Profiling genome-wide methylation in two maples: fine-scale approaches to detection with nanopore technology.

Susan L. McEvoy[1], Patrick G. S. Grady[2], Nicole Pauloski[2,3], Jill L. Wegrzyn[1#]

[1] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut, 06269
[2] Department of Molecular and Cell Biology, University of Connecticut, Storrs, Connecticut, 06269
[3] Institute for Systems Genomics, University of Connecticut, Storrs, Connecticut, 06269

Mailing address: 75 N. Eagleville Rd., Unit 3043, Storrs, CT 06269

[#]Corresponding author: Jill Wegryzn, email: jill.wegrzyn@uconn.edu,

Running head: Nanopore-based Methylation Profiling of Two Acers

Keywords: Acer saccharum, Acer negundo, nanopore, methylation, methylome, transposable elements, nutrient stress, abiotic stress, assembly, annotation, METEORE, DeepSignal-Plant, Nanopolish

ABSTRACT

DNA methylation is critical to the regulation of transposable elements and gene expression. Traditional methods of methylation quantification rely on bisulfite conversion, which requires treatments to the DNA that can compromise accuracy. Recent advances in long-read sequencing technologies allow for methylation detection in real time. The associated algorithms that interpret these modifications have evolved from strictly statistical approaches to Hidden Markov Models and, recently, deep learning approaches. Much of the existing software focuses on methylation in the CG context but methylation in other contexts is important to quantify, as it is extensively leveraged in plants. Here, we present methylation profiles for two maple species across the full range of 5mC sequence contexts using Oxford Nanopore long-reads. Hybrid and reference-guided assemblies were generated for two new *Acer* accessions: *Acer negundo* (65x nanopore and 111X Illumina) and *Acer saccharum* (93x nanopore and 148X Illumina). The nanopore reads generated for these assemblies were re-basecalled, and methylation detection was conducted in a custom pipeline with the published *Acer* references (PacBio assemblies) and the new hybrid assemblies to generate four epigenomes. The abundance and distribution of genes (including those previously associated with abiotic stress response), repeats, and methylation contexts were examined and compared with those of recently characterized broadleaf tree species.

INTRODUCTION

The processes shaping plant development and growth are regulated by epigenetic modifications that impact gene expression, genomic stability, and plasticity (Kumar and Mohapatra 2021). Plants leverage methylation in sequence contexts beyond the CG dinucleotide (CHG and CHH, where H = C, A, or T), and these modifications are primarily regulated by transposable elements (TEs), which represent a significant portion of most plant genomes. Epigenetic modifications introduced through mobile elements contribute to genetic variation that is associated with biotic and abiotic stress adaptations (Ritter and Niederhuth 2021).

Methylation in model plants has been studied from its initiation in embryos in *Arabidopsis thaliana* (Jullien et al. 2012), to the accumulation of methylation variation in the independent branches of a single *Populus trichocarpa* individual (Hofmeister et al. 2020). In *Arabidopsis*, the genome-wide regulatory effects of methylation were demonstrated by the knockout of all five methyltransferases, impacting cell fate throughout the plant (He et al. 2022). Also in *Arabidopsis*, a heritable epiallele associated with climate was found to control leaf senescence, providing an example of local climate adaptation (He et al. 2018). In *P. trichocarpa*, epigenetic modifications were associated with changes in the circadian cycle (Liang et al. 2019). Increasingly, non-model plant systems have been investigated, including for methylation responsible for leaf shape and photosynthetic traits in *Populus simonii* (Ci et al. 2016) and salinity-induced methylation in mangroves (Miryeganeh et al. 2022).

Until recently, methylation was primarily investigated through treatments such as whole-genome bisulfite conversion (WGBS). This technique is prone to degradation of DNA, incomplete

conversion, and amplification bias (Gouil and Keniry 2019). Since WGBS libraries are primarily short-read sequenced, interpretation also suffers from poor resolution in repetitive regions. Long-read technologies, including Pacific Biosciences' (PacBio's) single-molecule real-time (SMRT) sequencing, and nanopore sequencing by Oxford Nanopore Technologies (nanopore), can detect methylation. SMRT sequencing can detect 5mC modifications based on polymerase dynamics at very high coverage, as well as methods that rely on bisulfite conversion for standard coverage. In comparison, nanopore sequencing can directly detect DNA or RNA modifications through a voltage-measured pore, enabling real-time, single-molecule sensitivity (Liu et al. 2021).

Being sessile, plants rely on heritable methylation as an evolutionary strategy, and this is particularly important in long-lived tree species. DNA methylation is known to have a critical role in the silencing of TEs, but environmental stress can reduce this activity and result in TE bursts (Cavrak et al. 2014). Epigenetic mechanisms related to transposition have been associated with responses to drought, temperature, and nutrient stress (Fan, Peng, and Zhang 2022). In the context of maples, sugar maple (*A. saccharum*) is susceptible to calcium deficiency, and this has led to a significant decline in natural populations (Bishop et al. 2015). The first comparative genomics study on North American maples identified candidate genes from the analysis of expression in the aluminum- and calcium-amended plots at the Hubbard Brook Experimental Forest (McEvoy et al. 2021). The interplay between TEs and gene expression is complex, but many forest tree species would benefit from a deeper examination to fully understand their adaptive potential for new and ongoing threats.

This study extends the previous work on *A. negundo* and *A. saccharum* (McEvoy et al. 2021), as well as that of Sork et al. (2021) and Niederhuth et al. (2016) in comparative plant methylomics. With nanopore sequencing from new *Acer* individuals, we completed genome assembly and annotation and detected methylation. We focused on methylation calling methods and, in doing so, generated comparative methylation profiles focusing on transposable elements and nutrient stress candidate genes.

## METHODS

### Sequencing

Leaves from two maple individuals, *A. negundo* (Accession 253-2013*B) and *A. saccharum* (1353-84*A), were shipped on dry ice from the Arnold Arboretum of Harvard University. High molecular weight (HMW) DNA was extracted from both samples using the nanopore protocol for *Arabidopsis* leaves (Vaillancourt and Robin Buell 2019). The resulting gDNA was checked for quality control via Thermo Scientific Nanodrop and then Agilent TapeStation (Agilent Technologies, Santa Clara, CA). Libraries were prepared with the nanopore Genomic DNA by Ligation protocol with additional Covaris shearing to improve coverage (Oxford Nanopore Technologies, Oxford, UK). A single nanopore Flongle sequencing run was conducted to evaluate the library quality prior to the PromethION run (*A. negundo*: 15 K reads, N50 16 Kb in 23 hours; *A. saccharum*: 39 K reads, N50 14 Kb after 21 hours). Two PromethION runs (one per individual) followed, using flow cell type FLO-PRO002, kit SQK-LSK110, and Guppy v4.0.11

(Oxford Nanopore Technologies 2020a) with the high-accuracy basecalling model and read filtering minimum qscore of 7. To reduce the error rate, particularly for methylation calling, the resulting FAST5s were re-basecalled with Guppy v5.0.16 (GPU) and the latest nanopore super-accuracy model. The same samples were short-read sequenced in a single Illumina NovaSeq 6000 SP v1.5 300 cycle run. TruSeq DNA Nano with Covaris shearing was used for Illumina library preparation in advance (Illumina Inc., San Diego, CA, USA).

**Assembly and annotation**

Nanopore long-reads were filtered for archaea, bacteria, fungi, and virus contaminants via Centrifuge v1.0.4-beta (Kim et al. 2016) as well as length (5 Kb minimum). Illumina short reads were quality-controlled with FASTP v0.22.0 (Chen et al. 2018). The filtered nanopore reads were combined with raw Illumina reads for hybrid assembly with MaSuRCA v4.0.3 (Zimin et al. 2017). This was followed by short-read polishing with Pilon v1.24 (Walker et al. 2014) using the trimmed short-reads that were aligned with Bowtie v2.3.5.1 (Langmead et al. 2009). Scaffolding was performed with RagTag v2.1.0 (Alonge et al. 2021) using the original chromosome-scale reference genomes for the same species (McEvoy et al. 2021).

Repeats were identified with RepeatModeler v2.0.1 (Flynn et al. 2020) and softmasked with RepeatMasker v4.0.6 (Smit, Hubley, and Green 2013–2015). ParseRM generated repeat summaries and abundance by divergence estimates. BRAKER2 v.2.1.6 predicted genes with the previously published RNA-Seq leaf tissue library provided as evidence (Brůna et al. 2021) The RNA-Seq leaf tissue libraries represented 31.2 M read pairs for *A. negundo* and 30.6 M read pairs for *A. saccharum*. Gene models were filtered with gFACs v1.1.3 (Caballero and Wegrzyn 2019) with the following: unique genes only, mono-exonics missing a start or stop codon or valid protein domain, mult-iexonics missing both a start and stop codon, and genes with exons smaller than 6 bp. Functional annotation of the final gene space was conducted with EnTAP v0.10.7 using RefSeq Complete and Uniprot NCBI databases, along with Eggnog v4.1 for gene family assignment (O'Leary et al. 2016; Hart et al. 2018; UniProt Consortium 2019).

**Methylation detection**

The METEORE pipeline was selected as the general approach to methylation detection based on its ability to generate a reliable consensus results from multiple tools (Yuen et al. 2021). Two tools were selected: Nanopolish v0.13.2 (Loman, Quick, and Simpson 2015) and DeepSignal-Plant v0.1.4 (Ni et al. 2021). This pairing was chosen as it had favorable results in the Yuen et al. (2021) benchmarking study and the ability to inform beyond the CG-context. Nanopolish is a well-supported tool that detects methylation in the CG context. DeepSignal-Plant is the top-performing, most accessible tool trained with plant-based models to detect methylation in CHG and CHH contexts from nanopore sequencing (Ni et al. 2021).

Nanopore reads from the new individual were aligned to both the new and original genomes for methylation calling. Methods for Nanopolish and DeepSignal-Plant proceeded according to the documentation for each tool and the Snakemake workflows provided by METEORE (Fig. 1). Each tool was run to create two output formats: 1) the input format necessary for integration in the consensus, and 2) the standardized form of independent tool output, calculating per-site and

per-strand frequencies, allowing for greater ease of interpretation across tools. This provided both tool-specific frequencies and a consensus for comparison.

To begin, the rebasecalled and filtered FASTQ reads were used to filter the FAST5 using ont_fast5_api ‘fast5_subset.py’ v4.0.0 (Oxford Nanopore Technologies 2021). Nanopolish requires indexing of FASTQ to FAST5 in the default multi-fast5 format, and then alignment with minimap2 with parameters -ax map-ont (H. Li 2018). These alignments were run with minimap2 version 2.22-r1101 outside of the METEORE pipeline because the within-METEORE version provided by Snakemake was older. The next step was Nanopolish 'call_mods', followed by METEORE scripts to convert the output of log-likelihood ratio values into a standardized format.

DeepSignal requires single-fast5 formatted files, so the multi-fast5 files were converted using ont_fast5_api ‘multi_to_single_fast5’. DeepSignal-Plant began with annotation of the FAST5s with FASTQ using Tombo ‘preprocess annotate_raw_with_fastqs’ v1.5.1 (Oxford Nanopore Technologies 2020b). This was followed by Tombo ‘resquiggle’ with the parameter --signal-align-parameters 4.2 4.2 2000 2500 20.0 40 750 2500 250 used for the original genomes only. It is possible to detect methylation on either reads or extracted features; reads were recommended, so this method was implemented. At this stage, results were split into separate files for CG, CHG, CHH, and CHH subcontexts by modifying METEORE split_freq_file_by_5mC_motif_sm.py. METEORE scripts were then used to standardize per-site and per-strand formats as described for Nanopolish. The CG file alone was formatted as input for the consensus script. Consensus predictions were generated using a Random Forest with provided models optimized at n-estimator = 3 and max_dep = 10. Results were analyzed with BEDTools v2.29.0 (Quinlan and Hall 2010) and karyoplotR (Gel and Serra 2017).

