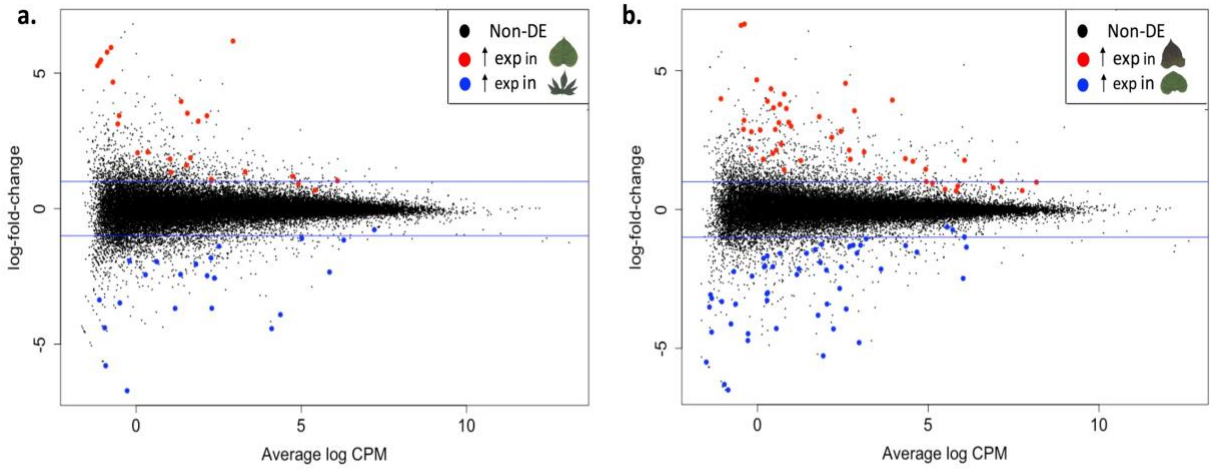

Figure 5: Plot of log-fold change against log CPM (counts per million) with differentially expressed transcripts highlighted (red and blue dots). **a**. Red and blue dots represent transcripts with higher expression in entire and lobed respectively. **b**. Red and blue dots represent higher expression in high aspect ratio and low aspect ratio individuals respectively.

Of the 47 DETs for circularity, 25 were upregulated in the entire type individuals as compared to the lobed type individuals whereas 22 were down-regulated. Among the differentially transcribed transcripts, we were able to annotate fifteen upregulated and four down-regulated transcripts (after discounting the "uncharacterized protein" annotations). We used the functional annotations of these transcripts to identify putative leaf shape genes (Table 3). The upregulated transcripts had some interesting functional annotations (all with P-value < 0.001). For example, FAR1-related sequence 5 (or FRS5), with log fold-change of 5.77, is a putative transcription factor which is involved in regulating light control of development and is upregulated in *Arabidopsis* by far-red treatment (Lin and Wang, 2004). Other putative leaf shape genes are exocyst complex component EXO70A1-like and extra-large guanine nucleotide-binding protein which are involved in regulating cell proliferation and organ morphogenesis.

Among the transcripts differentially expressed for AR, 48 were upregulated in high aspect ratio individuals as compared to the low aspect ratio individuals and 52 were down-regulated. From the set of upregulated transcripts, 20 were annotated with meaningful biological information; this number was higher for down-regulated transcripts (24/52 transcripts were annotated). Interestingly, we saw a trend of multiple transposase and reverse transcriptase related

proteins (5 transcripts) being upregulated in highAR vs lowAR. Contrastingly, multiple ribosomal units were downregulated in highAR vs lowAR. Functional annotations of the remaining transcripts were used to identify putative aspect ratio gene in sweetpotato. Some interesting annotations belonged to set of down-regulated genes: protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10 (LSH10), Xyloglucan endotransglucosylase hydrolase protein 9 (XTH9) and protein SABRE. Interestingly, we also found two proteins involved in disease resistance, RPP13 and R1A-4, that were down-regulated in high AR individuals compared to low AR individuals. These DETs form a set of putative genes maintaining AR variation in leaf shape of sweetpotato (Table 3). We did not find any evidence of differential expression of a previously reported leaf width gene BLF1 in Barley (Jost et al., 2016) and ANGUSTIFOLIA in *Arabidopsis* (Kim et al., 2002).

