Genomics to Increase Breeding Efficiency Example: Cedar Polycross

Omnia Gamal EL-Dien¹

NSERC Postdoctoral fellow Research Officer, CFS

Tal J. Shalev¹, Lise van der Merwe², Carol Ritland¹, Matias Kirst³, Alvin D. Yanchuk², John H. Russell², Joerg Bohlmann¹

¹Michael Smith Laboratory (UBC), ²Minstry of Forests, ³U of Florida



ITAC Extension Meeting Virtual, January 10, 2023

CED A R

futurecedarforests.ca





Quantitative Genetics/Genomics Journey



Traditional Tree Breeding Phenotypic-based selection



- Produces improved individuals, and
- increases genetic gain over generations



- Time & cost consuming, mainly due to the testing phase.
- Depending on the trait, It can takes up to 25 years

1- Clonal Bank Management

2- Increasing BV Accuracy [Incomplete mating design] Increase Breeding Efficiency

Accuracy ↑ Time↓

Gain↑

3- Predicting BV

[Eliminate testing phase]

Cedar Enhanced Durability and Resistance

g

Objective Deliver elite selections to industry seed orchards decades ahead of traditional breeding methods using GS

Dr. John H. Russell World's leading expert on the western red and yellow cedars Years of Service 1985-2018









futurecedarforests.ca

CED 🗛 R

Western Redcedar (Thuja plicata)

- Cupressaceae
- Culturally and economically significant species in BC
- Range: From Northern California to Southern Alaska
- Known for natural wood durability (outdoor products)



WRC Improvement Program in British Columbia

Wil	ldstand Selection 1000 Parents	63 - 53 - 41 - 10002 22 - 22 - 10 -	Growth	He (j	cartwood durability fungal resistance).	De	er browsing resistance	
	Phenotype		Height & Diame	ter	Heartwood extractiv	es	Foliar extractive	S
	Selection age		7-year-old		25-year-old		1-year-old	
(Generation		1 st gen				2 nd gen	

Phenotype \rightarrow Breeding Value (BV) \rightarrow Selection \rightarrow Reforestation

CED **A** R

1st generation PX progeny trial

- **1000** female parent trees
- Polycrossed with a common set of pollen parents from (21 males)
- Seven testing series with four to seven field test sites per series.
- For this study, we sampled trees from series #3 (planted in 2000), which tested **111 PX families** for growth over six sites.

Training population

- 1,520 trees (~ 18 years old)
- 26 PX families (~58 HS offspring)
- 3 sites

CFDAR

futurecedarforests.ca

- Phenotyped (3 traits)
- Genotypes: Parents (21 + 26) + Offspring





1-Clonal Bank Management

Coefficient of Relatedness

Measure of genetic similarities/differences among individuals due to ADDITIVE genetic effects of shared alleles



(Pedigree $\rightarrow A$ matrix)

Average amount of shared alleles

- Parent-Offspring = 0.50
- Full-Sib (FS) = 0.50
- Half-Sib (HS) = 0.25

Average relatedness

Expected Genomic Relationship





Unexpected Error

TWO possible genotypes for 8 female parents

26 PX families (~58 HS offspring)





Take home message: Are we propagating the right winners?

Confirm that all the clones of the WINNERS are correct ③



2- Increasing BV Accuracy [Incomplete mating design]

Training population

- 1,520 trees (~ 18 years old)
- 26 PX families (~58 HS offspring)
- 3 sites
- Phenotyped (3 traits)
- Genotypes: Parents (21 + 26) + Offspring



Background

1st generation progeny trial

- Open-Pollinated (OP)
- Poly-Cross (PX)



INCOMPLETE pedigree Mating Design

- *Forward* selection (selection from *offspring*): Inaccurate offspring BVs estimate
- ✓ Backward selection (selection from parents): Female only



1. Pedigree Reconstruction

(Training population, N=1520, PX \rightarrow FS)



- Range: 7-187 offspring/male
- Mean: 68

- Range: 1-15 offspring/FS
- Mean: 3.3

CEDAR R (Polycross): Each female parent is pollinated with a pollen mix collected from a group of known males FS (Full-sib): Each female parent is pollinated with one known male

Traits

(Applied perspective: 3/ Academic perspective 8)

• Growth:

0	Height	[HT]
0	DBH	[DBH]

• Foliar specialized chemistry (34 compounds):

