RES-FOR HIGHLIGHT #13

February 2021

High-throughput SNP Genotyping-by-Sequencing (GBS) and Genomic Selection (GS)

Overview- GS shifts the paradigm

Genomic selection (GS) is a promising pathway to accelerate genetic gain per unit time by shortening the lengthy breeding and testing cycles typical in forestry. In addition to reducing the institutional investment in carefully maintaining expensive large-scale field trials each generation, GS further decreases the error inherent in selection decisions based solely on phenotypes with reliable genetic merit.

The foundation for this opportunity is grounded on two major developments: 1) the accumulated knowledge of functional genetic diversity, genes and QTLs (quantitative trait loci); and 2) the technological advancement in high-throughput, robotic, next-generation sequencing technologies. With the modern-day smart computing techniques, marker-data deployment can now effectively accommodate data from tens or hundreds of thousands of progenies evaluated in field trials and deliver the genomic estimated breeding values (GEBVs) in time for breeding decisions to be made. Breeding programs world-wide are going through a revolutionary GS transition.

The overwhelming success in animal breeding (eg: dairy industry) has encouraged a broader adoption of GS in cereal crops, and it is now raising hopes to improve efficiency and productivity for tree improvement programs. However, the genotyping technologies required for GS remain challenging for conifer tree improvement programs: compared with livestock and cereal crops, conifer genomes are massive in size and composed of extensive low-complexity, highly-repetitive, non-coding regions; and very little is known about their genomic architecture, such as positions of QTLs and genes.

Table 1. Examples of available high-throughput single nucleotide polymorphism (SNP) genotyping technologies.

Genotyping Platform	Technology	SNP x simple combinations	Capital Investment	Cost per sample	Advantages
Illumina Infinium BeadArray	Fixed array	3- 50K SNPs x 48 (maizeSNP50) 9-80K x 48 (Tomato and Wheat)	High	Moderate	High quality, pre-curated SNPs, high multiplexity
Affymetrix Axiom GeneChip	Fixed array	570K SNPs x 96 samples (corn); 50K SNPs x 384 samples (rice)	High	Moderate	High quality, pre-curated SNPs, high multiplexity
KASP™ Fluidigm Dynamic Arrays	Flexible PCR-based	96 SNPs x 96 samples	Moderate to Low	Low	High-throughput
GBS, Genotyping-by-Sequencing	Reduced representation	~100K x 48 samples (pine) 6 million SNPs x 48 samples (spruce)	Low to Moderate	Low to Moderate	Marker Discovery
Amplicon sequencing	Targeted-GBS	Variable (20-500 SNPs x 48-384 samples)	Low to Moderate	Low to moderate	Targeted genotyping

[1]ncbi.nlm.nih.gov/probe/docs/techbeadarray; [2]affymetrix.com/agrigenotyping; [3]KASP assay design; [4]Elshire et al., 2011; [5]Sato et al., 2019

Genotyping-by-sequencing (GBS) for resilient forest

GBS, genotyping-by-sequencing, is a robust, flexible, and affordable SNP discovery platform. It is robust because of its high reproducibility. Once the SNP discovery (identification) phase is completed, species-

wide genotyping can be performed for newly collected DNA samples (typically from needles). Moreover, the determination of genetic parameters for the newly genotyped individuals would not be biased by the limited genetic information in the original SNP discovery- the flexibility of GBS allows novel genetic diversity to be rediscovered. Furthermore, because GBS does not require any prior genomic knowledge, this technology can be applied across a wide range of species. This genotyping approach utilizes restriction enzymes to avoid low complexity, highly repetitive regions of the genome. Doing so provides GBS a straightforward means to target high-complexity genic (coding) regions, increase the throughput rate and reduce sequencing costs.