Transcript ID	**LogFC**	**FDR**	**NCBI RefSeq ID**	**Gene Description**
Circularity				
trn23120	5.77	0.003	XP_019184633	FAR1-RELATED SEQUENCE
trn27953	2.08	0.021	XP_019191314	Exocyst complex component EXO70A1
trn24736	1.33	0.033	XP_019169432	Extra-large guanine nucleotide-binding protein
Aspect Ratio				
trn25740	-2.14	0	XP_019159391	Protein LIGHT-DEPENDENT SHORT HYPOCOTYLS 10
trn1645	-1.28	0.024	XP_019194051	Xyloglucan endotransglucosylase hydrolase protein 9
trn27852	-2.99	0.025	XP_019174234	Protein SABRE

Table 3: Candidate genes maintaining variation in leaf traits (circularity and AR) identified from the set of differentially expressed transcripts (DETs) in *Ipomoea batatas*.

Field experiment

We performed a field experiment to examine the variability of leaf shape traits in different environments. Additionally, we wanted to identify the underlying components controlling this variability (genotype vs environment vs GxE) and determine the extent to which these components alter leaf phenotype.

To characterize leaf shape differences between environments, among genotypes and the potential effect of GxE on leaf shape, we performed ANOVA (Table 4) as described in the methods. We found that circularity, aspect ratio and solidity significantly differ due to accession or genotype (F_{73}=18.06, F_{73}=4.22, F_{73}=21.09; P <0.001) and that accession explains 73.23%, 38.40% and 77.18% of the total variation respectively. This high variance explained is reflected in high heritability values reported below. Interestingly, we also found evidence of significant

block effect with low variation (F_8=3.01, P = 0.002; η^2 = 1.33%) on circularity. In contrast, aspect ratio and solidity were not significantly influenced by block effects.

Garden differences between OH and MI contributed 1.93% (F_1=15.55, P <0.001) of the variability in AR whereas accession by garden interaction explained 12.95% (a significant GxE effect: F_{69}=5.01, P = 0.009) of the variation. Moreover, circularity and solidity were not significantly altered by environment and thus had no significant differences due to GxE.

Variable	**df**	**Circularity**			**Aspect Ratio**			**Solidity**		
		F	**P**	**η^2(%)**	***F***	**P**	**η^2(%)**	***F***	**P**	**η^2(%)**
Accession	73	18.06	<0.001***	73.23	4.22	<0.001***	38.40	21.09	<0.001 ***	77.18
Garden	1	3.64	0.056	0.20	15.55	<0.001***	1.93	3.37	0.067	0.16
Block	8	3.01	0.002 **	1.33	1.38	0.020	1.38	1.94	0.052	0.70
GxE	69	1.30	0.06	5.01	1.50	0.009 **	12.95	1.30	0.065	0.40
Residuals	364	NA	NA	20.2	NA	NA	45.31	NA	NA	17.56

Table 4: ANOVA table of the below ground biomass model showing significant explanatory variables. df: degrees of freedom; *F*: value of F-statistic; P: p-value; η^2: eta-squared value.

Since much of the variation in leaf shape traits are explained by the accession effect, which indicates a genetic basis underlying these traits, we wanted to determine a heritability value for each trait. We thus estimated broad-sense heritability across environments and replicates. We found the leaf shape traits to exhibit high heritabilities. Circularity and solidity exhibited similar and high heritability ($H^2_{MI_cir}$= 0.79, $H^2_{OH_cir}$= 0.73; $H^2_{MI_solidity}$= 0.82, $H^2_{OH_solidity}$= 0.76) whereas aspect ratio exhibited lower heritability ($H^2_{MI_AR}$= 0.39, $H^2_{OH_AR}$= 0.26). Heritability values in OH garden were consistently lower than in MI garden due to reduced genetic variance and increased environmental variance. Overall, the high heritability values indicate that leaf morphology is controlled by genetic factors to a great extent.

