Agave: a biofuel feedstock for arid and semi-arid environments

Stephen Gross [1], Jeffrey Martin [1], June Simpson [2], Zhong Wang [1,3], and Axel Visel [1,3]

[1] Department of Energy Joint Genome Institute, Walnut Creek, CA
[2] CINVESTAV, Irapuato, Guanajuato, Mexico
[3] Genomics Division, Lawrence Berkeley National Lab, Berkeley, CA

Abstract

Efficient production of plant-based, lignocellulosic biofuels relies upon continued improvement of existing biofuel feedstock species, as well as the introduction of new feedstocks capable of growing on marginal lands to avoid conflicts with existing food production and minimize use of water and nitrogen resources. To this end, species within the plant genus *Agave* have recently been proposed as new biofuel feedstocks. Many *Agave* species are adapted to hot and arid environments generally unsuitable for food production, yet have biomass productivity rates comparable to other second-generation biofuel feedstocks such as switchgrass and *Miscanthus*. Agaves achieve remarkable heat tolerance and water use efficiency in part through a Crassulacean Acid Metabolism (CAM) mode of photosynthesis, but the genes and regulatory pathways enabling CAM and thermotolerance in agaves remain poorly understood. We seek to accelerate the development of agave as a new biofuel feedstock through genomic approaches using massively-parallel sequencing technologies. First, we plan to sequence the transcriptome of *A. tequilana* to provide a database of protein-coding genes to the agave research community. Second, we will compare transcriptome-wide gene expression of agaves under different environmental conditions in order to understand genetic pathways controlling CAM, water use efficiency, and thermotolerance. Finally, we aim to compare the transcriptome of *A. tequilana* with that of other *Agave* species to gain further insight into molecular mechanisms underlying traits desirable for biofuel feedstocks. These genomic approaches will provide sequence and gene expression information critical to the breeding and domestication of *Agave* species suitable for biofuel production.

Overview

I. *Agave* can supplement other bioenergy feedstocks

Agave species are adapted to their native habitat in the semi-arid regions of Mexico and the United States. *Agave* thus holds promise as a biofuel feedstock [1], capable of growing on marginal lands where other proposed bioenergy plants such as switchgrass and *Miscanthus* cannot. Their ability to withstand high heat and semi-arid conditions largely relies upon their use of Crassulacean Acid Metabolism (CAM)—a specialized form of photosynthesis in which CO_2 uptake by the plant and CO_2 fixation by photosynthesis is temporally separated between day and night. CAM allows agaves to keep leaf stomata (pores) closed during the hot day, minimizing water loss through evapotranspiration.