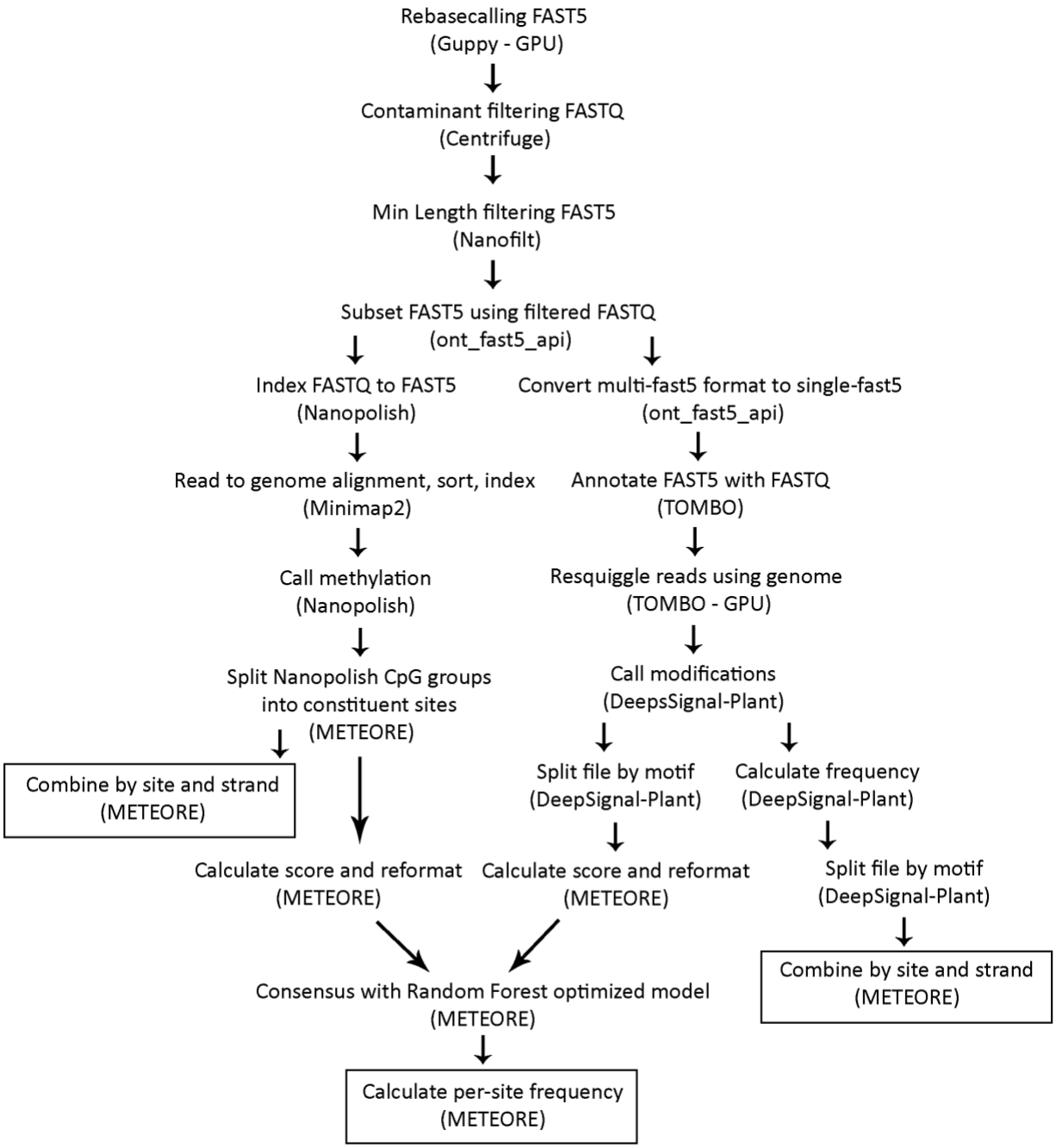



**Figure 1.** Stepwise method of methylation detection using the METEORE pipeline to create a consensus of DeepSignal-Plant and Nanopolish.

### Statistical analysis and visualization

BEDTools v2.29.0 (Quinlan and Hall 2010) was used to create 1 Mb windows and 100 Kb windows with 50 Kb overlaps across the genomes. It was then used to map methylation to these windows and calculate the frequency mean. BEDTools was also used with gene annotation files to intersect the gene regions and count methylated sites within the region. Plotting of chromosomal distributions was conducted with karyoplotR v1.21.3 (Gel and Serra 2017).

Pearson correlation was calculated with R Core Stats Package v4.2.0 (R Core Team 2013). Statistics and summaries were used to compare across reference genomes within *Acer* (McEvoy et al. 2021), as well as *Populus* (Hofmeister et al. 2020), *Quercus* (Sork et al. 2022), and the 34 angiosperms surveyed by Niederhuth (2016). Rapidly expanding, and contracting and expanding, families within *A. negundo* and *A. saccharum* were plotted along the distributions of methylation frequency and TE coverage. A total of 245 candidate genes associated with calcium response in *A. saccharum* were also investigated.

Scripts for all methods are available at https://gitlab.com/PlantGenomicsLab/acermethylation

RESULTS AND DISCUSSION

**Sequencing**

The nanopore sequencing of *A. negundo* resulted in 48.5 Gb (115x coverage) in 4.29 M reads with a read N50 of 16.3 Kb. Re-basecalling resulted in an expected loss, resulting in 29.6 Gb (70x) in 2.55 M reads with an N50 of 16.2 Kb. After contaminant (56 K reads) and length filtering, 65x coverage with an N50 of 16.6 Kb remained. Illumina raw reads had 111x coverage as 46.6 Gb in 310 M reads. After trimming for adaptors, length, and quality, 104x coverage (146 M total read pairs) remained (File S1). Nanopore sequencing of *A. saccharum* resulted in 104.5 Gb in 9.8 M reads with an N50 of 15.1 Kb. Re-basecalling resulted in 56.7 Gb (99x coverage) in 5.2 M reads with an N50 of 15.2 Kb. Filtering (135 K contaminant reads) generated coverage of 93x with a read N50 of 15.5 Kb. Illumina sequencing (150bp PE) produced 148x coverage of 52.9 Gb bases in 563 M read pairs and trimming reduced this to 139x (526 M total read pairs).

**Assembly and annotation**

After the initial draft assembly, *A. negundo* had a total length of 421 Mbp in 421 contigs and an N50 of 2.16 Mb. BUSCO embryophyta genes were 96.0% complete, with 2.9% of these in duplicate. Polishing only minimally reduced the length and N50 (File S1). Scaffolding with the original genome increased the length slightly, but it remained within 421 Mbp, and the N50 grew to 33 Mb. The assembly was in 32 scaffolds, with 13 chromosomes representing 99.7% of the assembled length. The genome size remained constant, though slightly smaller, than the published reference (442 Mbp) and closer to the kmer-based estimate of 319 Mbp (McEvoy et al. 2021). At 96%, the final complete assembly BUSCO scores remained similar to the draft, with duplicates dropping slightly to 2.8%, 0.7% fragmented, and 3.3% missing (Table 1, File S1). This was an improvement from the original reference duplication value of 3%, though the single copy completes dropped slightly. Structural annotation of the new genome identified 27,541 genes, of which 23,408 were functionally annotated by either similarity search or gene family assignment. The BUSCO score for annotated proteins was 92.1% complete, with 3% in duplicate (Table 1).

The *A. saccharum* draft assembly had a total length of 571 Mbp in 1194 contigs and an N50 of 805 Kb. BUSCO embryophyta genes were 96.0% complete, with 6.6% of those in duplicate. Similar to *A. negundo*, polishing slightly reduced the total length and N50 in this genome. The BUSCO duplicate score increased by 0.1%. The BUSCO of the scaffolded genome was 96.2%

complete, with 6.5% in duplicate, 0.7% fragmented, and 3.1% missing. The assembly had a total length of 571 Mbp in 160 contigs and an N50 of 41 Mb (Table 1, File S1). Scaffolding resulted in 13 chromosomes that represented 96.8% of the assembled length. The genome size dropped from the published reference of 626 Mbp, a more substantial decrease than seen in *A. negundo*, and below the original kmer-based estimate of 636 Mbp. A total of 35,834 genes were identified, and 29,858 were associated with functional information. The resulting annotation BUSCO score was 91.8% complete, containing 8.5% duplicates (Table 1, File S1).

The first versions of the *Acer* genomes were exclusively deep-coverage (>100X) PacBio Sequel II, and resulted in fairly contiguous references that assembled to chromosome-scale with the addition of HiC libraries (McEvoy et al. 2021). The two new accessions, assembled in a hybrid manner and scaffolded with the published PacBio references as described above, were smaller than the original references (Table 1). The difference in size could be structural variation between the different genotypes, but more likely reflects some of the differences in read inputs and methodology (assembler). *A. negundo* was most similar to its original genome, with an improved duplication rate, and the missing genes identified (~3 K) were not specific to the new genome. *A. saccharum* had an increase in BUSCO-estimated duplication and 4 K original genes were not observed in the new assembly. Similar to the assembly, differences in final gene number may partially reflect variation in the BRAKER software used for prediction. Whole-genome alignment between the new and original versions revealed that, of the existing assembled sequence, no major discrepancies were present in either species (Fig. 2a). Links between syntenic regions in both species showed similarities in spite of the larger genome size of *A. saccharum* (Fig. 2b).