0	Total monoterpenes	[F.TM]
0	α-thujone	[F.AT]

- Wood specialized chemistry (19 compounds):
 - Total **thujaplicins**
 - o Total **lignans**
 - Total extractives
 - \circ α -thujaplicin

[W.TT] [W.TL] [W.TE] [W.AT]

CED & R

2. Genetic estimates (Heritability, BV, GxE & genetic gains)

PX-pedigree (ABLUP) vs. Genomic (GBLUP)

(Training population)

	НТ	DBH	F.TM	F.AT
Heritability	0.13 (0.04)	0.09 (0.04)	0.28 (0.05)	0.25 (0.05)
Genetic Gain%	6.01	7.44	22	25.5
	W.TT	W.TL	W.TE	W.AT
	0.18 (0.05)	0.14 (0.04)	0.20 (0.05)	0.19 (0.05)
	6.15	55	11	7.7

ABLUP resulted in

- Up to 51% overestimation in genetic gain, and
- Up to 44% in genetic diversity

<u>Gains (%):</u> Gains are expressed as the percentage of the selected 75 (5% of 1520) trees' mean BV relative to the population phenotypic mean.

21



PX-pedigree (ABLUP) vs. Genomic (GBLUP) Up to 22% BV accuracy

GAMAL EL-DIEN ET AL. Evolutionary Applications									
TABLE 3 Estimates of theoretical accuracy (\hat{r}) of parents' and offspring's breeding values for selected models and all tested traits									
Model ABLUP-PX GBLUP-A									
Parents				Parents					
Trait ^a	Females	- Offspring		Males	Females	Offspring			
HT	0.74	0.47		0.63	0.63	0.56			
DBH	0.67	0.41		0.56	0.56	0.50			
F.AT	0.84	0.58		0.71	0.72	0.67			
F.TM	0.82	0.55		0.70	0.70	0.65			
W.AT	0.79	0.53		0.67	0.68	0.62			
W.TT	0.78	0.52		0.66	0.67	0.61			
W.TL	0.8	0.51		0.67	0.68	0.60			
W.TE	0.79	0.53		0.67	0.68	0.62			

Note: The number of female parents = 25 (26 - 1 non-genotyped parent), number of male parents = 20 (21 - 1 non-genotyped parent), number of offspring = 1506 (representing 26 PX families and 438 FS families after pedigree reconstruction), and the average number of offspring per family \approx 58/maternal HS family, 68/paternal HS family, and 3.3/FS family.

^aSee Table 1 for traits description.

 ★ Forward selection (selection from offspring): Inaccurate offspring BVs estimate
 ✓ Backward selection (selection from parents): Female only

3. Trade-offs?

Growth \leftrightarrow **Foliar extractives** \leftrightarrow **Wood extractives**

Trait ^a	нт	DBH	F.AT	F.TM	W.AT	W.TT	W.TL	W.TE
HT		0.81	0.05	0.04	-0.04	-0.09	-0.05	-0.09
DBH	0.78		0.10	0.10	-0.08	-0.14	-0.13	-0.16
F.AT	0.03	0.23		0.97	-0.02	-0.04	-0.05	-0.05
F.TM	0.02	0.21	0.97		0.00	-0.02	-0.06	-0.03
W.AT	0.16	0.03	0.14	0.16		0.85	0.41	0.81
W.TT	0.09	0.00	0.13	0.14	0.82		0.66	0.97
W.TL	0.47	0.43	-0.02	-0.01	0.47	0.59		0.74
W.TE	0.12	0.03	0.11	0.12	0.83	1.00	0.67	

TABLE 4 Phenotypic (above diagonal) and genetic correlation (below diagonal) between tested traits

Note: Significance of both correlations was assessed differently. For phenotypic correlation, we used cor.test function in *R* for the correlation between the adjusted phenotypes to estimate *p*-value. We used an α -level of 0.05 to determine significance. Bold type reflects strong significant correlation, while italics reflect small significant correlations. For genetic correlation estimated from multivariate GBLUP-A models using CORGH structure, we identified significance as having a correlation estimate at least double the SE. Bold type reflects significant correlation. Correlation cut-offs: small, <0.4; medium, 0.4–0.7; and strong, >0.7.

^aSee Table 1 for traits description.

Growth for the indirect selection of Lignans? Thujaplicins for the indirect selection of Lignans?