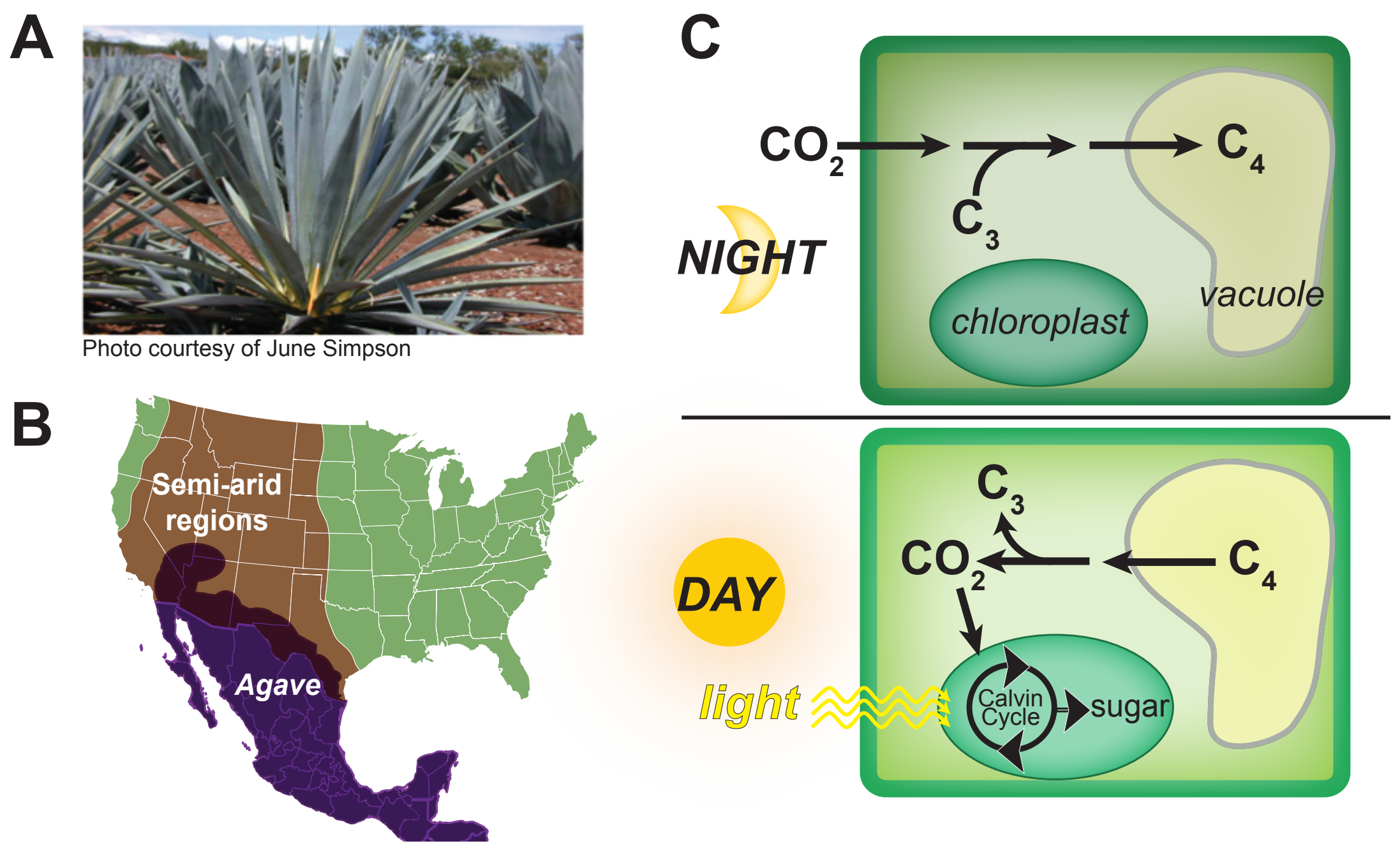


Photo courtesy of June Simpson

A. *Agave tequilana* cultivated in Mexico.

B. Semi-arid regions of the United States (brown) are unsuitable for cultivation of other bioenergy plants, which require more temperate regions (green). Most *Agave* species are adapted to semi-arid regions in Mexico and the extreme southwestern USA (purple).

C. Crassulacean Acid Metabolism (CAM). CO_2 enters plant cells at night, joins with a 3-carbon molecule (C_3) and is stored in the vacuole as a 4-carbon molecule (C_4). During the day, C_4 molecules diffuse out of the vacuole, and CO_2 is relased and assimilated into sugar in the chloroplast.

II. Agaves are productive on marginal lands

Agaves are capable of efficiently producing lignocellulosic biomass with little water and nitrogen (fertilizer) inputs. Some species of *Agave*, such as *A. salmiana* and *A. mapisaga* have been reported to produce up to 40 metric tonnes (Mg) of dry biomass per hectare per year [2].

