New Phytologist

***New Phytologist* Supporting Information**

Article title: **The remarkable morphological diversity of leaf shape in sweetpotato (*Ipomoea batatas*): the influence of genetics, environment, and GXE**

Authors: Sonal Gupta, David M. Rosenthal, John R. Stinchcombe, Regina S. Baucom

Article acceptance date: 19 October 2019

The following Supporting Information is available for this article:

Method S1: RNA-Seq data processing and transcriptome analysis

Fig. S1: Green-house grown accessions selected for transcriptomic analysis.

Fig. S2: Correlation plot between leaf shape traits (traditional and EFD PCs).

Fig. S3: Leaf shape variation captured by EFDs from MI and OH differing significantly in their order of variation explained**.**

Table S1: Accession IDs with their source and location of origin used in this study (see separate file).

Table S2: Differentially expressed transcripts associated with leaf shape traits found in this study (see separate file).

Table S3: Raw read counts of orthologs of homeobox domain genes within the assembled transcriptomes, for accessions chosen for circularity RNA-Seq analysis (see separate file).

Method S1: RNA-Seq data processing and transcriptomic analysis

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 v1.4 (Martin, 2011) to remove the adapters, and trimmomatic v0.36 (Bolger *et al.*, 2014) to clean the reads based on length and quality score. Further, we performed error correction of the RNA-Seq data using rcorrecter(Song & 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 v2.2.0(Grabherr *et al.*, 2011), with k=25, and multi k-mer approach, using Velvet/Oases v1.2.10(Zerbino & 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 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 v1.0.3(Smith-Unna *et al.*, 2016) and BUSCO v3- 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.

References:

Bolger AM, Lohse M, Usadel B. **2014**. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. **2005**. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research.

Bioinformatics **21**: 3674-3676.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* **29**: 644–652.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.

Schulz MH, Zerbino DR, Vingron M, Birney E. **2012**. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* **28**: 1086–1092.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. **2015**. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S. **2016**. TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome research* **26**: 1134–1144.

Song L, Florea L. **2015**. Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* **4**: 48.

Zerbino DR, Birney E. **2008**. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* **18**: 821–829.