Environment	**Circularity H^2 (SE)**	**Aspect Ratio H^2 (SE)**	**Solidity H^2 (SE)**
MI	0.79 (0.03)	0.39 (0.07)	0.82 (0.03)
OH	0.73 (0.04)	0.26 (0.08)	0.76 (0.04)

Table 5: Broad-sense heritability values for leaf shape traits in differing environments. The values in the parentheses represents the standard error (SE) associated with the calculations.

DISCUSSION:

In this study, we examined leaf shape variation among replicated accessions of sweetpotato in the greenhouse and across two common garden field experiments. We found evidence of extensive morphological variation, with shape differences due to lobing, length-to-width ratio of leaves and the prominence of tip and petiolar sinuses (Figure 3 and Figure 4). We combined this morphological assessment with an examination of the gene expression underlying leaf development, and identified potential candidate genes associated with two leaf shape traits--circularity and aspect ratio (Figure 5 and Table 3). Further, we examined these same leaf shape traits in two common gardens in the field to assess the environmental influence on each trait and determine the extent to which each is controlled by genetic versus environmental factors. From this we found leaf shape to have a strong genetic basis with most of the variation in phenotype attributed to accessional variation, with low or limited influence of GxE (Table 4 and Table 5). Below, we expand on each of our findings, and place them in the context of current knowledge about leaf shape development as well as what is known about the environmental influence on leaf shape in other species.

High morphological diversity in I. batatas *leaf shape*

Leaf shape is highly variable and shows a wide phenotypic continuum in the domesticated *I. batatas*. We focused our morphometric study on three classical shape descriptors--circularity, aspect ratio and solidity--each of which captures a distinct aspect of shape. Circularity, a measure of leaf dissection, was found to be the most variable leaf shape trait in *I. batatas*, ranging from 0.09 to 0.71. In comparison, aspect ratio, a measure of the length to width ratio of the leaf, was found to be the least variable trait (range of 1.03-1.26); solidity ranged from 0.44 to 0.95. These easy to measure shape descriptors have been found to be of evolutionary significance in other plant species. For example, in a comparison of domesticated tomato and its wild counterparts, the three traits were shown to have distinct values and occupy diverse regions in morphospace (Chitwood et al., 2014); interestingly, in tomato, aspect ratio explained over 40% of total shape variation (Chitwood et al., 2013). In apple, an increase in leaf size post-domestication is due to increased leaf width, and thus with reduced aspect ratio (Migicovsky et al., 2017). Furthermore, the three shape descriptors have also been shown to successfully differentiate and classify leaves between species (Li et al., 2018). Thus, the extensive variation in leaf shape found in sweetpotato might be indicative of the ongoing evolution associated with the diversification phase (or the crop improvement phase) of domestication.

Although important, the classical shape descriptors do not capture the entirety of shape variation. For example, leaf shape variation due to prominence of the tips and petiolar sinus, and even the variation associated with the positioning of the lobes is not accounted for by the three shape descriptors. However, variation in shape due to these traits can be quantified by tracing the leaf outlines. Thus, to capture leaf shape variation not accounted for by circularity, aspect ratio and solidity, we performed a more global shape analysis using leaf outlines *via* Elliptical Fourier Descriptor (EFDs). EFD is a morphometric technique that comprehensively quantifies the outlines of a shape and is normalized to differences in leaf size. EFDs have been used to quantify leaf shape variation in tomato (Chitwood et al., 2012), grape (Chitwood et al., 2014), apple (Migicovsky et al., 2017) and *Passiflora* (Chitwood and Otoni, 2017). In our analysis, the first two principal components of the EFD analysis together accounted for 77.46% of the total

variation in leaf shape, and described variation associated with petiolar sinus, tips and positioning of lobes. This variation would have been missed by simply quantifying classical shape descriptors, suggesting that the use of comprehensive morphometric techniques can help quantify the full extent of shape variation across species.

Gene transcripts underlying leaf shape variation

To further our understanding of gene expression changes involved in leaf development in sweetpotato, we sequenced the transcriptomes of 19 accessions and assembled a high-quality gene expression database for performing a differential expression analysis in *I. batatas*. We adopted an integrated *de novo* assembly approach using multiple k-mers and combined the assemblies to obtain a final set of non-redundant transcripts forming the reference transcriptome for expression analysis. Additionally, our reference transcriptome was sufficiently complete, with only 6.32% BUSCO single-copy genes missing from our transcriptome as compared to 30.6% and 27.6% missing from the two other assemblies previously reported in the literature (Tao et al., 2012, Yang et al., 2017).