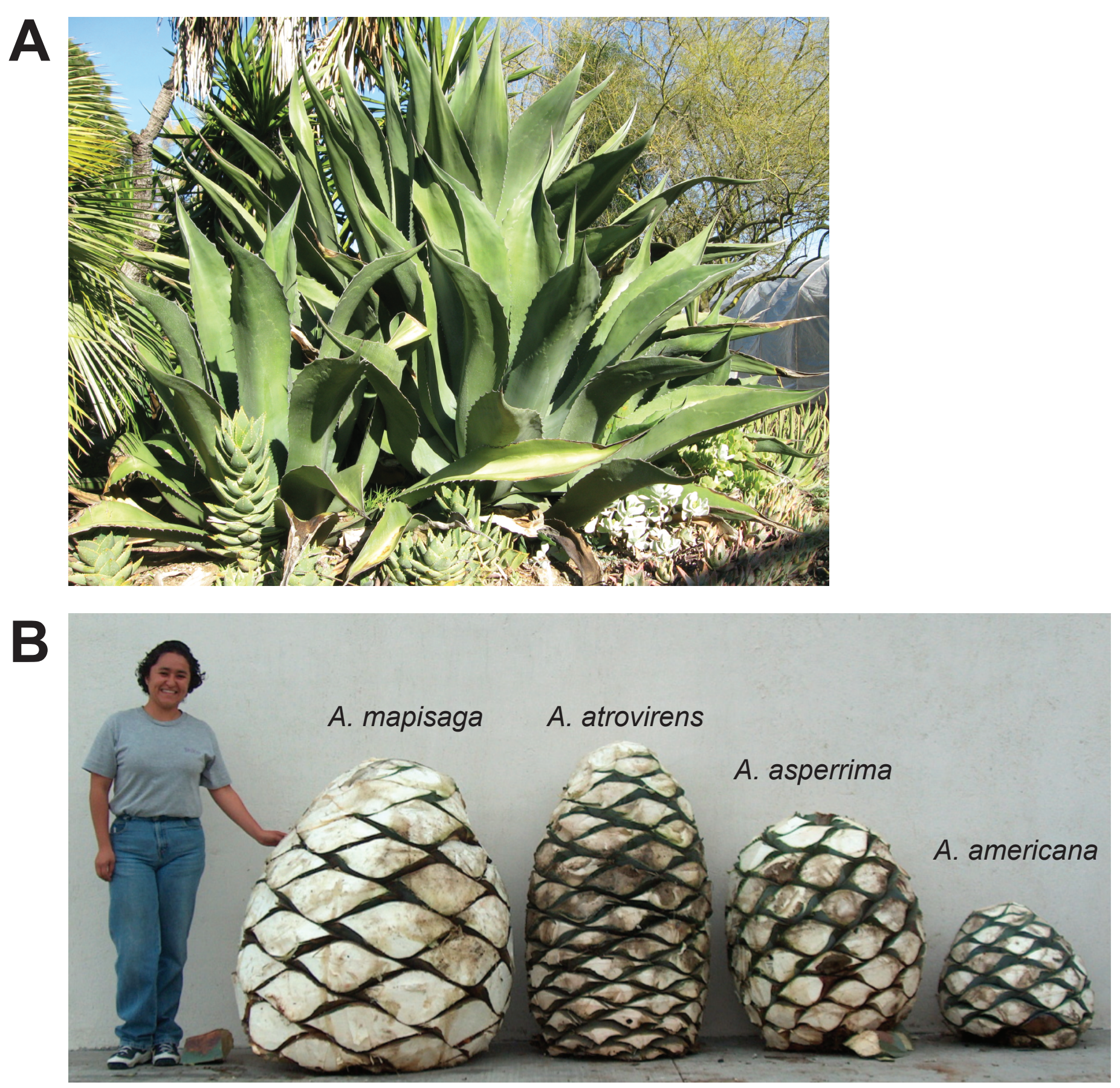


Photo courtesy of June Simpson and reference [3]

A. *Agave salmiana*, one of the most productive agaves known. *A. salmiana* grows well outside its native habitat in areas where winter temperatures rarely reach freezing. The specimen shown here is growing at the Ruth Bancroft Garden in Walnut Creek, CA.

B. Some *Agave* species are very large. Removing the leaves exposes the stem, or *piña* (Spanish for pinecone or pineapple). The *piña* of the largest species can weigh hundreds of pounds.

C. Comparison of inputs (water and nitrogen) and outputs (biomass and ethanol) of agaves and other biofuel feedstock species. Table is modified from reference [4].

C

	Inputs			Outputs	
Feedstock	**Water cm yr^{-1}**	**Drought tolerance**	**Nitrogen kg ha^{-1} yr^{-1}**	**Biomass Mg ha^{-1} yr^{-1}**	**Ethanol liters yr^{-1}**
Corn (grain)	50–80	low	90–120	7–10	2900
(stover)				3–6	900
Sugarcane (sugars)	150–250	moderate	0–100	11	6900
(bagasse)				10	3000
Miscanthus	75–120	low	0–15	15–40	4600–12,400
Poplar coppice	70–105	moderate	0–50	5–11	1500–3400
***Agave* spp.**	**30–80**	**high**	**0–12**	**10–34**	**3000–10,500**

Table adapted from reference [4]

Building a genomics resource for *Agave*

III. *de novo* assembly of an agave transcriptome

A bottleneck in *Agave* research is the paucity of sequences available, making genomic-scale molecular analyses, marker development for mapping and breeding, and candidate gene approaches difficult or impossible. We seek to enhance sequence resources for *Agave* research, and chose *A. tequilana*, the agave currently cultivated for the tequila industry, as our reference species. As agaves have large genomes (~4–7 Gb) [5, 6], we propose to sequence the transcriptome using Illumina massively-parallel RNA-seq technologies. A new assembly pipeline developed at JGI, Rnnotator [7], will be used to assemble raw RNA-seq data for *de novo* transcriptome assembly without a reference genome.