PX-pedigree (ABLUP) vs. Genomic (GBLUP) (Training population, N=1520, 3 measured phenotypes)

Breeding Values

– PX limitation

CEDAR

- ✓ Backward selection: Female only
- **×** Forward selection: Inaccurate offspring BVs estimate
- Backward selection (from 45 parents)
 - ↑ Selection intensity (Female and male BVs)
- Forward selection (from 1520 offspring)
 - 22%
 BV accuracy
 - 35% ↑ Expected genetic gain

PX (Polycross): Each female parent is pollinated with a pollen mix collected from a **group** of known males

Take home message:

PX mating design will benefit the most from GBLUP

Each tree \rightarrow ~ 58 HS from the female parent

~ 72 HS from the male parent

~ 3 FS



3- Predicting BV

Eliminate testing phase \rightarrow Early selection at the SEEDLING phase

Genomic Selection

From Proof-of-Concept to Application

Target



Genomic Selection (GS) Predicting phenotypes from genotypes



Prediction Accuracy (PACC): Correlation between measured BV and predicted GBV for validation populationGenomic Breeding ValueEl-Kassaby lecture notes

GS approaches (Y) → GBV (Genomic Breeding Value)

Models	GBLUP Genomic Best Linear Unbiased Prediction	Regression Models Ridge Regression, Bayes-A,-B,-C, Bayesian Lasso, ,and Kernel regression
SNPs <i>(Xs)</i>	Genomic relationship Prediction + Increasing Accuracy	SNPs effects Prediction

Genomic Selection Efforts in Canada (A) Proof-of-Concept (Eastern Canada)



Genomic Selection Efforts in Canada (A) **Proof-of-Concept (Western Canada)**



B Ratcliffe¹, OG El-Dien¹, J Klápště^{1,2}, I Porth¹, C Chen³, B Jaquish⁴ and YA El-Kassaby¹

Genomic Selection Efforts in Canada (A) **Proof-of-Concept (Western Canada)**



black spruce, Norway spruce, Douglas fir, and lodgepole pine.

(A) Proof-of-Concept Conclusions



- GS is very promising for conifers breeding and has the potential to reduce the TESTING phase into one-year→ increase genetic gain/unit time.
- Prediction accuracy is driven by **RELATEDNESS** between Training and Validating populations.
- **Growth** and **wood density** traits are complex quantitative traits.

Genomic Selection Efforts in Canada (B) **Applications (i. Quebec)**



FastTRAC (Fast Tests for Rating and Amelioration of Conifers) is a three-year project funded by Genome Canada, Génome Québec and multiple organizations through the Genomics Applied Partnership Program (GAPP). It brings together scientists, tree breeders, foresters, and economists to demonstrate and apply genomic-assisted selection at the operational scale of tree improvement programs as a method for more rapidly rating candidate trees for productivity and resilience attributes, so to hasten the selection and reforestation of improved and adapted stock.

The approach, which is based on rapid screening of natural genetic variation available in tree breeding programs, will allow more flexibility to tree breeders in monitoring and changing spruce stock in the context of changing climate and economic markets, as well as reducing rotation age and risk in the face of an uncertain future. The reforestation of improved stock together with more intensive forestry on a limited part of the forest landscape shall also contribute towards reducing harvesting pressures on pristine natural forests in the future.



g







Canada

des fonits (Canadian Forest Se





GenomeQuébec les salutions sont dans les gènes



PROJECT LEADERS

Jean Bousquet Guy Smith

(2015-2018)

ded with the...

FastTRAC rewar-

PUBLICATIONS AND MEDIAS

PROJEI

PROJECT

TEAM

LINKS

CONTACT

FRANCAIS

NEWS

Four new publications on genomic...

TV report ' Un coup de pouce de...

Azaiez et al. 2018: new publication...

http://fasttracproject.ca

Genomic Selection Efforts in Canada (B) Applications (ii. Quebec, Alberta, and British Columbia)



Home	Overview	Activitie	es Team	Resources	Publications
Media	Events	Links	Contacts		

PROJECT LEADERS

Joerg Bohlmann Jean Bousquet

```
(2016-2022)
```

About

The goal of the **Spruce-Up** Project is to produce new and enhanced genomics and socioeconomic knowledge for the development, validation and implementation of biomarkers and genomic selection systems to accelerate spruce breeding programs.