The early success with high-throughput SNP genotyping was achieved with SNP arrays (Illumina BeadArray and Affymetrix GeneChip, Table 1). The SNP identification and selection processes are especially critical in designing fixed arrays such as a SNP chip, which makes the technology expensive upfront, and uncompromising when the family- or population-specific diversity needs to be captured. The trade-off of employing technologies like GBS is the considerable bioinformatics support needed to properly analyze and curate the sequence data. Nonetheless, in many cases, genetics and bioinformatics knowledge established from developing GBS could lead to the further production of higher throughput genotyping systems such as KASP™ assays using a Fluidigm system (Table 1).

SNP density and the estimation of genetic parameters

Equipped with the genotyping capacity, the RES-FOR team has investigated GBS technology in estimating genetic parameters for the large lodgepole pine and white spruce genomes. As shown in Figure 1A (pine) and 1C (spruce), with ~ 25K SNPs, for all traits measured in the RES-FOR trees including growth, wood quality attributes, pest resistance, drought tolerance and chemical phenotypes, heritability estimates were subpar in all progeny test sites. However, increasing the SNP number up to ~ 500K improved the heritability estimates which outperformed the conventional pedigree estimations (Figure 1B). These results highlight the impact of adequate genome coverage, provided by the GBS technology, on the proportion of genetic variance captured in these two coniferous species.

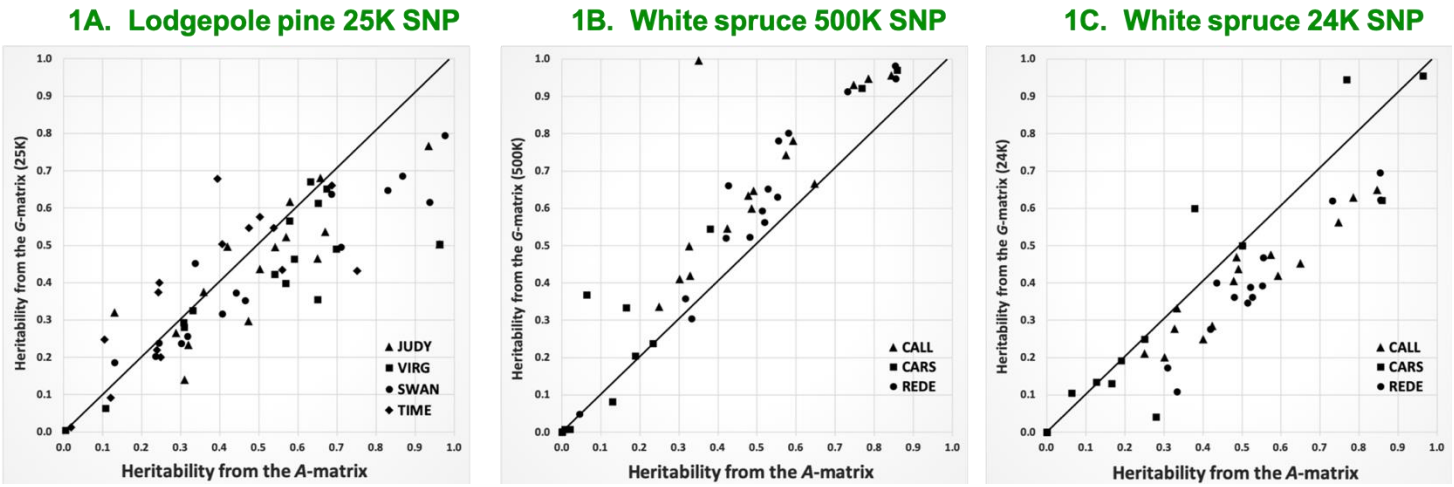

Figure 1. Estimation of heritability of all RES-FOR lodgepole pine and white spruce trees using pedigree information and different GBS genotyping capacities (ie: number of SNPs). The solid lines represent the 1:1 relationship between these two heredity estimates, pedigree versus genomics (see Highlight sheet #12).

For more information on the RES-FOR project please contact:
Barb Thomas: bthomas@ualberta.ca or Stacy Bergheim: sberghei@ualberta.ca

For more information on this highlight sheet please contact:
Charles Chen: charles.chen@okstate.edu or Eduardo Pablo Cappa: cappa.eduardo@inta.gob.ar

RES FOR