From the analyses of differential expression, we found 47 genes that were differentially expressed between entire and lobed leaved individuals. Functional annotations of these genes identified potential candidates that could be responsible for maintaining leaf shape variation (Table 3). For example, homolog FAR1-related sequence is a putative transcription factor which is responsible for regulating development in response to light control. Thus, upregulation of FRS5 might be a response of the plant to promote survival under shade leading to less dissected more entire leaves. Similarly, trn27953 and trn24736, homologs of exocyst complex component EXO70A1 and extra-large guanine nucleotide-binding protein, respectively, are involved in cell proliferation and organ morphogenesis (Synek et al., 2006; Ding et al., 2008) and are thus very likely candidates regulating dissection of leaves in sweetpotato. We did not find any evidence that genes related to circularity in other species--*e.g.*, PTS in tomato (Kimura et al., 2008), STM in *Arabidopsis* (Piazza et al., 2010), RCO in *C. hirsuta* and other Brassicaceae (Vlad et al., 2014; Sicard et al., 2014) and LMI1 in cotton (Andres et al., 2017)--were differentially expressed in sweetpotato accessions that varied for circularity (*i.e.* lobed vs entire). The genes listed above in tomato, Arabidopsis, and cotton alter the complexity and lobedness of the leaves and are mostly genes that contain a homeobox domain. Thus, although preliminary, our data indicate that the degree of lobing in *I. batatas* is not conferred by the action of a gene containing a homeobox domain as in these other species; in fact, none of the differentially expressed transcripts that we identified contained a homeobox domain.

In contrast, we found a total of 100 differentially expressed genes associated with aspect ratio. Based on the function of the homologs of these genes, we identified promising putative candidate genes (Table 3). For example, LIGHT-DEPENDENT SHORT HYPOCOTYLS 10 (LSH10) in *A. thaliana* is hypothesized to be a probable transcription regulator, as its function is to regulate development in response to light by promoting cell growth, indicating that leaf aspect ratio might be determined early in the developmental process. The other promising candidate genes, xyloglucan endotransglucosylase/hydrolase protein 9-like and protein SABRE-like, are involved in cell elongation (Hyodo et al., 2003) and radial expansion (Aeschbacher et al., 1995). Since aspect ratio is a measure of how long-vs-round a leaf is, differential expression of genes involved in radial expansion (causing rounder leaves) and cell elongation (causing longer leaves) are plausible for achieving a difference in AR of a leaf. Finally, differentially expressed

transcripts associated with aspect ratio also included genes having role in thermotolerance and freezing survival. Interestingly, functional plausibility for this idea has already been established in a closely related *Ipomoea* species: in *I. hederacea*, lobed-leaf accessions experienced less damage from freezing night temperatures compared to heart-shaped accessions (Campitelli et al., 2013). We think that the identified thermotolerance genes within this study likely indicate downstream genes that are differentially expressed due to the difference in AR (effect rather than cause). Overall, this finding supports the role of leaf shape in regulating the surrounding temperature, which is the hypothesized path through which leaf shape influences photosynthetic efficiency (Andries et al., 1969).

Factors influencing leaf shape traits in multiple environments

While studies often examine the potential for plasticity in leaf shape traits (McLellan, 2000, Viscosi, 2015, Royer et al., 2009), the relative influence of genetic background, environment and gene by environment interactions are less commonly examined. To fill in this gap, we performed a common garden field study of 74 accessions in two geographically separate common gardens. We show that leaf shape traits (circularity, aspect ratio and solidity) in sweetpotato are highly variable and are influenced by multiple effects. Variation in circularity and solidity were mostly attributed to accession (or genotype) and showed little to no effect due to environment or gene by environment interaction. This is confirmed by the lack of significant changes in circularity and solidity values between gardens. Circularity and solidity have exceptionally high broad-sense heritability values (0.76 and 0.79 respectively, averaged between gardens). These traits have likewise been shown to be highly heritable in tomato with values of h^2 being 0.65 and 0.67, respectively (Chitwood et al., 2013). The high PCV for circularity and solidity in *I. batatas* (22.61% and 11.85%) along with high broad-sense heritability indicates that there is a lot of standing variation for these traits that can be actively selected for (or against) by breeders. Furthermore, the lack of plasticity and GxE demonstrate the stability of these simple leaf shape descriptor traits, at least in the environments tested.

