What is Next in the Application of Genome-Wide Information to Tree Breeding?

Matias Kirst

Professor of Quantitative Genetics and Genomics
Coordinator Plant Molecular and Cellular Biology Graduate Program
Co-Director Cooperative Forest Genetics Research Program
University of Florida Genetics Institute
School of Forest Resources and Conservation

SFTIC
Melbourne, Florida – June 20 2017
Will we fail (again)?

Matias Kirst

Professor of Quantitative Genetics and Genomics
Coordinator Plant Molecular and Cellular Biology Graduate Program
Co-Director Cooperative Forest Genetics Research Program
University of Florida Genetics Institute
School of Forest Resources and Conservation

SFTIC
Melbourne, Florida – June 20 2017
1990 Map

After USDA Plant Hardiness Zone Map, USDA Miscellaneous Publication No. 1475, Issued January 1990

2006 Map

National Arbor Day Foundation Plant Hardiness Zone Map published in 2006.

Zone

© 2006 by The National Arbor Day Foundation®
No. 34. Complete destruction of chestnut trees in a nearly pure stand. Many of the trunks have lost their bark. Scene in Forest Park, near Brooklyn, New York.—Photograph by Prof. Collins. (1912)
1990’s

QTL

Parents

1    2    3    4     5           n

Current population

2010’s

Genome-wide selection

Breeding parents

1    2    3    4     5           n

Current population

2000’s

GWAS

Ancestral population

1    2    3    4     5           n

Current population
Fit all SNPs in a prediction model:

\[ Y = SNP + e \]

Training population

Genotypes

Phenotypes

Validation

Define multi-loci models to predict phenotypes

Meuwissen et al. (2001) Genetics 157: 1819-1829
Genome-Wide Selection

Training population

Phenotypes \rightarrow Genotypes

Fit all SNPs in a prediction model

\[ Y = SNP + e \]

Define multi-loci models to predict phenotypes

Meuwissen et al. (2001) Genetics 157: 1819-1829
GWS
Genome-Wide Selection

Lower \( N_e \) \( \sim \) lower allelic range
Higher LD = lower resolution
Application to breeding

GWAS
Genome-Wide Association Studies

Higher \( N_e \) \( \sim \) higher allelic range
Lower LD = higher resolution
Application to discovery of gene function / functional genomics
Lower $N_e \sim$ lower allelic range
Higher LD = lower resolution
Application to breeding

Higher $N_e \sim$ higher allelic range
Lower LD = higher resolution
Application to discovery of gene function / functional genomics
Lower $N_e \sim$ lower allelic range
Higher LD = lower resolution
Application to breeding
Markers needed $\approx 10-20$/cM

Higher $N_e \sim$ higher allelic range
Lower LD = higher resolution
Application to discovery of gene function / functional genomics
Markers needed $\approx 1M+$
Genomic selection in forest species

Dario Grattapaglia · Marcos D. V. Resende

Prospects for genomic selection: a simulation study of

Hiroyoshi Iwata · Takeshi Hayashi · Yūki Kume

Methods in a Standard Populus (Pinus taeda L.)

V. Resende, D. J. Garrick, R. L. Fernando, T. A. Martin, G. F. Peter, and M. Kirst

Genetics, Vol. 190, 1503–1510 April 2012

Accelerating the domestication process: improving the accuracy of prediction

M. F. R. Resende Jr, P. Muñoz, M. D. V. Resende, and M. Kirst

Species within and between breeding groups in

André Rainville and Jean Bousquet

New Phytologist

BMC Genomics

Open Access
Genomics incorporated into pine breeding

Genomic Selection in Conifers

- "Comparing CLonal Lines on Experimental Sites = CCLONES"
- 900 replicated loblolly pine clones developed from 62 elite full-sib families
- $Ne \approx 40$
- Four field locations
- $\sim 3$ SNP markers/cM