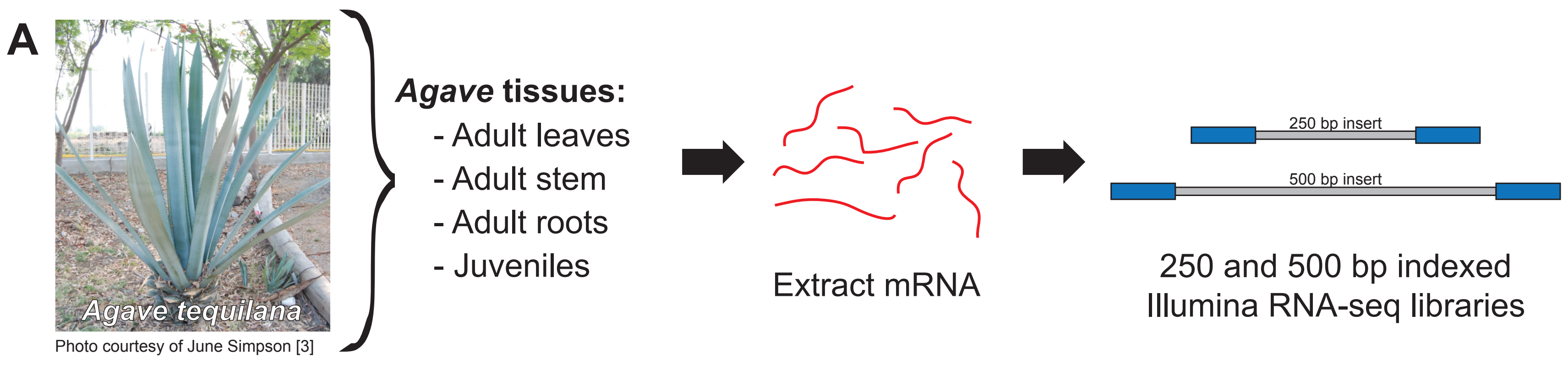
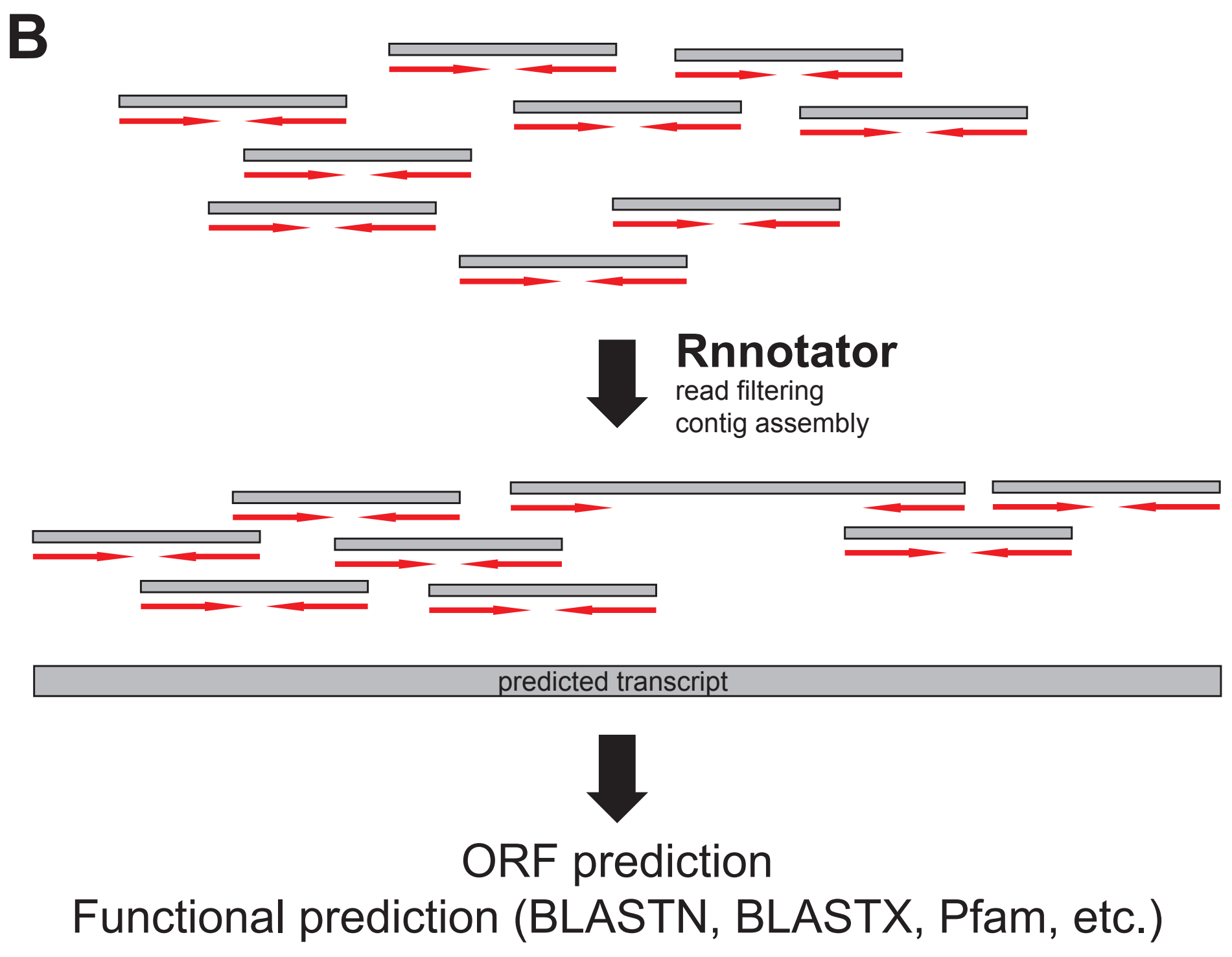