**Table 1.** Assembly and annotation statistics for new and original genomes

| **Assembly** | **Genome** | | | **Genes** | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Length** | **N50** | **BUSCO*** | **Total** | **Mono-exonic** | **Average length** | **Avg. CDS length** | **BUSCO*** |
| **ACNE2** | 421 Mbp | 33 Mb | 96.0 (2.8) | 27,541 | 3474 | 3098 | 1122 | 92.1 (3.0) |
| **ACNE** | 442 Mbp | 32 Mb | 97.7 (3.0) | 30,491 | 5558 | 5386 | 1174 | 94.1 (6.8) |
| **ACSA2** | 571 Mbp | 41 Mb | 96.2 (6.5) | 35,834 | 5366 | 2918 | 1087 | 91.8 (8.5) |
| **ACSA** | 626 Mbp | 46 Mb | 97.4 (5.6) | 40,074 | 8765 | 6761 | 1190 | 93.1 (7.7) |

* Complete % (Duplicate %), Embryophyta 10

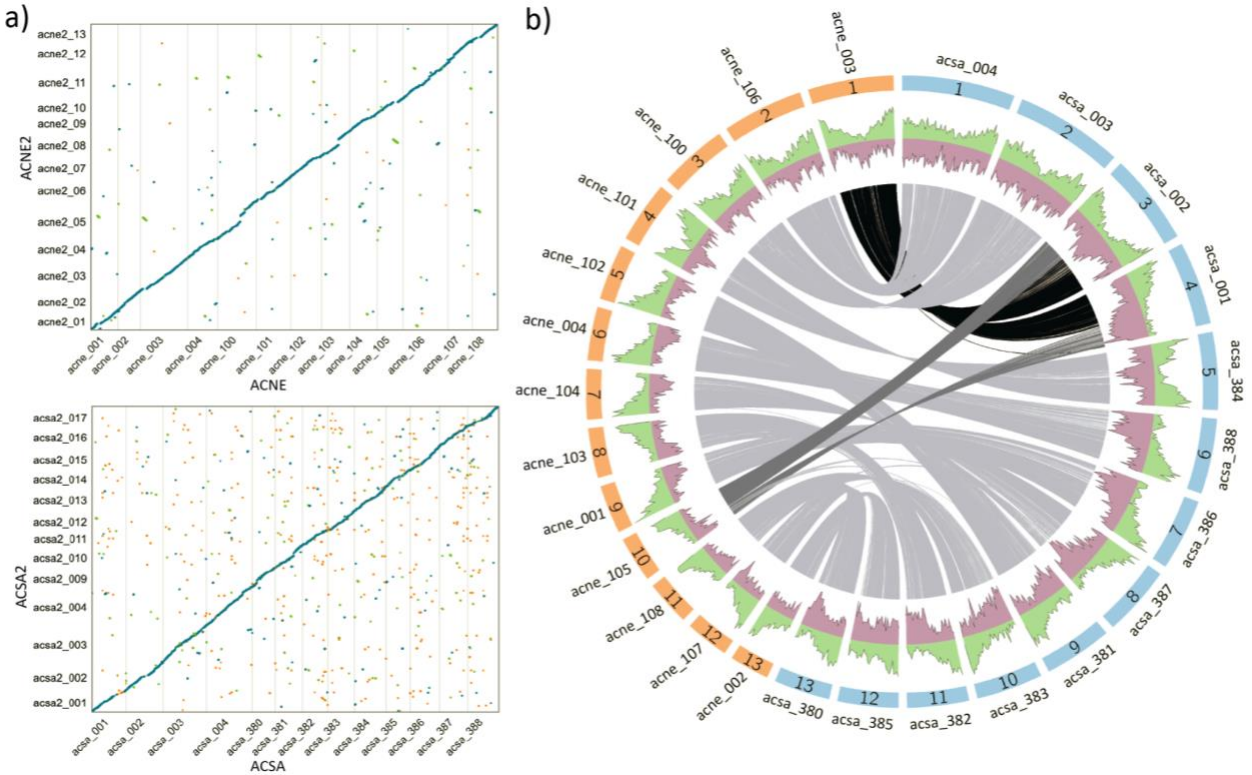



**Figure 2.** a) Alignment of reference genomes comparing original (x-axis) and new (y-axis) assemblies for *A. negundo* (ACNE) and *A. saccharum* (ACSA). b) Gray lines show blocks of syntenic genes between *A. negundo* (orange) and *A. saccharum* (blue) using the original reference genomes (reprinted from McEvoy et al. (2021)). The green area depicts gene density and purple is repeat frequency.

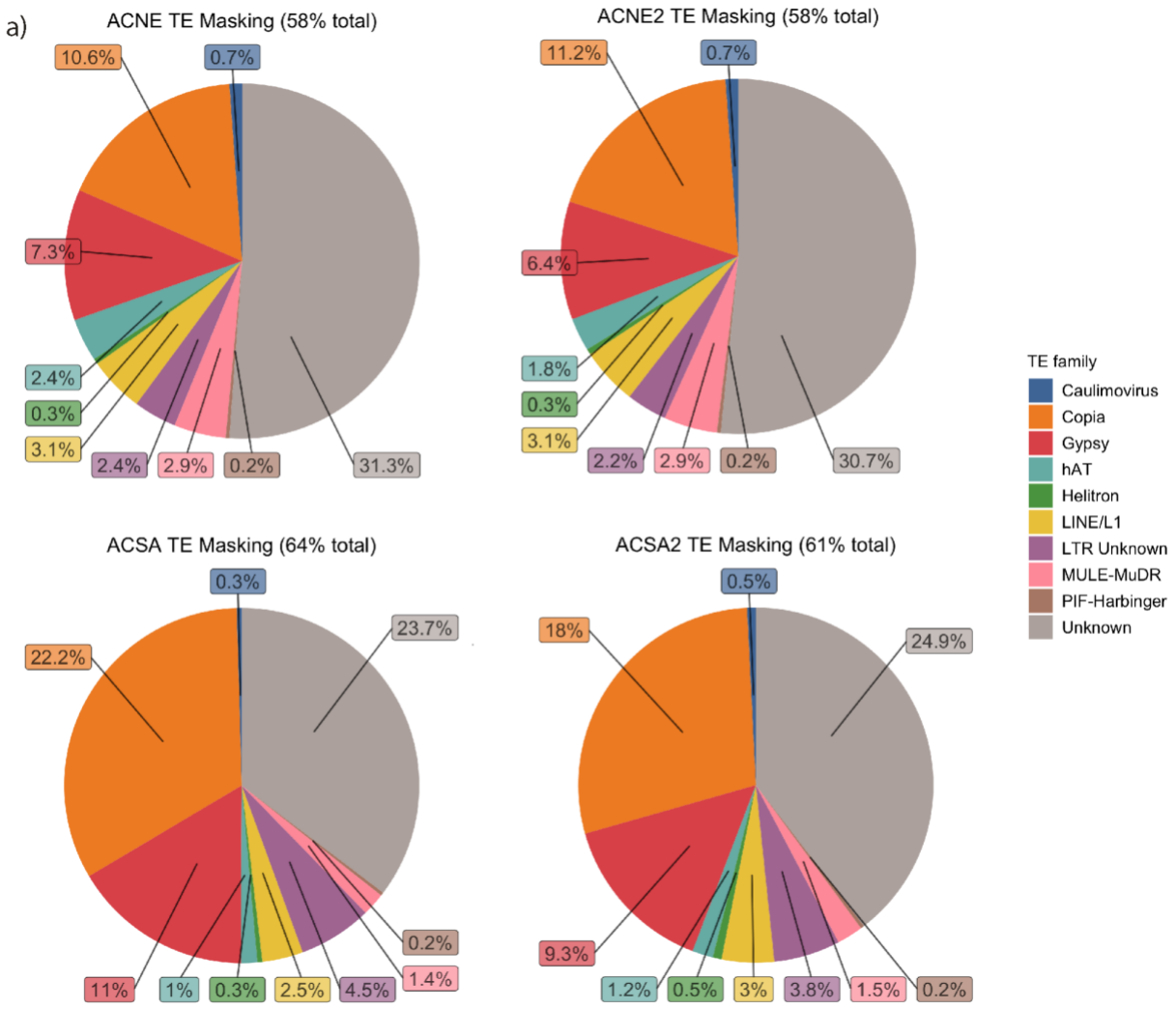
**Repeats**

LTR superfamilies Ty1-copia and Ty3-gypsy typically make up the greatest proportion of TEs in land plants. Insertions of LTRs in and around genes can be responsible for alternative splicing, duplication, recombination, and epigenetic control (Galindo-González et al. 2017). Biotic and abiotic stress, as well as other external stimuli that may result in polyploidization, can lead to TE movement and the rapid proliferation of several subfamilies (Mhiri, Borges, and Grandbastien 2022). The exact mechanisms of activation and repression are not fully understood, but generally involve methylation in any of three sequence contexts. Activation of TEs leads to an initial response of post-transcriptional gene silencing involving siRNA, which is followed by establishment and maintenance of silencing with DNA and histone methylation in RNA-directed DNA methylation (RdDM) pathways. This leads to chromatin modification making the TE inaccessible to transcriptional machinery (Erdmann and Picard 2020). Given this, the intersection of methylation and TEs can be informative for understanding control of expression and architectural changes to the genome that may play a role in evolution.

Repeat detection and classification were similar for both the new and original *A. negundo*. Whole-genome coverage was 58% for both, 20% as LTRs, with minor differences in the family representation. Among LTRs, Ty1-copias were present in the greatest amount (11.2%, 10.6%),

followed by Ty3-gypsy (6.4%, 7.3%). LTR repeats annotated as unknown represented only 2% of the *A. negundo* genome (Fig. 3a). In addition to Ty1-copia being the largest superfamily, it also appears to have had a burst of recent activity based on the amount with minimal sequence divergence (Fig. 3b). This subfamily was found in more copies in the new genome, which was otherwise very similar to the original. Compared to *A. negundo*, *A. saccharum* was more repetitive overall, at 61% in the new assembly and 64% in the original version, with LTRs representing 32% and 37% in the new and original assemblies, respectively. Repeat family representation followed the same general pattern of *A. negundo*, but with more Ty1-copia (18%, 22%) Ty3-gypsy (9.3%, 11%), and unknown LTRs (3.8%, 4.5%). Ty1-copia subfamilies with low divergence were even more abundant in these genomes, but the very recent burst noted in *A. negundo* was less significant.

In most plants, Ty3-gypsy elements are more abundant and more likely to insert in heterochromatic regions (Cossu et al. 2012) while Ty1-copia elements are typically more closely associated with genes, more transcriptionally active, and insert in a seemingly random pattern (Galindo-González et al. 2017; Qiu and Ungerer 2018). It should be noted that all other *Acer* recently characterized, including *A. pseudosieboldianum, A. catalpifolium, A. yangbiense,* and *A. truncatum*, demonstrated the same trend, with higher representation of Ty1-copia elements (J. Yang et al. 2019; Ma et al. 2020; Yu et al. 2021; X. Li et al. 2022). Notable exceptions to the Ty3-gypsy dominance have also been observed in cacao, grape, banana, and *Cucumis sativus* (Moisy et al. 2008; Argout et al. 2011; Castanera et al. 2019; Pratama, Dwivany, and Nugrahapraja 2021). As more genomes become available, more variation in these ratios has been noted within and across genera (Zagorski et al. 2020). It is possible that some of this new variation reflects improvements in long-read sequencing to resolve and quantify these elements, as seen with *Cucumis melo* (Castanera et al. 2019). Improvement of the *C. melo* genome with long reads modified the ratio of Ty3-gypsy to Ty1-copia elements in favor of Ty3-gypsy, compared to the previous publication (Ruggieri et al., 2018). In the maples, despite the fact that the original PacBio reads and the new nanopore reads shared similar read N50, the inclusion of some longer nanopore reads appeared to have improved the resolution of repeat space.


a)
ACNE TE Masking (58% total)
10.6%
0.7%
7.3%
2.4%
0.3%
3.1%
2.4%
2.9%
0.2%
31.3%
ACNE2 TE Masking (58% total)
11.2%
0.7%
6.4%
1.8%
0.3%
3.1%
2.2%
2.9%
0.2%
30.7%
ACSA TE Masking (64% total)
0.3%
23.7%
22.2%
0.2%
1.4%
11%
1%
0.3%
2.5%
4.5%
ACSA2 TE Masking (61% total)
0.5%
18%
24.9%
9.3%
1.2%
0.5%
3%
3.8%
1.5%
0.2%
TE family
Caulimovirus
Copia
Gypsy
hAT
Helitron
LINE/L1
LTR Unknown
MULE-MuDR
PIF-Harbinger
Unknown