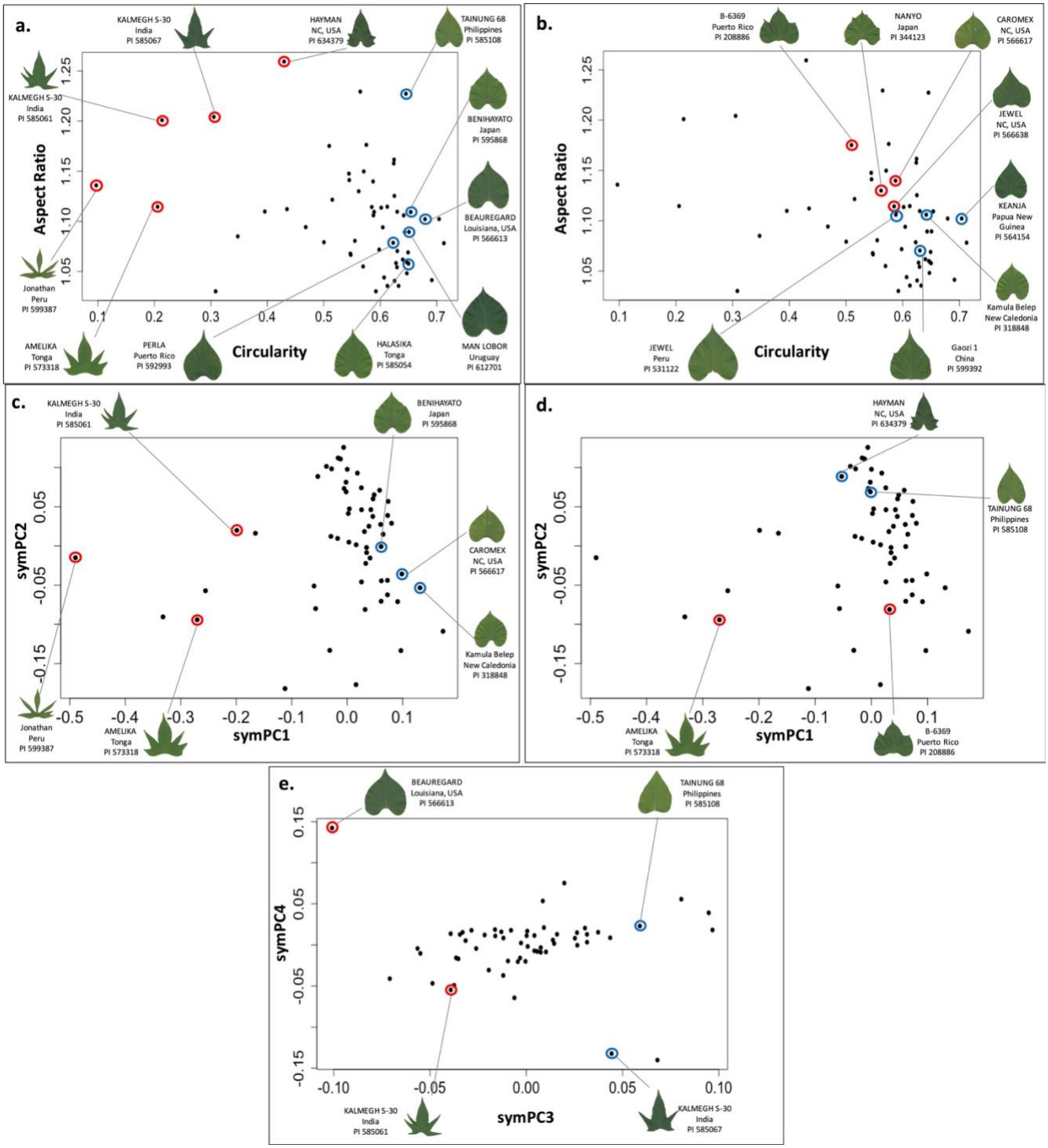

Fig. S1: Green-house grown accessions selected for transcriptomic analysis for leaf shape traits a. Circularity, b. Aspect Ratio, c. symPC1, d. symPC2, and e. symPC3. Red circles represent the accessions chosen for low ends of the trait spectrum and blue circle represents the accessions chosen for high ends of the trait spectrum.