Contrary to our results, multiple studies have found that leaf dissection--captured here by our measure of circularity--is a highly plastic trait that responds to changes in mean annual temperature (Royer et al., 2009, Peppe et al., 2010, Chitwood et al., 2016, Nakayama et al., 2014, Sicard et al., 2014). In general, leaves grown in colder environments are more dissected (*i.e.* lobed) as compared to leaves grown in warmer environments (Royer et al., 2009, Peppe et al., 2010, Chitwood et al., 2016, Nakayama et al., 2014, Sicard et al., 2014). The prevalence of higher dissection in colder environments is probably an adaptation to increase carbon assimilation due to increased transpiration and photosynthate production in toothed species (Royer and Wilf, 2006). Interestingly, researchers have successfully used measures of circularity and solidity to reconstruct paleoclimates in multiple species (Bailey and Sinnott, 1916, Wolfe, 1978, Royer et al., 2009, Peppe et al., 2010), indicating that leaf dissection is often environmentally influenced. The absence of a strong environmental component on our values of circularity and solidity in *I. batatas* is thus relatively surprising, and perhaps indicates that such paleoclimatic relationships might not be easily reconstructed for all species. It may also reflect that our common gardens were not different enough to lead to plastic responses in these two measures of leaf shape. The Ohio garden was consistently warmer (by 2°C on average) and experienced less precipitation than the Michigan garden--the difference between the two gardens was 662.43 mm/month on average throughout the growing season. Thus, although there were

significant environmental differences between gardens, before we conclude that circularity in *I. batatas* is not strongly environmentally responsive, multiple studies in environments that range more widely for temperature will need to be performed.

Comparatively, we found significant variation in aspect ratio due to environment and GxE, explaining 1.93% and 12.95% of the total observed variation in this measure of leaf shape, respectively. This is reflected in the significant alteration of trait values between environments. Aspect ratio in the greenhouse was found to be higher as compared to the gardens but this variation was small with values varying of the clones ranging from 0.02-0.06 ($P < 0.001$). Even between the gardens, there was small yet significant differences observed ($P < 0.001$; 95% CI = 0.009-0.03), with clones grown in Michigan consistently showing less round, more elliptical leaves than clones grown in the Ohio garden. However, we still found that 38.40% of the variation in the trait was due to accessional variation which was also indicated in the estimated heritability value of the trait ($h^2 = 0.24$). Aspect ratio has been found to be a major source of leaf shape variation in apples and tomatoes with high heritabilities of 0.75 and 0.63, respectively (Migicovsky et al., 2017; Chitwood et al., 2013). In contrast, we found that this important leaf shape trait is globally not as variable in sweetpotato (4.76% PCV), but it still presents a selection potential. Since aspect ratio has considerable GxE, this indicates that this trait has a genetic component that interacts with the environment leading to varied values between environment.

Conclusion

For leaf shape to be a crop improvement trait, there should be evidence of phenotypic variation among cultivated accessions and this variation should be genetically controlled and thus able to respond to selection. Overall, the work we present here suggests leaf shape has the potential to serve as a crop improvement trait in sweetpotato: we show that the extensive natural variation in different measures of leaf shape is largely controlled by genetic factors, with a low proportion of variance in leaf shape attributable to environmental differences between gardens. We also identified transcript expression patterns that underlie broad differences in two common descriptors of leaf shape variation--aspect ratio and circularity. This comprehensive and multi-level examination of leaf shape, which is important for a broad understanding of crop-environment interactions, provides somewhat surprising and unexpected results. While we identified differentially expressed genes that underlie leaf shape, they were not the same as those previously identified in other species. Further, and also unexpectedly, we found remarkably little evidence for differences in heritability values between our two common gardens. As a next step, we plan to perform a multi-year, multi-garden experiment with a diverse set of sweetpotato accessions to investigate potential correlational patterns between leaf traits, physiological traits, as well as extend our analysis to yield and quality traits.