Photo courtesy of June Simpson [3]

***Agave* tissues:**
- Adult leaves
- Adult stem
- Adult roots
- Juveniles

Extract mRNA

250 and 500 bp indexed Illumina RNA-seq libraries



Rnnotator
read filtering
contig assembly

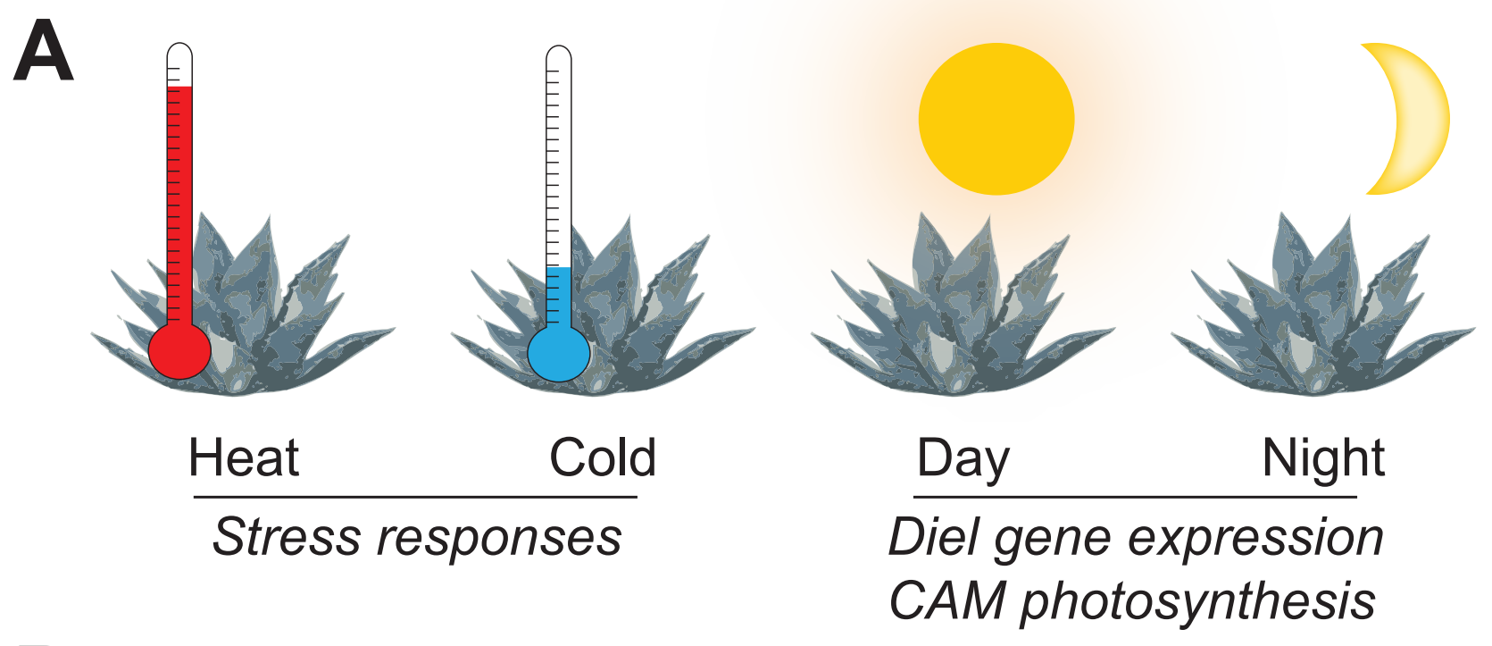
ORF prediction
Functional prediction (BLASTN, BLASTX, Pfam, etc.)