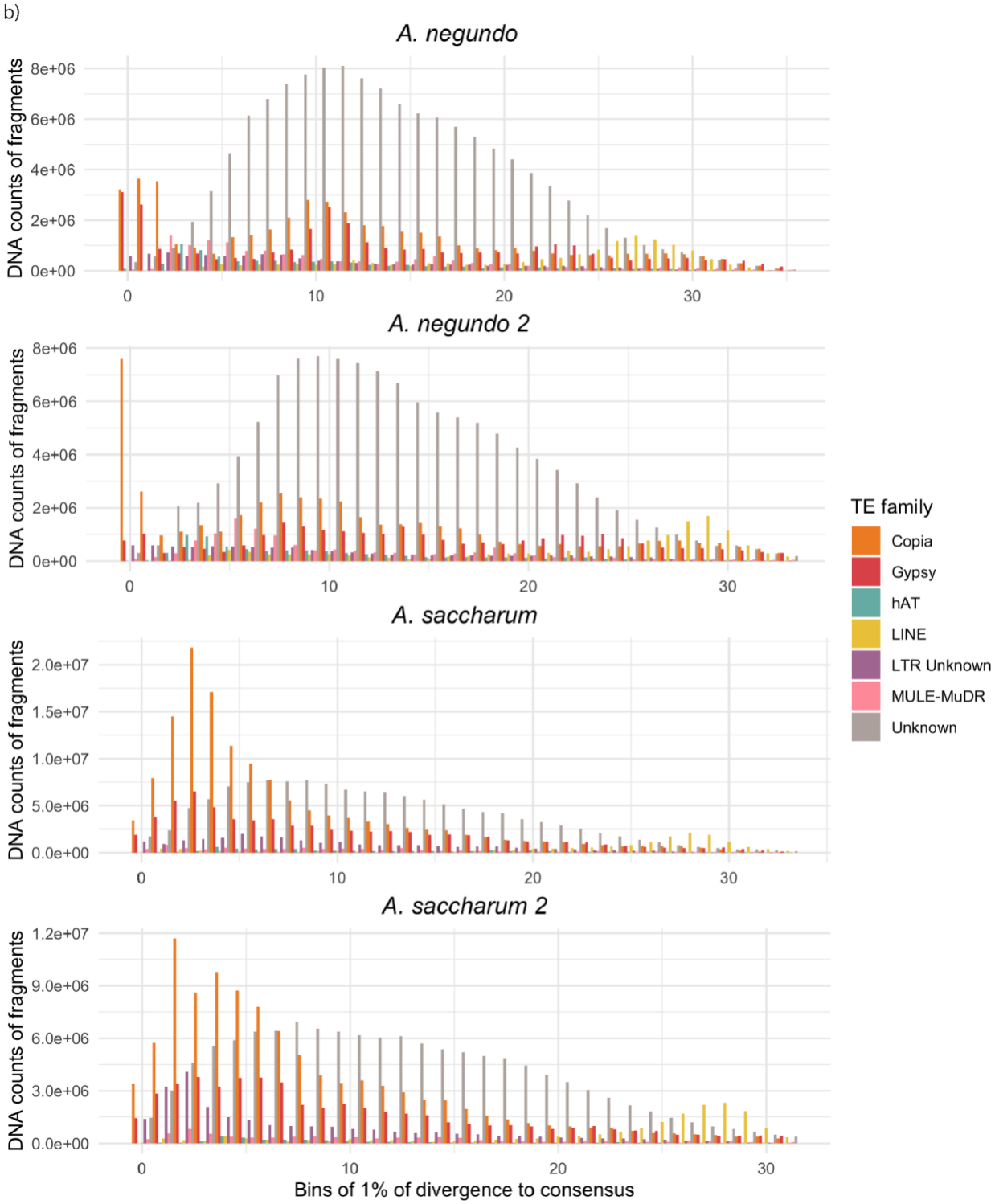



**Figure 3.** a) Genome-wide coverage by repeat class or superfamily for *A. negundo* (ACNE) and *A. saccharum* (ACSA). Only select groups are shown; see File S1 for a complete listing. b) Transposable element abundance for selected repeat families. Abundance of each repeat class or family is plotted in bins by its amount of divergence, with the least divergence representing putative recent activity.

**Methylation within *Acer***

Long-read sequencing for whole-genome methylation detection and quantification has not yet been widely adopted in plants. *Brassica nigra* was among the first plants on which such analyses were conducted, with nanopore reads enabling the demarcation of the centromeres (Perumal et al. 2020); an improved nanopore-sequenced radish genome provided the same centromere-level resolution (Cho et al. 2022). Both centromeres and stress responses were studied in *Gossypium thurberi* and *Gossypium davidsonii* through 5mC and 6mA sites (Yang et al. 2021). The first release of DeepSignal was used with nanopore reads to detect novel repeats and characterize the subtelomeres in algae (Chaux-Jukic et al. 2021).

The DeepSignal-Plant tool utilized nanopore data from *A. thaliana, Oryza sativa,* and *B. nigra* for development and benchmarking (Ni et al. 2021). This tool represents the first machine-learning approach for plants that can achieve accuracy for all three important 5mC states: CG, CHG, and CHH. CG methylation is most often located near and within gene bodies; CHG methylation plays a primary role in silencing TEs, and CHH is responsible for regulating both CG- and CHG-modified TEs (Ni et al. 2021). Studies on transgenerational inheritance of these forms of 5mC modifications in *Arabidopsis* determined that asymmetric CHH must be re-established de novo, while symmetric CG and CHG methylation is maintained (Hsieh, 2016). Using the consensus-based approach from METEORE, we combined DeepSignal-Plant with one of the best implementations of an HMM-based approach, Nanopolish, leveraging the strengths of both to detect methylation across all contexts in nanopore sequencing (Yuen et al. 2021).

Independently-assessed methylation levels for CG sites reported by Nanopolish and DeepSignal-Plant were both high relative to the consensus results (Fig. 4). DeepSignal-Plant detected the most CG-methylation across all four genomes, as was seen in benchmarking studies comparing DeepSignal (v1) and Nanopolish on *E. coli* and *H. sapiens* data (Ni et al. 2019). The CG levels for both the new and original *A. negundo* and original *A. saccharum* were quite similar (~70%), and the new *A. saccharum* genome was higher, at 75%. Nanopolish results followed a similar trend, but with slightly lower values of ~67%, and 71% for the new *A. saccharum* genome. Consensus results for all were around 53%. In particular, the original *A.*

*saccharum* genome was considerably lower, at 34%—close to the CHG value, which was unexpected.

The CHH and CGH values estimated by DeepSignal-Plant were in the 40% range but, again, the original *A. saccharum* genome value was lower than those of the other genomes, at 33%. CHH values were the lowest, and more consistent across all four, ranging from 9 to 11%.

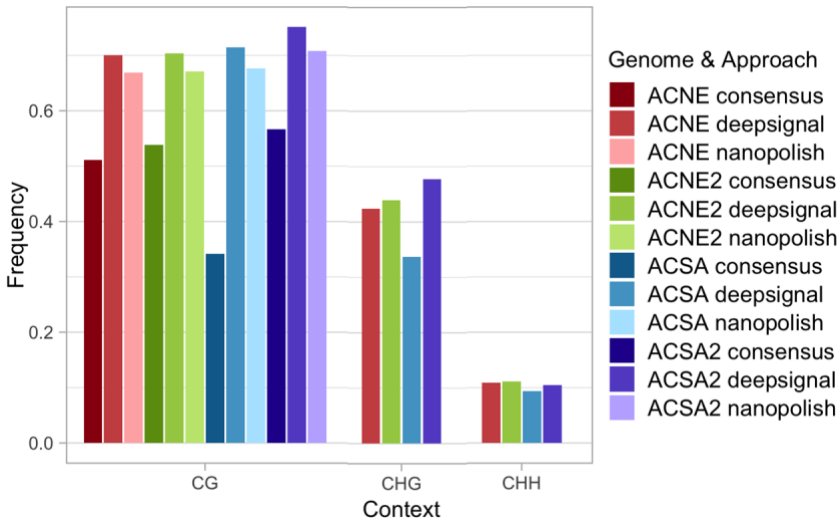



**Figure 4.** Comparison of global methylation levels by context for DeepSignal-Plant, Nanopolish, and the METEORE-generated random forest consensus.

The original *A. saccharum* genome resulted in much lower estimates of methylation compared to the other three genomes. This genome had the lowest retention of sequences during resquiggling, a preliminary step for DeepSignal-Plant that leverages the reference genome to correct base calling inaccuracies (Oxford Nanopore Technologies 2020b). When resquiggling the reads to the original and new *A. negundo* genomes, 21.2% and 18.0% reads were unsuccessfully processed. For *A. saccharum* these numbers were 54.3% and 31.5%, resulting in 43X coverage for the original genome (File S1). This could potentially have reduced coverage to levels affecting methylation calling in DeepSignal-Plant or Nanopolish; however, the primary impact was observed by the consensus statistics generated by METEORE (Fig. 4).

Given the reduced methylation levels for the consensus using the original *A. saccharum*, the methylation datasets based on the new genomes were selected as the best representation. As such, downstream analyses were conducted with the METEORE consensus results. Figures displaying results for both new and original genomes are available in the supplementary material (Fig. S1–5).

**Comparative methylomes**

Independent estimates from DeepSignal-Plant and Nanopolish were on the higher end of the spectrum of global methylation levels reported for 34 angiosperms in Niederhuth et al. (2016). Consensus results were more similar, but still higher than many angiosperms —similar in total to *Fragaria vesca*, *Manihot esculenta*, *Vitis vinifera*, *Brachypodium distachyon*, and *Setaria viridis*. WGBS in the form of MethylC-Seq was used in the Niederhuth et al. (2016) study, so increased signal detection could result from the long-reads, which are able to span low-complexity regions. A recent study compared short-read and long-read assemblies of the *Brassica* genome with both WGBS and nanopore-based approaches (Nanopolish). The direct

CG methylation profiling using the nanopore reads was strongly correlated (>93%) with their WGBS data for the same accession. In addition, centromeric repeats and were identified using nanopore long-reads and accompanying hypermethylated signal (Perumal et al. 2020).

The distribution of methylation, genes, and TEs is of interest due to the implications for evolutionary strategies involving methylation's regulation of genes, TEs, and genes neighboring TEs. Many plant species exhibit a pattern of gene density along the arms of the chromosome and TE density in centromeric regions (Comai et al., 2017). As an exception, Sork et al. (2022) demonstrated that *Q. lobata* has a more uniform distribution of genes and CHH methylation along the chromosome arm, similar to several *Poaceae*. This pattern of distribution, resulting in more intermixed genes, methylation, and TEs, perhaps hints at regulatory strategies not found in most angiosperms that enabled its widespread biogeographic success. The Pearson correlation (R) of genes to methylation for each sequence context was plotted along with the previous analyses (Niederhuth et al. 2016; Sork et al. 2022) (Fig. 5a). With the inclusion of *Acer*, there is a moderate negative correlation between CHH methylation and genes—almost 50%—similar to about half of angiosperms surveyed, while the remaining angiosperm species have stronger negative correlation (~75%). CHH was the context with lowest negative correlation of genes to methylation for *Acer*. Less correlation of methylation with genes means more intersection with TEs in intergenic space, lending support to the hypothesis that the CHH context is more frequently associated with TE regulation in less gene-dense regions, and perhaps has more of a role regulating key genes at methylated TE boundary regions.