ACKNOWLEDGEMENTS:

We thank Dan York and Tyler Marrs for growing the plants in the field and greenhouse, and Tilottama Roy for taking leaf measurements. We are truly grateful for their invaluable field assistance in the UM garden. We thank Andrew Fox, Jordan Francisco, Yufei Gao, Abby Singletary and Nicholas Tomeo for data collection and biomass harvest at the OU garden. We are also extremely thankful to Daniel H. Chitwood for his suggestions and comments on the manuscript. University of Michigan and Ohio University provided funding for this project.

References

Abràmoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with imageJ. *Biophotonics International* (2004).

Aeschbacher, R. A., Hauser, M. T., Feldmann, K. A. & Benfey, P. N. The SABRE gene is required for normal cell expansion in Arabidopsis. *Genes Dev.* (1995).

Andersson, S. Geographical variation and genetic analysis of leaf shape in Crepis tectorum (Asteraceae). *Plant Syst. Evol.* (1991).

Andres, R. J. *et al.* Modifications to a LATE MERISTEM IDENTITY1 gene are responsible for the major leaf shapes of Upland cotton (Gossypium hirsutum L.). *Proc. Natl. Acad. Sci.* **114,** E57–E66 (2017).

Andries, J. A., Jones, J. E., Sloane, L. W & Marshall, J. G. Effects of Okra Leaf Shape on Boll Rot, Yield, and Other Important Characters of Upland Cotton, Gossypium hirsutum L. *Crop Science Society of America* 9: 705-710 (1969).

Bailey, I. W. & Sinnott, E. W. The Climatic Distribution of Certain Types of Angiosperm Leaves. *Source Am. J. Bot.* (1916).

Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014).

Borlaug, N. The Green Revolution, peace, and humanity. AgBioWorld (1970).

Borlaug, N. E. & Dowswell, C. R. Feeding a world of ten billion people : A 21 st century challenge. in *Proceedings of the International Congress "In the Wake of the Double Helix: From the Green Revolution to the Gene Revolution", 27-31 May 2003, Bologna, Italy, 3-23,* (2005).

Brown, V. K., Lawton, J. H. & Grubb, P. J. Herbivory and the Evolution of Leaf Size and Shape. *Philos. Trans. R. Soc. B Biol. Sci.* (1991).

Campitelli, B. E. & Stinchcombe, J. R. Natural selection maintains a single-locus leaf shape cline in Ivyleaf morning glory, Ipomoea hederacea. in *Molecular Ecology* **22,** 552–564 (2013).

Chitwood, D. H. *et al.* The Developmental Trajectory of Leaflet Morphology in Wild Tomato Species. *PLANT Physiol.* (2012).

Chitwood, D. H. *et al.* A Quantitative Genetic Basis for Leaf Morphology in a Set of Precisely Defined Tomato Introgression Lines. *Plant Cell* **25,** 2465–2481 (2013).

Chitwood, D. H. *et al.* A Modern Ampelography: A Genetic Basis for Leaf Shape and Venation Patterning in Grape. *PLANT Physiol.* (2014).

Chitwood, D. H. *et al.* Resolving Distinct Genetic Regulators of Tomato Leaf Shape within a Heteroblastic and Ontogenetic Context. *Plant Cell* (2014).

Chitwood, D. H. *et al.* Climate and developmental plasticity: interannual variability in grapevine leaf morphology. *Plant Physiol.* (2016).

Chitwood, D. H. & Otoni, W. C. Morphometric analysis of Passiflora leaves: The relationship between landmarks of the vasculature and elliptical Fourier descriptors of the blade. *Gigascience* (2017).

Conesa, A. *et al.* Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* (2005).

Covarrubias-Pazaran, G. Genome-Assisted prediction of quantitative traits using the r package sommer. *PLoS One* (2016).