THE UNIVERSITY OF BRITISH COLUMBIA Alberta Innovates Technology Futures SCIENCES Canada GenomeBritishColumbi INNOVATION CENTRE UNIVERSITÉ Alberta and Fores g Canada **К**ТМІС ALBERTA Geological Si IBIS Ouébec GenomeAlbert TORONTO JGI BC Cancer Agency © 2017 Spruce-Up

https://spruce-up.ca/en/

Genomic Selection Efforts in Canada (B) **Applications (iii. Alberta)**



(B) Applications (iv. British Columbia) Cedar Enhanced Durability and Resistance



ABOUT THE PROJECT

PROJECT LEADERS

JOHN RUSSELL

World's leading expert on western red and yellow cedars Years of Service :1985-2018

> Joerg Bohlmann Alvin Yanchuk

(2014-2015) (2016-2021)

Canada's iconic western redcedar (WRC) export industry will be facing a future decline, as the durable and aesthetically appealing timber supply moves from old growth to second growth and faces climate-driven challenges. The CEDaR (Cedar Enhanced Durability and Resistance) project aims to develop genomic selection (GS) statistical models associating genomic and phenotypic variation for multiple traits and to apply GS in the WRC breeding program at a very early age The goal is to deliver

elite selections to industry orchards to produce seed for reforestation.



Genome BritishColumbia



futurecedarforests.ca





WRC unique biology

• **Selfing** + glacial refugia (small N_e)

CEDAR

 Seedlings (< one year) can be reproductively induced

Five selfing generations in 10 years! Hypothesis: WRC has a stronger LD compared to other conifers

LD: non-random association between molecular markers and causal genes

Tal Shalev, Omnia Gamal El-Dien et al. *Western redcedar genome reveals excessive low genetic diversity in a self-compatible conifer*. *Genome Research*

WRC Improvement Program in British Columbia

Wil	ldstand Selection 1000 Parents	63 - 53 - 41 - 10002 22 - 22 - 10 -	Growth	He (j	cartwood durability fungal resistance).	De	er browsing resistance	
	Phenotype		Height & Diame	ter	Heartwood extractiv	es	Foliar extractive	S
	Selection age		7-year-old		25-year-old		1-year-old	
(Generation		1 st gen				2 nd gen	

Phenotype \rightarrow Breeding Value (BV) \rightarrow Selection \rightarrow Reforestation

CED **A** R

Traits

(Applied perspective: 3/ Academic perspective 8)

• Growth:

0	Height	[HT]
0	DBH	[DBH]

• Foliar specialized chemistry (34 compounds):

0	Total monoterpenes	[F.TM]
0	α-thujone	[F.AT]

- Wood specialized chemistry (19 compounds):
 - Total **thujaplicins**
 - o Total **lignans**
 - Total extractives
 - \circ α -thujaplicin

[W.TT] [W.TL] **[W.TE]** [W.AT]

CED **A** R

CEDAR Cedar Enhanced Durability and Resistance

• Fitting models



• Validating models

• Predicting phenotypes



- \rightarrow Training population (PX progeny trial, 21 males)
 - 45 Parents trees
 - 1,520 trees (26 PX)
 - 3 sites
 - Phenotyped (3 traits) + genotypes
- → A. Training population (cross-validation) All traits
 - B. Target population (independent-validation) Foliar extractives
- → Target population (1-year-old)
 - 3,000 seedlings (168 FS families)
 - Phenotyped (foliar extractives) + genotypes

Genotypes: **45K** filtered SNPs (genic and non-genic) PX: Polycross FS: Full-sib

Prediction Accuracy

Correlation between (predicted and true BV) for validation population

Correlation between **predicted BV** (GBLUP, CV) and?

BV (GBLUP) \rightarrow Prediction Accuracy (PACC)



GS validation in TRAINING population (N=1520, All traits) A. Random cross-validation (pooled sites)



B. Removing relatedness PACC= Relatedness + LD



Others CV

- C. Across-Sites
 - Small or no decrease in PACC

- **D. Across-Ages**
 - Small decrease in PACC
- **E. Across-Generations**
 - Small decrease in PACC

2. GS validation in TARGET population independent-validation

(Foliar extractives, α-thujone, GBLUP)



CEDAR

Prediction Accuracy: Correlation between **measured BV** (GBLUP, full data) and **predicted BV** (GBLUP, CV) for validation population