Previous analysis of CHH subcontexts across chromosomes detected differences between *Q. lobata*, in which CHH is generally distributed, and *Populus trichocarpa*, in which CHH is localized around the centromere and decreases across the gene-dense chromosome arms (Sork et al. 2022). *Acer* presents an intermediate pattern, particularly *A. saccharum* (Fig. 5b, Fig. S1). *A. saccharum* does not have the same observable enrichment of CHH adjacent to centromeres as *P. trichocarpa*, nor is CHH as generally distributed as in *Q. lobata*. In both *Acer* species, there is a clear pattern for a preference of CTA, followed by CAA, along the lengths of the chromosome arms. Cytosine methyltransferases are known to have sequence context preferences. In *Arabidopsis,* CMT2 prefers CHH sites, specifically CAA and CTA. However, little is known about these preferences outside of a handful of model species (Kenchanmane Raju, Ritter, and Niederhuth 2019).

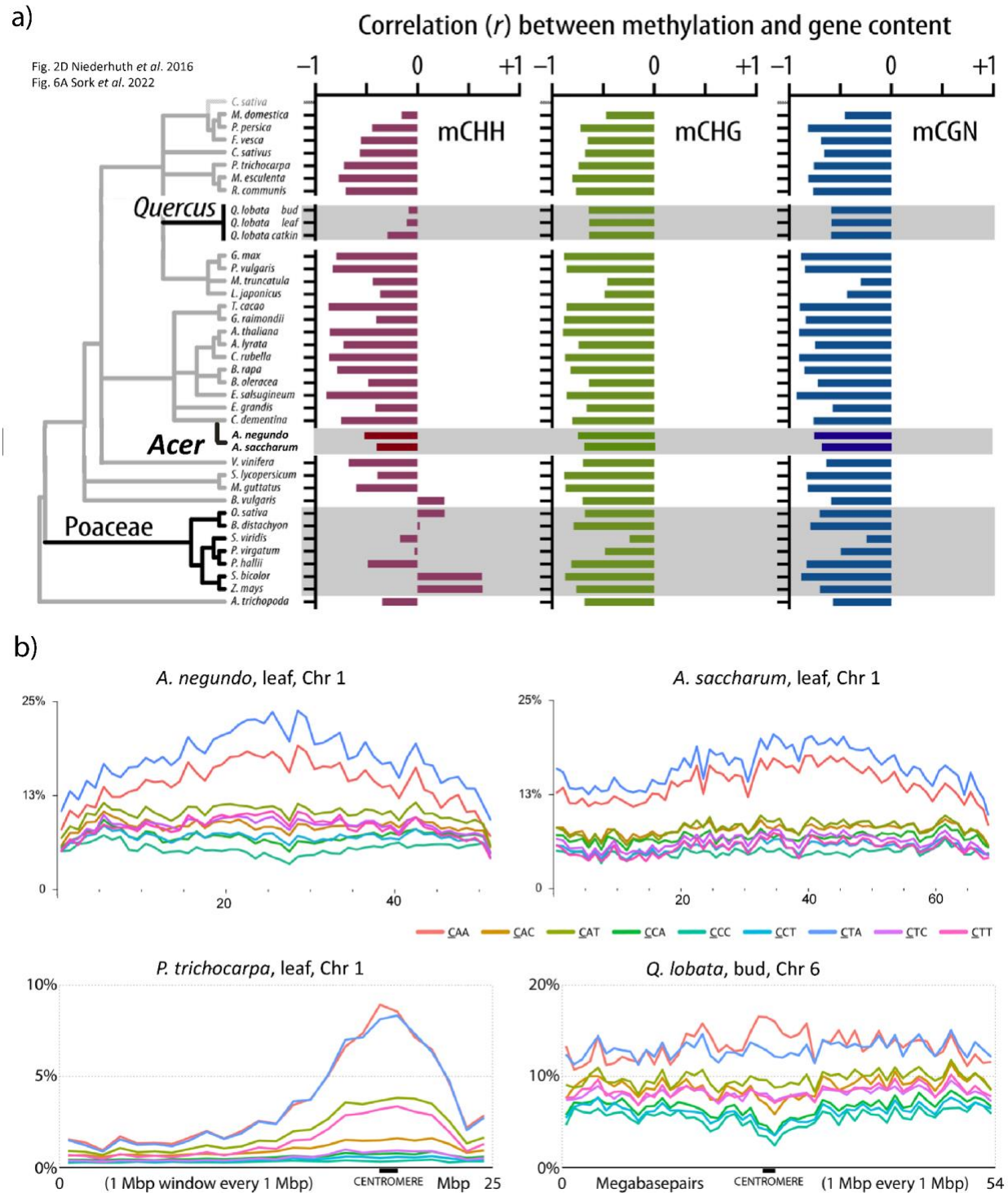



**Figure 5.** a) Modified from Sork et al. (2022) to add the *Acer* genomes. *Acer* has a negative correlation between methylation and gene content, similar to other species. In contrast, *Quercus* or *Poaceae* are less negatively correlated, or even positively correlated in some species. b) Distribution of CHH methylation by subcontext across the longest chromosomes for *A. negundo* and *A. saccharum*. *Populus* and *Q. lobata*, reprinted from Sork et al. (2022), are shown for comparison of distribution patterns, where *Populus* is

localized and *Q. lobata* is generally distributed. Distributions for all chromosomes are in Fig. S1. CHG distributions are in Fig. S2 and gene densities plotted with all contexts in 100 Kb windows are in Fig. S3.

Methylation across repeat superfamilies was also plotted along with results from Sork et al. (2022) for comparison (Fig. 6a, b). The repeat analysis of the *Acer* found Ty1-copia was present in greater amounts than Ty3-gypsy, 4.8% more coverage in *A. negundo* and 8.7% in *A. saccharum*, and much of it recent based on diversity estimates. There was a higher frequency of methylation over Ty3-gypsy elements in *A. negundo*, particularly for the CHH context. In both species, CG and CHG methylation were higher among Ty1-copias. SINEs were heavily methylated (>40% for CHH) in *Quercus* but SINEs were not clearly identified in the *Acer* for comparison. On the other hand, LINEs were highly methylated (CG and CHG) in *Acer*.

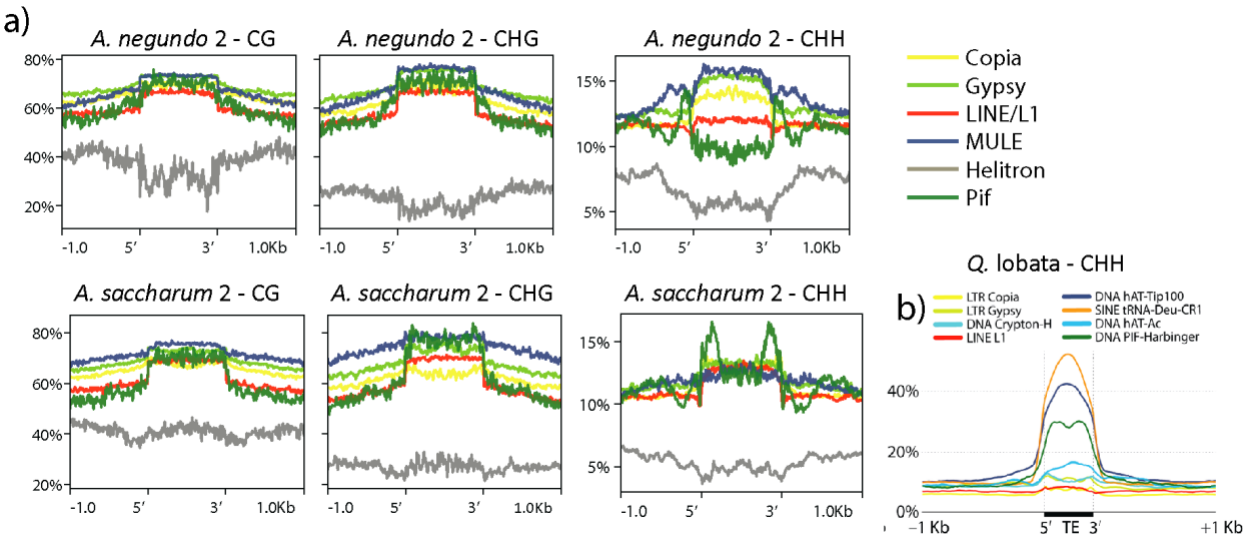



**Figure 6.** a) Genome-wide methylation distribution by sequence context and TE family. b) Genome-wide CHH distribution (bud tissue) over TE regions, each line representing one of the largest superfamilies of each class. Reprinted from Sork et al. (2022).

Plots of methylation frequencies across gene regions for all three contexts were somewhat different than that observed in *Q. lobata*, but there is a spectrum of variation in frequency distributions across gene regions in different angiosperms, as seen in Niederhuth et al. (2016) (Fig. 7a, c). The two *Acer* species were very similar, but *A. negundo*, which has the shorter genome, had a slightly higher frequency of CG. The higher frequencies across genes in the original genomes compared to the new genomes may be due to differences in average gene length, as seen in *Q. lobata* (Fig. 7c, Fig. S4). Both *Acer* species also shared the same sharp increase in CHH just upstream of the transcription start site, which was greater in *A. negundo*.

CHH methylation is not retained due to a lack of symmetry and must be established with each new generation (Hsieh, 2016). Studies on CHH methylation in *Arabidopsis* have characterized the unique pathways responsible, and their TE targets (Bouyer et al., 2017). The RdDM pathway targets Class I TEs, specifically RC/Helitron and DNA/MuDR, and the CMT2 pathway targets the LTRs (primarily Ty3-gypsy and Ty1-copia) (Sasaki et al. 2019). Focusing on CHH levels across genic regions, higher frequencies were found in portions of the upstream flanking

region (Fig. 7a). The next largest fraction was in the first intron region, where both new *Acer* genomes had a significant drop in number of methylated sites, especially for those at 90% methylation (Fig. S5). This pattern was seen in other introns, but at lower levels, while the levels in downstream flanking regions were consistent across frequencies. In certain plant genes—often highly expressed genes—the first intron can contain regulatory elements, though it is observed much less frequently than upstream promoters (Rose 2018).