Ding, L., Pandey, S. & Assmann, S. M. Arabidopsis extra-large G proteins (XLGs) regulate root morphogenesis. *Plant J.* (2008).

de Vries, C. A., J. D. Ferwerda, and M. Flach. Choice of food crops in relation to actual and potential production in the tropics. *Neth. J. Agric. Sci*. **15**:241–248 (1967).

Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29,** 644–652 (2011).

Gurevitch, J. Variation in leaf dissection and leaf energy budgets among populations of Achillea from an altitudinal gradient. *Am. J. Bot.* **75,** 1298 (1988).

Huaman Z. Current status on maintenance of sweet potato genetic resources at CIP, in Exploration, maintenance and utilization of sweetpotato genetic resources. CIP: Lima, Peru. 101-120. (1987).

Hue, S. M., Chandran, S. & Boyce, A. N. Variations of leaf and storage roots morphology in Ipomoea batatas L. (Sweet Potato) cultivars. in *Acta Horticulturae* **943,** 73–80 (2012).

Hyodo, H. *et al.* Active gene expression of a xyloglucan endotransglucosylase/hydrolase gene, XTH9, in inflorescence apices is related to cell elongation in Arabidopsis thaliana. *Plant Mol. Biol.* (2003).

Ialongo, C. Understanding the effect size and its measures. *Biochem. Medica* (2016).

Iwata, H. & Ukai, Y. SHAPE: a computer program package for quantitative evaluation of biological shapes based on elliptic Fourier descriptors. *J. Hered.* (2002).

Jones, C. Does shade prolong juvenile development? A morphological analysis of leaf shape changes in Cucurbita argyrosperma subsp. sororia (Cucurbitaceae). *Am. J. Bot.* (1995).

Jöst, M. *et al.* The INDETERMINATE DOMAIN Protein BROAD LEAF1 Limits Barley Leaf Width by Restricting Lateral Proliferation. *Curr. Biol.* (2016).

Kays, S. J., and S. E. Kays. Sweetpotato chemistry in relation to health. Proceedings of the Sweet Potato Production System Towards the 21st Century 231–272. Kyushu National Agricultural Experimental Station Miyakonojo (1998).

Khoury, C. K. *et al.* Distributions, ex situ conservation priorities, and genetic resource potential of crop wild relatives of sweetpotato [ipomoea batatas (L.) lam., I. series batatas]. *Front. Plant Sci.* **6,** 1–14 (2015).

Kim, G. T. *et al.* The ANGUSTIFOLIA gene of Arabidopsis, a plant CtBP gene, regulates leaf-cell expansion, the arrangement of cortical microtubules in leaf cells and expression of a gene involved in cell-wall formation. *EMBO J.* (2002).

Kimura, S., Koenig, D., Kang, J., Yoong, F. Y. & Sinha, N. Natural Variation in Leaf Morphology Results from Mutation of a Novel KNOX Gene. *Curr. Biol.* (2008).

Klein, L. L. *et al.* Digital Morphometrics of Two North American Grapevines (Vitis: Vitaceae) Quantifies Leaf Variation between Species, within Species, and among Individuals. *Front. Plant Sci.* (2017).

Levine, T. R. & Hullett, C. R. Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research. *Human Communication Research* (2002).

Li, M. *et al.* Topological Data Analysis as a Morphometric Method: Using Persistent Homology to Demarcate a Leaf Morphospace. *Front. Plant Sci.* (2018).

Li H. Aligning sequences reads, clone sequences and assembly contigs with BWA-MEM. arXviv:1303.3997v1. (2013).

Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

Lin, R & Wang. H. Arabidopsis FHY3/FAR1 Gene Family and Distinct Roles of Its Members in Light Control of Arabidopsis Development. *PLANT Physiol.* (2004).

Long, S. P., Zhu, X. G., Naidu, S. L. & Ort, D. R. Can improvement in photosynthesis increase crop yields? *Plant, Cell and Environment* (2006).

Mangiafico, S.S.. rcompanion: Functions to Support Extension Education Program Evaluation, R package version 2.0.0. https://CRAN.R-project.org/package=rcompanion (2018).

Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* (2011).

Mathan, J., Bhattacharya, J. & Ranjan, A. Enhancing crop yield by optimizing plant developmental features. *Development* (2016).