A. Illumina RNA-seq library production. Four distinct tissues of *A. tequilana* were chosen: adult leaves, stems, and roots; as well as juvenile roots and shoots. mRNA was purified from each tissue, fragmented to either 250 or 500 nt lengths, reverse transcribed into DNA, and ligated to Illumina indexed adaptors (blue). Our procedure preserves information about the mRNA source and strand specificity to aid transcriptome assembly.

B. Transcriptome assembly and annotation. Raw paired-end reads (red arrows) of the agave RNA-seq libraries (grey bars) are filtered and assembled into contigs by Rnnotator. Resulting contigs, representing specific splice variants of gene transcripts, will be annotated with predicted function using several bioinformatic tools.

IV. Pathway discovery using RNA-seq

With a reference agave transcriptome in hand, we plan to investigate genetic pathways providing agaves with their ability to survive the stresses of semi-arid environments. RNA-seq [8] is an exceptional tool to quantitatively investigate gene expression. *A. tequilana* plants will be subjected to distinct environmental conditions, and gene expression will be analyzed to uncover genes and pathways involved in stress responses. We expect to apply this same approach to other *Agave* species which have traits that may be bred or engineered into future biofuel-optimized agave cultivars.



A. Some possible experiments to compare gene expression of agaves under different environmental conditions. Quantitative RNA-seq can be used to detect differences in gene transcript levels, providing a way to uncover genetic pathways controlling stress responses or circadian rhythms of agave.

B. *Agave parryi*, native to New Mexico and Arizona, is one of the most cold-tolerant species of *Agave*. Studying its response to cold stress may provide insight into genes that can be used to control frost damage in agave biofuel cultivars grown in colder climates.

V. Acknowledgements

The authors wish to thank Chia-Lin Wei, Natasha Zvenigorodsky, and Cindy Choi (JGI) for technical assistance and project planning. This work is supported by a Lawrence Berkeley National Lab LDRD grant to A. Visel. This work performed at the U.S. Department of Energy Joint Genome Institute was supported in part by the Office of Science of the U.S. Department of Energy under contract DE-AC02-05CH112.^{31}

VI. References

[1] Davis, A. S. *et al.* The global potential for *Agave* as a biofuel feedstock. GCB Bioenergy 3, 68–78, (2011).
[2] Nobel, P. S. *et al.* High annual productivity of certain agaves and cacti under cultivation. Plant Cell Environ 15, 329-35, (1992).
[3] Simpson, J. *et al.* Genomic resources and transcriptome mining in *Agave tequilana*. GCB Bioenergy 3, 25–36, (2011).
[4] Somerville, C. *et al.* Feedstocks for lignocellulosic biofuels. Science 329, 790-2, (2010).
[5] Palomino, G. *et al.* Nuclear genome size analysis of *Agave tequilana* Weber. Caryologia 56, 37-46, (2003).
[6] Robert, M. L. *et al.* Wild and agronomically important *Agave* species (Asparagaceae) show proportional increases in chromosome number, genome size, and genetic markers with increasing ploidy. Bot J Linn Soc 158, 215-22, (2008).
[7] Martin, J. *et al.* Rnnotator: an automated *de novo* transcriptome assembly pipeline from stranded RNA-seq reads. BMC Genomics 11, 663, (2010).
[8] Wang, Z. *et al.* RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63, (2009).