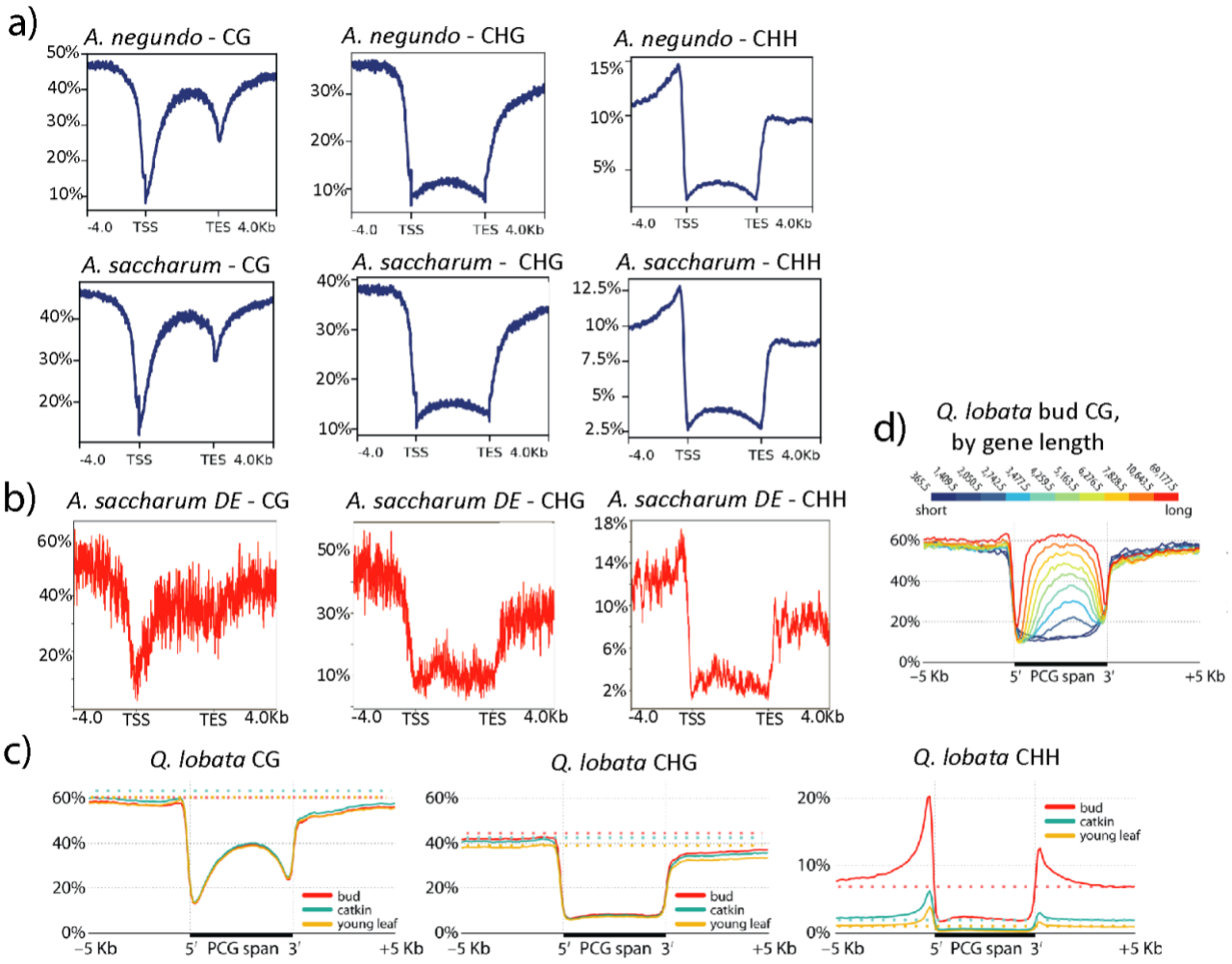



**Figure 7.** a) Methylation frequency distribution across protein coding regions, shown by assembly and sequence context. TSS = transcription start site; TES = transcription end site, b) methylation frequency distribution across 240 (of the original 245) genes differentially expressed in response to calcium and aluminum treatments in stem as seen in McEvoy et al. (2021), c) methylation frequency distribution for *Q. lobata* across protein coding regions (PCG) and d) CG for *Q. lobata* genes in deciles by gene length, reprinted from Sork et al. (2022).

### Methylome and gene regulation

By combining distributions of methylation, select TEs, and gene density, trends amongst the elements can be observed. In Fig. 8, each row contains chromosomes that are largely syntenic between *A. negundo* (ACNE) and *A. saccharum* (ACSA), as seen in Fig. 2b. Regions of low

gene density are co-located with peaks of LTRs and methylation in all contexts, including CHH. It is likely these contain centromeres and pericentromeric regions. Included in the gene density track (Fig. 8) are gene families previously identified as significantly expanding (26 in *A. negundo*, 99 in *A. saccharum*) or significantly contracting (52, 18) when evaluated in terms of gene family evolution across 22 land plant species (McEvoy et al. 2021).

This particular subset of syntenic chromosomes provides a few different examples of rapid gene family evolution in chromosomes. In the first row, the two chromosomes are inverse in terms of sequence, but largely similar, with no previously identified gene family evolution. The middle row shows many instances of gene family expansion in *A. saccharum*, which tends to be the predominant pattern throughout the full set of chromosomes. In the bottom row, *A. negundo* has more evidence of gene family dynamics. In the previous *Acer* study, *A. saccharum* was shown to have more families associated with rapid expansion, while *A. negundo* primarily had contracting families. This analysis shows that the distribution is generalized to the whole genome and does not seem to be restricted to specific chromosomes. The exception is Chr 8, which has numerous expanding families but no associated changes in methylation in any context observable within the window used to generate the chromosome-scale visualization (Fig. S6).

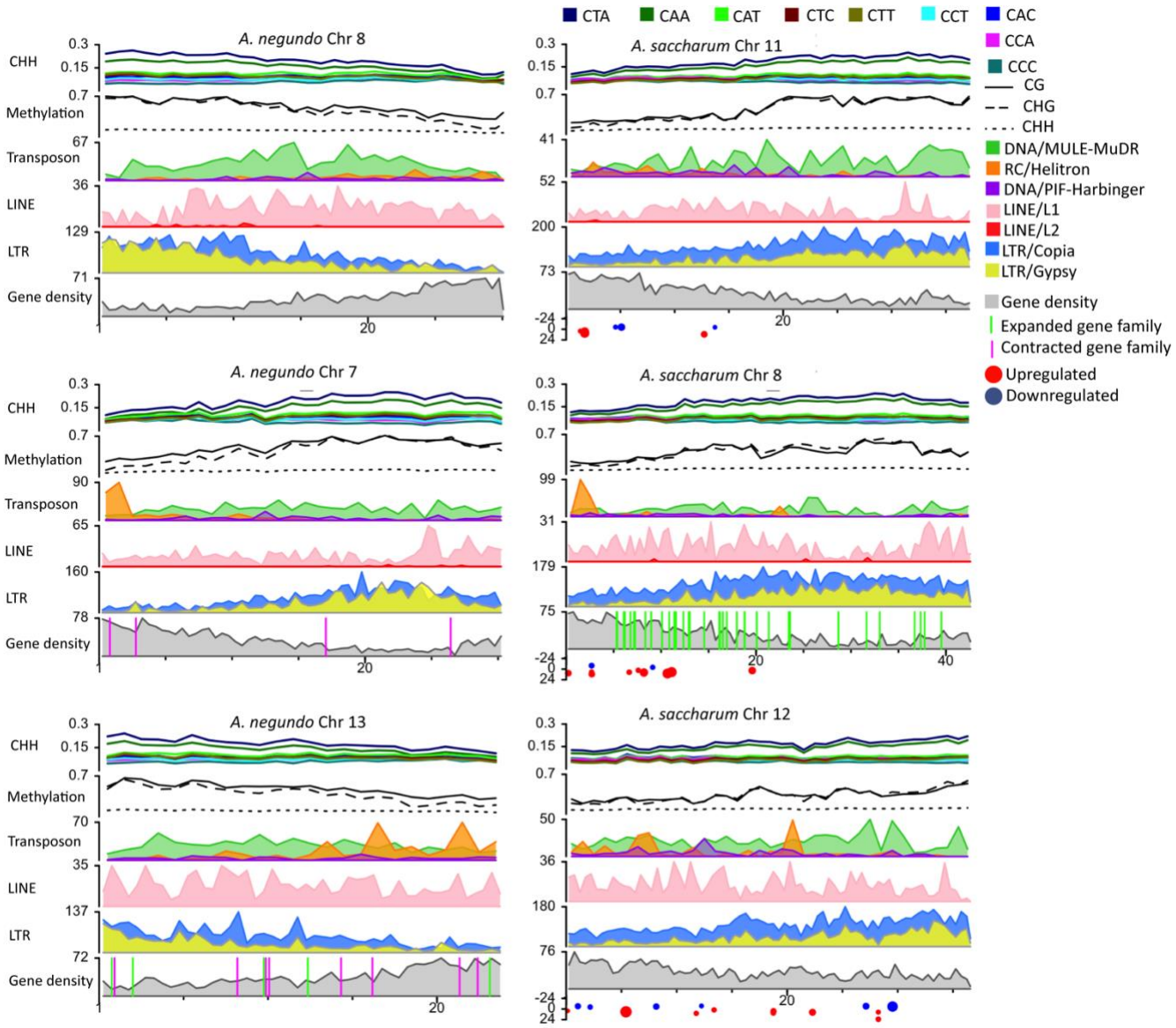



**Figure 8.** Distribution of gene density, repeat families, methylation by context, CHH methylation subcontexts, gene family dynamics, and gene expression results for select chromosomes. Middle row, right, shows multiple gene family expansions (lime green vertical bars) in *A. saccharum*. Bottom row shows gene family expansions and contraction (purple bars) in *A. negundo*. Gene expression results are from aluminum and calcium treatments at Hubbard Brook Experimental Forest as detailed in McEvoy et al. (2021). X-axis indicates log2 fold change, while dot size represents the *p*-adjusted value.

Several studies have examined the relationship between methylation changes and metal toxicity, and the associated nutrient stress, in both model and non-model plant systems. One such study examined the aquatic plant *Hydrilla verticillata* and noted demethylation in response to copper, as well as blocked ROS interactions that could cause remethylation (Shi et al. 2017). In *Hibiscus cannabinus*, increased methylation was associated with increasing chromium levels (Ding et al. 2016). In *Arabidopsis halleri*, cadmium treatments increased CpG methylation of

genes responsible for symmetric methylation (Galati et al. 2021). In rice, a model for nutrient stress, the expression of Heavy Metal Transporting ATPases (HMAs) was modified and retained over generations and modulated by the methylation of specific TEs (Cong et al. 2019). While our study does not lend itself to proper examination of species-specific patterns of methylation that can be correlated to gene expression changes, the potential for gene regulation among candidate genes, previously identified in RNA-Seq experiments in *A. saccharum,* could be examined. Gene expression results focused on candidates associated with nutrient stress (or heavy metal toxicity) were mapped to the new *A. saccharum* genome (240 in total; Figure 8). A total of 115 genes were downregulated in trees grown in aluminum-amended plots while 130 genes were up-regulated (in both cases, compared with trees grown in calcium-amended or control plots). These genes are distributed across all chromosomes at different densities. To further examine the patterns of methylation in all three contexts, the 240 differentially expressed genes were compared with the patterns observed for the full gene space in Figure 6B. The subset of differentially expressed genes appeared enriched for CG and CHG methylation in the upstream regions when compared with the full gene space. The most noticeable difference, however, was the higher frequency of CHH methylation in the promoter region of differentially expressed genes. Values here peak at around 16% compared to 12.5% for the whole genome mean. This higher value matches what is seen in *Q. lobata* bud tissue relative to catkins or young leaves, which have a greater portion of meristematic tissue, as mentioned in Sork et al. (2022), as well as undergoing developmental processes. Increased methylation could be a sign of gene networks requiring more flexibility in expression to meet the challenges of development or biotic and abiotic stress (Lang et al., 2017).

## CONCLUSION

Integration of epigenomic data is important for a full understanding of the genomic mechanics in effect, particularly in long-lived plant species, which are highly dependent on epigenetic strategies for adaptation. This study developed two improved reference genomes for two new accessions of *A. saccharum* and *A. negundo*. These hybrid assemblies benefitted from the inclusion of deep nanopore coverage. Methylation was conducted with a custom pipeline that leveraged improved base calling and considerations for false positives. Methylation frequencies and distributions were compared with other recent broadleaf tree methylomes and uncovered clear differences among the species. Differences in repeat content across the species was reflected in the patterns of methylation observed. Preliminary analysis of candidate genes associated with nutrient stress in *A. saccharum* were evaluated and patterns of methylation were variable, with increased upstream methylation observed in all three contexts. Further investigations require parallel expression and tissue-specific studies, as well as pan-genome (population scale) analysis to understand how the methylome is contributing to genome evolution.