McLellan, T. Geographic variation and plasticity of leaf shape and size in Begonia dregei and B. homonyma (Begoniaceae). *Bot. J. Linn. Soc.* (2000).

Meyer, R. S. & Purugganan, M. D. Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* **14,** 840–852 (2013).

Migicovsky, Z., Li, M., Chitwood, D. H. & Myles, S. Morphometrics Reveals Complex and Heritable Apple Leaf Shapes. *Front. Plant Sci.* (2018).

Murugan, S., Paramasivam, S. K. & Nedunchezhiyan, M. Sweet Potato as animal feed and fodder. *Fruit, Veg. Cereal Sci. Biotechnol.* (2012).

Nakayama, H. *et al.* Regulation of the KNOX-GA Gene Module Induces Heterophyllic Alteration in North American Lake Cress. *Plant Cell Online* (2014).

Navarro, D. Learning statistics with R:A tutorial for psychology students and other beginners. *Ua.Edu.Auccsteachinglsr Www.Lulu.Comcontent* (2013).

Peppe, D. J. *et al.* Sensitivity of leaf size and shape to climate: Global patterns and paleoclimatic applications. *New Phytol.* (2011).

Piazza, P. *et al.* Arabidopsis thaliana leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep. *Curr. Biol.* (2010).

Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–40 (2010).

Royer, D. L., Meyerson, L. A., Robertson, K. M. & Adams, J. M. Phenotypic plasticity of leaf shape along a temperature gradient in Acer rubrum. *PLoS One* (2009).

Royer, D. L. & Wilf, P. Why Do Toothed Leaves Correlate with Cold Climates? Gas Exchange at Leaf Margins Provides New Insights into a Classic Paleotemperature Proxy. *Int. J. Plant Sci.* (2006).

Schuepp, P. H. Tansley Review No. 59 Leaf boundary layers. *New Phytol.* **125,** 477–507 (1993).

Schulz, M. H., Zerbino, D. R., Vingron, M. & Birney, E. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28,** 1086–1092 (2012).

Sicard, A. *et al.* Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. *Curr. Biol.* (2014).

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31,** 3210–3212 (2015).

Simmonds, N. W. *Evolution of Crop Plants* (Longman, 1976).

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* **26,** 1134–1144 (2016).

Song, L. & Florea, L. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *Gigascience* (2015).

Suh, S. K, Cho, Y., Park, H. K., Scott, R. A. Gene action and heritability of leaf and reproductive characteristics in soybean. *Breed Sci* 50:45–51 (2000).

Synek, L. et al. AtEXO70A1, a member of a family of putative exocyst subunits specifically expanded in land plants, is important for polar growth and plant development. *Plant J.* (2006).

Tao, X. *et al.* Digital gene expression analysis based on integrated De Novo transcriptome assembly of sweet potato [Ipomoea batatas (L.) Lam.]. *PLoS One* (2012).

Viscosi, V. Geometric morphometrics and leaf phenotypic plasticity: Assessing fluctuating asymmetry and allometry in European white oaks (Quercus). *Bot. J. Linn. Soc.* (2015).

Vogel, S. "Sun Leaves" and "Shade Leaves": Differences in Convective Heat Dissipation. *Ecology* 49:1203-1204 (1968).

Vlad, D. *et al.* Leaf shape evolution through duplication, regulatory diversification and loss of a homeobox gene. *Science (80-.).* (2014).

Vuolo, F. *et al.* Coupled enhancer and coding sequence evolution of a homeobox gene shaped leaf diversity. *Genes Dev.* (2016).

Wolfe, J. a. A paleobotanical interpretation of Tertiary climates in the Northern Hemisphere. *Am. Sci.* (1978).

Wyatt, R. & Antonovics, J. Butterflyweed re-revisited: Spatial and temporal patterns of leaf shape variation in Asclepias tuberosa. *Evolution* **35,** 529–542 (1981).

Yang, J. *et al.* Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants* (2017).

Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18,** 821–829 (2008).

Zhu, X.-G., Long, S. P. & Ort, D. R. Improving Photosynthetic Efficiency for Greater Yield. *Annu. Rev. Plant Biol.* (2010).