

Wavelet-based Genomic Signal Processing Locates Centromeres and helps reveal an unidentified CENH3 gene in *P. trichocarpa*.

Background

Accurate identification of centromere positions, which are important factors in chromosome segregation, is challenging. To date their positions are approximately identified from analysis of single data sets such as gene density or recombination rates. However, there now exist multi-omic data sets from a large number of *Populus trichocarpa* genotypes that can provide multiple lines of evidence to detect and characterize centromeres with greater resolution.

Approach

- We performed wavelet-based signal processing across the genome of methylation, gene, single nucleotide polymorphism (SNP) and assembly gap density signals

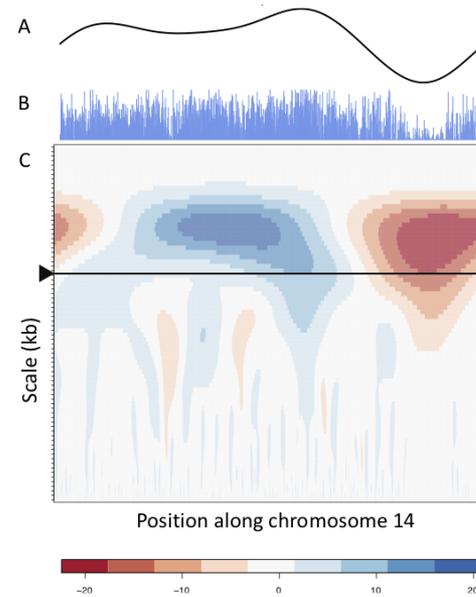
Outcomes

- Detected centromeric regions in *P. trichocarpa* and identified a specific wavelet characteristic signature of centromere position from SNP density profiles.
- Co-evolution analysis identified concentrations of SNPs in the centromeric regions correlated with SNPs in a putative *P. trichocarpa* CENH3 gene, a centromere-specific histone. This may alter the annotation of the primary function of the CENH3 gene.

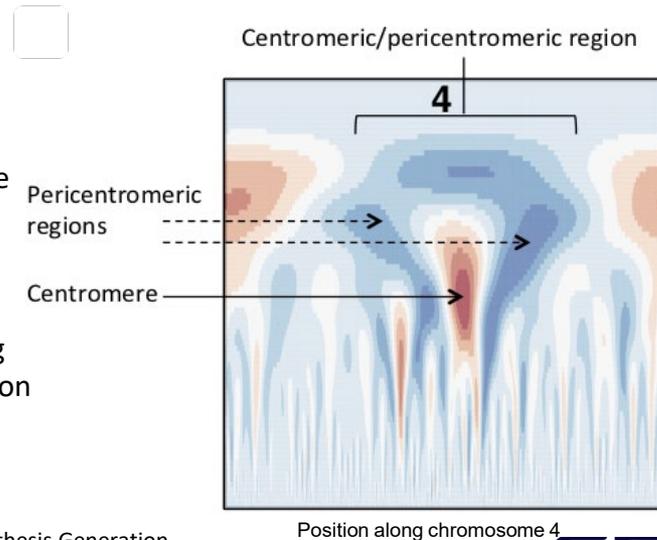
Significance

- Centromere location is important in investigations involving recombination, such as genomic selection. CENH3 correlation analysis supports the hypothesis that the sequence of the centromere is co-evolving with the centromere-interacting CENH3.

Wavelet-based Genomic Signal Processing for Centromere Identification and Hypothesis Generation, Weighill D, Macaya-Sanz D, DiFazio S, Joubert W, Shah M, Schmutz J, Sreedasyam A, Tuskan G and Jacobson D (2019) *Front. Genet.* 10:417. doi: 10.3389/fgene.2019.00487



Continuous wavelet transform (CWT) and Smooth Peaks. CWT landscape of the gene density profile of chromosome 14. (B) is the original gene density signal, (C) is the CWT coefficient landscape of the signal and (A) shows the vector of wavelet coefficients of the scale corresponding to the large scale valley, as shown by the arrow in C.



CWT of SNP density profile showing Wavelet Centromere Signature in chromosome 4. “Tooth-x-ray” centromeric signature for SNP density consisting of a broad scale peak encompassing the centromeric/pericentromeric regions, and the lower scale valley within the large peak indicating the centromeric region.