## DATA AVAILABILITY

New and original *Acer saccharum* sequencing, assemblies, and annotations are available in BioProject PRJNA748028. *Acer negundo* data is available in BioProject PRJNA750066. Scripts

are available in the Plant Computational Genomics GitLab, AcerMethylation repository https://gitlab.com/PlantGenomicsLab/acermethylation/

ACKNOWLEDGEMENTS
We would like to thank the Arnold Arboretum of Harvard University for providing leaf samples that were difficult to acquire during the pandemic. We thank the Institute for Systems Genomics Center for Genome Innovation for molecular resources and support. We would also like to thank the Computational Biology Core for access to HPC resources. We would also like to thank members of the University of Connecticut MCB and EEB departments, including Christine McCann who advised on extractions and library prep, and the many who provided collaborative brainstorming on methylation calling methods, including Savannah Hoyt, Gabby Hartley, Michelle Neitzey, Vidya Vuruputoor, and Ross Whetten at NCSU.

FUNDING
Funding was provided by the Botanical Society of America, Bill Dahl Graduate Student Research Award and The Ronald Bamford Endowment Fund for Botany Research to the Department of Ecology and Evolutionary Biology. PGSG was supported by National Institutes of Health R01GM123312-02 to R. O'Neill.

## REFERENCES


Alonge, Michael, Ludivine Lebeigle, Melanie Kirsche, Sergey Aganezov, Xingang Wang, Zachary B. Lippman, Michael C. Schatz, and Sebastian Soyk. 2021. “Automated Assembly Scaffolding Elevates a New Tomato System for High-Throughput Genome Editing.” *bioRxiv*. https://doi.org/10.1101/2021.11.18.469135.

Argout, Xavier, Jerome Salse, Jean-Marc Aury, Mark J. Guiltinan, Gaetan Droc, Jerome Gouzy, Mathilde Allegre, et al. 2011. “The Genome of Theobroma Cacao.” *Nature Genetics* 43 (2): 101–8.

Bishop, Daniel A., Colin M. Beier, Neil Pederson, Gregory B. Lawrence, John C. Stella, and Timothy J. Sullivan. 2015. “Regional Growth Decline of Sugar Maple (Acer Saccharum) and Its Potential Causes.” *Ecosphere* 6 (10): art179.

Brůna, Tomáš, Katharina J. Hoff, Alexandre Lomsadze, Mario Stanke, and Mark Borodovsky. 2021. “BRAKER2: Automatic Eukaryotic Genome Annotation with GeneMark-EP+ and AUGUSTUS Supported by a Protein Database.” *NAR Genomics and Bioinformatics* 3 (1): lqaa108.

Bouyer, D., Kramdi, A., Kassam, M., Heese, M., Schnittger, A., Roudier, F., & Colot, V. (2017). DNA methylation dynamics during early plant life. *Genome Biology*, *18*(1), 179.

Caballero, Madison, and Jill Wegrzyn. 2019. “gFACs: Gene Filtering, Analysis, and Conversion to Unify Genome Annotations Across Alignment and Gene Prediction Frameworks.” *Genomics, Proteomics & Bioinformatics* 17 (3): 305–10.

Castanera, Raúl, Valentino Ruggieri, Marta Pujol, Jordi Garcia-Mas, and Josep M. Casacuberta. 2019. “An Improved Melon Reference Genome with Single-Molecule Sequencing Uncovers a Recent Burst of Transposable Elements With Potential Impact on Genes.” *Frontiers in Plant Science* 10: 1815.

Cavrak, Vladimir V., Nicole Lettner, Suraj Jamge, Agata Kosarewicz, Laura Maria Bayer, and Ortrun Mittelsten Scheid. 2014. “How a Retrotransposon Exploits the Plant’s Heat Stress Response for Its Activation.” *PLoS Genetics* 10 (1): e1004115.

Chaux-Jukic, Frédéric, Samuel O’Donnell, Rory J. Craig, Stephan Eberhard, Olivier Vallon, and Zhou Xu. 2021. “Architecture and Evolution of Subtelomeres in the Unicellular Green Alga Chlamydomonas Reinhardtii.” *Nucleic Acids Research* 49 (13): 7571–87.

Chen, Shifu, Yanqing Zhou, Yaru Chen, and Jia Gu. 2018. “Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor.” *Bioinformatics* 34 (17): i884–90.

Cho, Ara, Hoyeol Jang, Seunghoon Baek, Moon-Jin Kim, Bomi Yim, Sunmi Huh, Song-Hwa Kwon, Hee-Ju Yu, and Jeong-Hwan Mun. 2022. “An Improved Raphanus Sativus Cv. WK10039 Genome Localizes Centromeres, Uncovers Variation of DNA Methylation and Resolves Arrangement of the Ancestral Brassica Genome Blocks in Radish Chromosomes.” *Theoretical and Applied Genetics* 135 (5): 1731–50.

Ci, Dong, Yuepeng Song, Qingzhang Du, Min Tian, Shuo Han, and Deqiang Zhang. 2016. “Variation in Genomic Methylation in Natural Populations of Populus Simonii Is Associated with Leaf Shape and Photosynthetic Traits.” *Journal of Experimental Botany* 67 (3): 723–37.

Comai, L., Maheshwari, S., & Marimuthu, M. P. A. (2017). Plant centromeres. *Current Opinion in Plant Biology*, *36*, 158–167.

Cong, Weixuan, Yiling Miao, Lei Xu, Yunhong Zhang, Chunlei Yuan, Junmeng Wang, Tingting Zhuang, et al. 2019. “Transgenerational Memory of Gene Expression Changes Induced by Heavy Metal Stress in Rice (Oryza Sativa L.).” *BMC Plant Biology* 19 (1): 282.

Ding, Han, Guodong Wang, Lili Lou, and Jinyin Lv. 2016. “Physiological Responses and Tolerance of Kenaf (Hibiscus Cannabinus L.) Exposed to Chromium.” *Ecotoxicology and Environmental Safety* 133 (November): 509–18.

Erdmann, Robert M., and Colette L. Picard. 2020. “RNA-Directed DNA Methylation.” *PLoS Genetics* 16 (10): e1009034.

Fan, Xiaoru, Lirun Peng, and Yong Zhang. 2022. “Plant DNA Methylation Responds to Nutrient Stress.” *Genes* 13 (6): 992.

Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. “RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (17): 9451–57.

Galati, Serena, Mariolina Gullì, Gianluigi Giannelli, Antonella Furini, Giovanni DalCorso, Rosaria Fragni, Annamaria Buschini, and Giovanna Visioli. 2021. “Heavy Metals Modulate DNA Compaction and Methylation at CpG Sites in the Metal Hyperaccumulator Arabidopsis Halleri.” *Environmental and Molecular Mutagenesis* 62 (2): 133–42.

Galindo-González, Leonardo, Corinne Mhiri, Michael K. Deyholos, and Marie-Angèle Grandbastien. 2017. “LTR-Retrotransposons in Plants: Engines of Evolution.” *Gene* 626 (August): 14–25.

Gel, Bernat, and Eduard Serra. 2017. “karyoploteR: An R/Bioconductor Package to Plot Customizable Genomes Displaying Arbitrary Data.” *Bioinformatics* 33 (19): 3088–90.

Gouil, Quentin, and Andrew Keniry. 2019. “Latest Techniques to Study DNA Methylation.” *Essays in Biochemistry* 63 (6): 639–48.

Hart, Alexander J., Samuel Ginzburg, Muyang (sam) Xu, Cera R. Fisher, Nasim Rahmatpour, Jeffry B. Mitton, Robin Paul, and Jill L. Wegrzyn. 2018. “EnTAP: Bringing Faster and Smarter Functional Annotation to Non-Model Eukaryotic Transcriptomes.” *bioRxiv*, April, 307868.

He, Li, Huan Huang, Mariem Bradai, Cheng Zhao, Yin You, Jun Ma, Lun Zhao, Rosa Lozano-Durán, and Jian-Kang Zhu. 2022. “DNA Methylation-Free Arabidopsis Reveals Crucial Roles of DNA Methylation in Regulating Gene Expression and Development.” *Nature Communications* 13 (1): 1335.

He, Li, Wenwu Wu, Gaurav Zinta, Lan Yang, Dong Wang, Renyi Liu, Huiming Zhang, et al. 2018. “A Naturally Occurring Epiallele Associates with Leaf Senescence and Local Climate Adaptation in Arabidopsis Accessions.” *Nature Communications* 9 (1): 460.

Hofmeister, Brigitte T., Johanna Denkena, Maria Colomé-Tatché, Yadollah Shahryary, Rashmi Hazarika, Jane Grimwood, Sujan Mamidi, et al. 2020. “A Genome Assembly and the Somatic Genetic and Epigenetic Mutation Rate in a Wild Long-Lived Perennial Populus Trichocarpa.” *Genome Biology* 21 (1): 259.

Hsieh, P.-H. (2016). *Maintenance and Inheritance of DNA Methylation in Arabidopsis* [UC Berkeley]. https://escholarship.org/uc/item/9r85v5k7

Jullien, Pauline E., Daichi Susaki, Ramesh Yelagandula, Tetsuya Higashiyama, and Frédéric Berger. 2012. “DNA Methylation Dynamics during Sexual Reproduction in Arabidopsis

Thaliana." *Current Biology: CB* 22 (19): 1825–30.
Kenchanmane Raju, Sunil K., Eleanore Jeanne Ritter, and Chad E. Niederhuth. 2019. "Establishment, Maintenance, and Biological Roles of Non-CG Methylation in Plants." *Essays in Biochemistry* 63 (6): 743–55.
Kim, Daehwan, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. 2016. "Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences." *Genome Research* 26 (12): 1721–29.
Kumar, Suresh, and Trilochan Mohapatra. 2021. "Dynamics of DNA Methylation and Its Functions in Plant Growth and Development." *Frontiers in Plant Science* 12 (May): 596236.
Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., Zhang, Y., Handa, A. K., & Zhu, J.-K. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(22), E4511–E4519.
Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25.
Liang, Lixiong, Yingying Chang, Junqian Lu, Xiaojuan Wu, Qi Liu, Weixi Zhang, Xiaohua Su, and Bingyu Zhang. 2019. "Global Methylomic and Transcriptomic Analyses Reveal the Broad Participation of DNA Methylation in Daily Gene Expression Regulation of Populus Trichocarpa." *Frontiers in Plant Science* 10 (February): 243.
Li, Heng. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34 (18): 3094–3100.
Liu, Yang, Wojciech Rosikiewicz, Ziwei Pan, Nathaniel Jillette, Ping Wang, Aziz Taghbalout, Jonathan Foox, et al. 2021. "DNA Methylation-Calling Tools for Oxford Nanopore Sequencing: A Survey and Human Epigenome-Wide Evaluation." *Genome Biology* 22 (1): 295.
Li, Xiang, Kewei Cai, Zhiming Han, Shikai Zhang, Anran Sun, Ying Xie, Rui Han, et al. 2022. "Chromosome-Level Genome Assembly for Acer Pseudosieboldianum and Highlights to Mechanisms for Leaf Color and Shape Change." *Frontiers in Plant Science* 13 (March): 850054.
Loman, Nicholas J., Joshua Quick, and Jared T. Simpson. 2015. "A Complete Bacterial Genome Assembled de Novo Using Only Nanopore Sequencing Data." *Nature Methods* 12 (8): 733–35.
Ma, Qiuyue, Tianlin Sun, Shushun Li, Jing Wen, Lu Zhu, Tongming Yin, Kunyuan Yan, et al. 2020. "The Acer Truncatum Genome Provides Insights into Nervonic Acid Biosynthesis." *The Plant Journal: For Cell and Molecular Biology* 104 (3): 662–78.
McEvoy, Susan L., U. Uzay Sezen, Alexander Trouern-Trend, Sean M. McMahon, Paul G. Schaberg, Jie Yang, Jill L. Wegrzyn, and Nathan G. Swenson. 2021. "Strategies of Tolerance Reflected in Two North American Maple Genomes." *The Plant Journal: For Cell and Molecular Biology*, December. https://doi.org/10.1111/tpj.15657.
Mhiri, Corinne, Filipe Borges, and Marie-Angèle Grandbastien. 2022. "Specificities and Dynamics of Transposable Elements in Land Plants." *Biology* 11 (4). https://doi.org/10.3390/biology11040488.
Miryeganeh, Matin, Ferdinand Marlétaz, Daria Gavriouchkina, and Hidetoshi Saze. 2022. "De

Novo Genome Assembly and in Natura Epigenomics Reveal Salinity-Induced DNA Methylation in the Mangrove Tree Bruguiera Gymnorhiza." *The New Phytologist* 233 (5): 2094–2110.