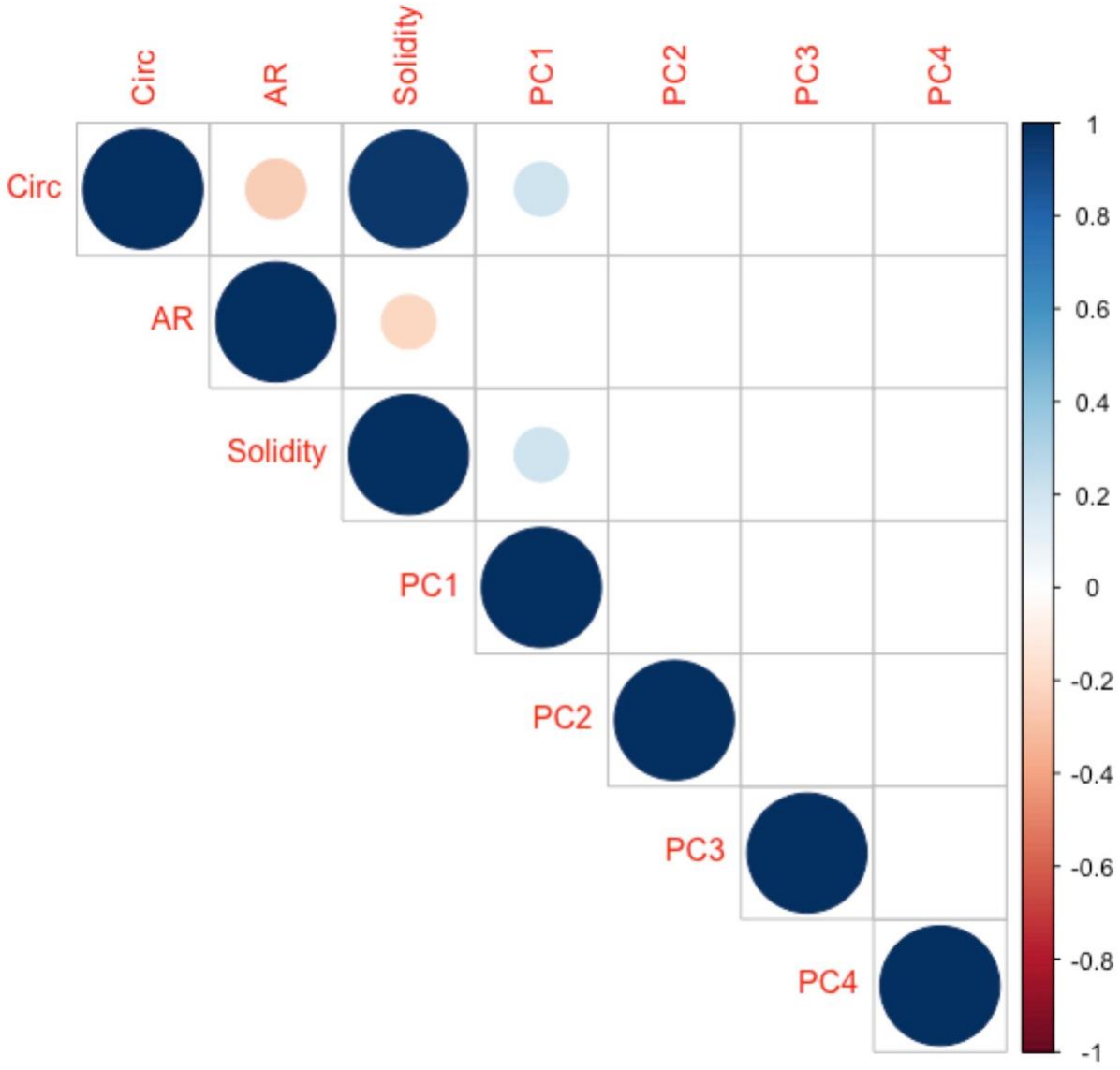

Fig. S2: Correlation plot between leaf shape traits (traditional and EFD PCs) showing that only symPC1 is slightly correlated with circularity and solidity; other traditional leaf shape traits (circularity, aspect ratio and solidity) are not correlated with symPCs.

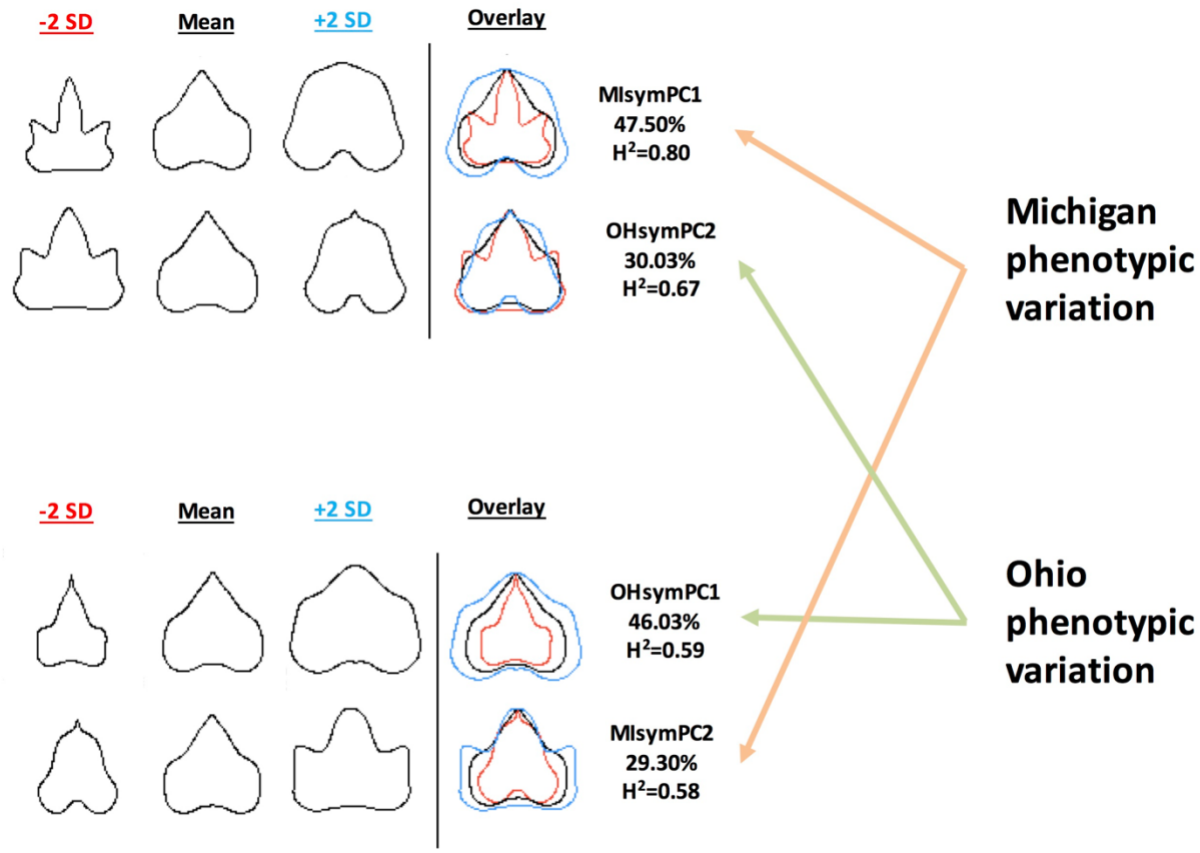

Fig. S3 Leaf shape variation captured by EFDs from MI and OH differing significantly in their order of variation explained**.** MIsymPC1 explains variation in leaf shape that is attributed to lobing, tip and petiolar sinus differences, similar to OHsymPC2 (which only explains ~30% of the variation in OH).