3. Predicting growth & wood extractives for TARGET A. GBLUP (genomic) vs. ABLUP (pedigree) prediction



4. Genomic Selection Application Early selection from the target population

- 1) Predicting BV for growth and wood resistance traits
- 2) Multi-trait selection(Selection index: GROWTH, WOOD, foliar)
- 3) Optimal selection for the best 100 seedlings:
 - Seed orchard
 - Field testing



futurecedarforests.ca

CEDAR

Optimal selection: balancing between genetic gain and diversity

CED**A**R **Conclusions**

Training (measured phenotypes)



Overcome limitation of PX

Pedigree reconstruction/GBLUP

- Forward selection (1520 offspring)

- 22% ↑ BV accuracy
- 35%
 Genetic gain

Omnia Gamal El-Dien et al. *Genomic selection reveals hidden relatedness and increased breeding efficiency in western redcedar polycross breeding* (Published in Evolutionary Applications)

Target (predicted phenotypes)



Predicting BV (wood + growth)

Multi-traits selection

Early selection + Accelerated breeding Breeding cycle $25 \rightarrow 2$ years

↑ Genetic gain

Omnia Gamal El-Dien et al. *Genomic Selection: From proof-of-concept to application in western redcedar (Thuja plicata) breeding program (in preparation)*

Future GS projects CEDaR-B2

Cedar Enhanced Durability and Resistance to Blight and Browsing

• Forest health issues

- Cedar Leaf Blight.
 - Significant loss of incremental growth
- Wildlife Browse

Joerg Bohlmann Alvin Yanchuk

PROJECT LEADERS

(2022-2024)

• The industry spends ~ \$6 million to protest seedlings

• **Phenotyping TRAINING population:**

- Cedar leaf blight (in-vitro)
- Browsing resistance (field experiment)



FastTRACII

Fast Tests for Rating and Amelioration of Conifers

- Target species: **black spruce** and **red spruce**
- 95 million planted seedlings per year in Quebec and Maritime provinces
- Goals:
 - Multi-trait genomic selection for <u>growth</u> productivity, <u>wood quality</u>, and **RESILIENCE** to <u>drought</u> and <u>cold stress</u> in relation to climate instability



PROJECT LEADERS Lead Jean Bousquet (ULaval) Co-lead Patrick Lenz (NRCan) Bruce Stewart (Nova-Scotia DNR) (2022-2026)







Overview of my current research project





Canada's National Forest Inventory Inventaire Forestier National du Canada

Multi-Agency Ground Plot (MAGPlot) database: Agglomerating forest sampling data for pan-Canadian research projects



Ressources naturelles Canada

s Natural Resources Canada





MAGPlot Database



Objectives

- Collate Canadian forest ground-plot data from multiple sources into a single, centralized, and a Canada-wide standard format
- Serve data in standardized, analysis-ready format.
- Data Sharing Agreement
 - → Receiving Data

[Plot & Tree measurements, Ecological, Site Treatments & Disturbance, Growth, Age, etc.]

→ Quality Control

 \rightarrow Standardization





NFIS

Current status



Canada

Government

of

Saskatchewan



Can we have the same for Genetic Trials?

PhD Team (El-Kassaby Lab, 2012-2017)









GS Team Quebec



- J.-P. Laverdière, graduate student, Laval University (now CWFC)
- Claire Depardieu, postdoc, Laval University
- Simon Nadeau, GS analyst, Canadian Wood Fibre Centre
- Martin Perron, scientist, Quebec Ministry of Forests
- Nathalie Isabel, scientist, Canadian Forest Service
- Jean Beaulieu, invited professor, Laval University
- Patrick Lenz, scientist, Canadian Wood Fibre Centre
- Jean Bousquet, professor and CRC, Laval University

















University of British Columbia

- Dr. Joerg Bohlmann
- Dr. Carol Ritland
- Tal Shalev
- Mack Yuen
- Dr. Inanc Birol
- Bohlmann Lab





University of Florida

• Dr. Matias Kirst

Joint Genome Institute

• Jeremy Schmutz

BC Ministry of Forests, Lands and Natural Resource Operations

- Dr. John Russell
- Dr. Alvin Yanchuk
- Lise van der Merwe

RAPiD Genomics

- Dr. Leandro Neves
- Dr. Jesse Breinholt

FPInnovations

• Dr. Rod Stirling









CED A R

futurecedarforests.ca



Traditional breeding Late selection Measured BV

Molecular breeding Early selection Predicted BV



Forest Statistics, NFI, CFS