Moisy, Cédric, Keith E. Garrison, Carole P. Meredith, and Frédérique Pelsy. 2008. "Characterization of Ten Novel Ty1/copia-like Retrotransposon Families of the Grapevine Genome." *BMC Genomics* 9 (October): 469.

Niederhuth, Chad E., Adam J. Bewick, Lexiang Ji, Magdy S. Alabady, Kyung Do Kim, Qing Li, Nicholas A. Rohr, et al. 2016. "Widespread Natural Variation of DNA Methylation within Angiosperms." *Genome Biology* 17 (1): 194.

Ni, Peng, Neng Huang, Fan Nie, Jun Zhang, Zhi Zhang, Bo Wu, Lu Bai, et al. 2021. "Genome-Wide Detection of Cytosine Methylations in Plant from Nanopore Sequencing Data Using Deep Learning." *bioRxiv*. https://doi.org/10.1101/2021.02.07.430077.

Ni, Peng, Neng Huang, Zhi Zhang, De-Peng Wang, Fan Liang, Yu Miao, Chuan-Le Xiao, Feng Luo, and Jianxin Wang. 2019. "DeepSignal: Detecting DNA Methylation State from Nanopore Sequencing Reads Using Deep-Learning." *Bioinformatics* 35 (22): 4586–95.

O'Leary, Nuala A., Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, et al. 2016. "Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation." *Nucleic Acids Research* 44 (D1): D733–45.

Oxford Nanopore Technologies. 2020a. *Guppy*. Github. https://github.com/nanoporetech.

———. 2020b. *Tombo* (version 1.5.1). https://github.com/nanoporetech/tombo.

———. 2021. *ont_fast5_api* (version 4.0.0). https://github.com/nanoporetech/ont_fast5_api.

Perumal, Sampath, Chu Shin Koh, Lingling Jin, Miles Buchwaldt, Erin E. Higgins, Chunfang Zheng, David Sankoff, et al. 2020. "A High-Contiguity Brassica Nigra Genome Localizes Active Centromeres and Defines the Ancestral Brassica Genome." *Nature Plants* 6 (8): 929–41.

Pratama, Sigit Nur, Fenny Martha Dwivany, and Husna Nugrahapraja. 2021. "Comparative Genomics of Copia and Gypsy Retroelements in Three Banana Genomes: A, B, and S Genomes." *Pertanika Journal of Tropical Agricultural Science* 44 (4). https://doi.org/10.47836/pjtas.44.4.01.

Qiu, Fan, and Mark C. Ungerer. 2018. "Genomic Abundance and Transcriptional Activity of Diverse Gypsy and Copia Long Terminal Repeat Retrotransposons in Three Wild Sunflower Species." *BMC Plant Biology* 18 (1): 6.

Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." *Bioinformatics* 26 (6): 841–42.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org/.

Ritter, Eleanore J., and Chad E. Niederhuth. 2021. "Intertwined Evolution of Plant Epigenomes and Genomes." *Current Opinion in Plant Biology* 61 (June): 101990.

Rose, Alan B. 2018. "Introns as Gene Regulators: A Brick on the Accelerator." *Frontiers in Genetics* 9: 672.

Ruggieri, V., Alexiou, K. G., Morata, J., Argyris, J., Pujol, M., Yano, R., Nonaka, S., Ezura, H., Latrasse, D., Boualem, A., Benhamed, M., Bendahmane, A., Cigliano, R. A., Sanseverino, W., Puigdomènech, P., Casacuberta, J. M., & Garcia-Mas, J. (2018). An improved

assembly and annotation of the melon (Cucumis melo L.) reference genome. *Scientific Reports*, *8*(1), 8088.

Sasaki, Eriko, Taiji Kawakatsu, Joseph R. Ecker, and Magnus Nordborg. 2019. “Common Alleles of CMT2 and NRPE1 Are Major Determinants of CHH Methylation Variation in Arabidopsis Thaliana.” *PLoS Genetics* 15 (12): e1008492.

Shi, Danlu, Kai Zhuang, Yan Xia, Changhua Zhu, Chen Chen, Zhubing Hu, and Zhenguo Shen. 2017. “Hydrilla Verticillata Employs Two Different Ways to Affect DNA Methylation under Excess Copper Stress.” *Aquatic Toxicology* 193 (December): 97–104.

Smit, A. F. A., R. Hubley, and P. Green. 2013-2015. *RepeatMasker Open-4.0*. http://www.repeatmasker.org.

Sork, Victoria L., Shawn J. Cokus, Sorel T. Fitz-Gibbon, Aleksey V. Zimin, Daniela Puiu, Jesse A. Garcia, Paul F. Gugger, et al. 2022. “High-Quality Genome and Methylomes Illustrate Features Underlying Evolutionary Success of Oaks.” *Nature Communications* 13 (1): 2047.

UniProt Consortium. 2019. “UniProt: A Worldwide Hub of Protein Knowledge.” *Nucleic Acids Research* 47 (D1): D506–15.

Vaillancourt, Brieanne, and C. Robin Buell. 2019. “High Molecular Weight DNA Isolation Method from Diverse Plant Species for Use with Oxford Nanopore Sequencing.” *bioRxiv*. https://doi.org/10.1101/783159.

Walker, Bruce J., Thomas Abeel, Terrance Shea, Margaret Priest, Amr Abouelliel, Sharadha Sakthikumar, Christina A. Cuomo, et al. 2014. “Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement.” *PloS One* 9 (11): e112963.

Yang, Jing, Hafiz Muhammad Wariss, Lidan Tao, Rengang Zhang, Quanzheng Yun, Peter Hollingsworth, Zhiling Dao, et al. 2019. “De Novo Genome Assembly of the Endangered Acer Yangbiense, a Plant Species with Extremely Small Populations Endemic to Yunnan Province, China.” *GigaScience* 8 (7). https://doi.org/10.1093/gigascience/giz085.

Yang, Zhaoen, Xiaoyang Ge, Weinan Li, Yuying Jin, Lisen Liu, Wei Hu, Fuyan Liu, Yanli Chen, Shaoliang Peng, and Fuguang Li. 2021. “Cotton D Genome Assemblies Built with Long-Read Data Unveil Mechanisms of Centromere Evolution and Stress Tolerance Divergence.” *BMC Biology* 19 (1): 115.

Yuen, Zaka Wing-Sze, Akanksha Srivastava, Runa Daniel, Dennis McNevin, Cameron Jack, and Eduardo Eyras. 2021. “Systematic Benchmarking of Tools for CpG Methylation Detection from Nanopore Sequencing.” *Nature Communications* 12 (1): 3438.

Yu, Tao, Yiheng Hu, Yuyang Zhang, Ran Zhao, Xueqing Yan, Buddhi Dayananda, Jinpeng Wang, Yuannian Jiao, Junqing Li, and Xin Yi. 2021. “Whole-Genome Sequencing of Acer Catalpifolium Reveals Evolutionary History of Endangered Species.” *Genome Biology and Evolution* 13 (12). https://doi.org/10.1093/gbe/evab271.

Zagorski, Danijela, Matthias Hartmann, Yann J. K. Bertrand, Ladislava Paštová, Renata Slavíková, Jiřina Josefiová, and Judith Fehrer. 2020. “Characterization and Dynamics of Repeatomes in Closely Related Species of Hieracium (Asteraceae) and Their Synthetic and Apomictic Hybrids.” *Frontiers in Plant Science* 11 (November): 591053.

Zimin, Aleksey V., Daniela Puiu, Ming-Cheng Luo, Tingting Zhu, Sergey Koren, Guillaume Marçais, James A. Yorke, Jan Dvořák, and Steven L. Salzberg. 2017. “Hybrid Assembly of the Large and Highly Repetitive Genome of Aegilops Tauschii, a Progenitor of Bread

Wheat, with the MaSuRCA Mega-Reads Algorithm." *Genome Research* 27 (5): 787–92.

SUPPLEMENTARY MATERIAL

**Supplementary Files**

File S1: Assembly statistics, annotation statistics, summaries of repeat coverage by class and family, tombo resquiggle summaries
https://gitlab.com/PlantGenomicsLab/acermethylation/-/blob/main/manuscript/supplemental/FileS1-summaries.xlsx

**Supplementary Figures**

Figure S1: Distribution of CHH methylation by subcontext across all chromosomes in new and original *A. negundo* (acne) and *A. saccharum* (acsa) genomes.

Figure S2: Distribution of CHG methylation by subcontext across all chromosomes in new and original *A. negundo* (acne) and *A. saccharum* (acsa) genomes.

Figure S3: Gene densities (red) with CG (blue), CHG (green), and CHH (maroon) contexts in 100 Kb windows across new and original *A. negundo* (acne) and *A. saccharum* (acsa) genomes.

Figure S4: a) Methylation frequency distribution across protein coding regions, shown by assembly and sequence context for original *A. negundo* (acne) and *A. saccharum* (acsa) genomes. TSS = transcription start site; TES = transcription end site, b) methylation frequency distribution across 240 (of the original 245) genes differentially expressed in response to calcium and aluminum treatments in stem as seen in McEvoy et al. (2021), c) methylation frequency distribution for *Q. lobata* across protein coding regions (PCG) and d) TEs by family, reprinted from Sork et al. (2022).

Figure S5: Per-site CHH methylation for gene regions. Y-axis indicates the portion of the region methylated at the rates indicated by circle size. ACNE = *A. negundo*; ACSA = *A. saccharum*.

Figure S6: Distribution of gene density, repeat families, methylation by context, CHH methylation subcontexts, gene family dynamics, and gene expression results for select chromosomes. New and original *A. negundo* (acne) and *A. saccharum* (acsa) are shown. Middle row, right, shows multiple gene family expansions (lime green vertical bars) in *A. saccharum*. Bottom row shows gene family expansions and contraction (purple bars) in *A. negundo*. Gene expression results are from aluminum and calcium treatments at Hubbard Brook Experimental Forest as detailed in McEvoy et al. (2021). X-axis indicates log2 fold change, while dot size represents the p-adjusted value.