Márcio Resende Jr.  Patricio Munoz

Genomic Selection Accuracies in CCLONES

DBH & Height

Phenotypic selection

Expectation from empirical calculation:
N ~1,000
QTL ~50-100
h² ~ 0.2-0.4
Other trait accuracies

<table>
<thead>
<tr>
<th>Trait</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cellulose</td>
<td>0.9</td>
</tr>
<tr>
<td>Lignin</td>
<td>0.6</td>
</tr>
<tr>
<td>Density</td>
<td>0.8</td>
</tr>
<tr>
<td>Wood stiffness</td>
<td>0.6</td>
</tr>
<tr>
<td>Rust</td>
<td>0.7</td>
</tr>
<tr>
<td>Pitch canker</td>
<td>0.5</td>
</tr>
</tbody>
</table>
## Genomic Selection Accuracies in CCLONES

<table>
<thead>
<tr>
<th>Trait category</th>
<th>Trait</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RR-BLUP</td>
</tr>
<tr>
<td>Growth</td>
<td>Height</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Diameter</td>
<td>0.46</td>
</tr>
<tr>
<td>Development</td>
<td>Crown width</td>
<td>0.48</td>
</tr>
<tr>
<td></td>
<td>Branch diameter</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Branch angle</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>Root number</td>
<td>0.24</td>
</tr>
<tr>
<td>Disease</td>
<td>Rust</td>
<td>0.29</td>
</tr>
<tr>
<td>resistance</td>
<td>Rust volume</td>
<td>0.23</td>
</tr>
<tr>
<td>Wood quality</td>
<td>Stiffness</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Lignin</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Late Wood</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>Density</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>C5C6</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Resende *et al.* Genetics 190:1503-10.
Genomic Selection Accuracies in CCLONES

<table>
<thead>
<tr>
<th>Trait</th>
<th>Methods</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RR-BLUP</td>
<td>BLASSO</td>
<td>Bayes A</td>
<td>Bayes Cπ</td>
<td>RR-BLUP B</td>
</tr>
<tr>
<td>Rust</td>
<td>0.29</td>
<td>0.28</td>
<td>0.34</td>
<td>0.34</td>
<td>0.33</td>
</tr>
<tr>
<td>Rust volume</td>
<td>0.23</td>
<td>0.24</td>
<td>0.28</td>
<td>0.29</td>
<td>0.37</td>
</tr>
</tbody>
</table>

How stable are prediction models across years?

Height

Genomic Selection Accuracies in CCLONES
Genomic Selection Accuracies in CCLONES

How stable are prediction models across sites?

<table>
<thead>
<tr>
<th>DBH</th>
<th>Nassau</th>
<th>Palatka</th>
<th>BFGrant</th>
<th>Cuttberth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassau</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatka</td>
<td>0.70</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFGrant</td>
<td></td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuttberth</td>
<td></td>
<td></td>
<td>0.57</td>
<td></td>
</tr>
</tbody>
</table>
Genomic Selection Accuracies in CCLONES

**How stable are prediction models across sites?**

<table>
<thead>
<tr>
<th>DBH</th>
<th>Nassau</th>
<th>Palatka</th>
<th>BFG Grant</th>
<th>Cuttberth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassau</td>
<td>0.71</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatka</td>
<td>0.48</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFG Grant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuttberth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How stable are prediction models across sites?

<table>
<thead>
<tr>
<th>DBH</th>
<th>Nassau</th>
<th>Palatka</th>
<th>BFGrant</th>
<th>Cuttberth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassau</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatka</td>
<td></td>
<td></td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>BFGrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuttberth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genomic Selection Accuracies in CCLONES

Map showing the distance between sites: 550 Km
### Genetic Gain Relative to Phenotypic Selection

- Increase in genetic gain using genomic selection relative to phenotypic selection.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Site</th>
<th>h (BLUP)</th>
<th>h (GS)</th>
<th>Efficiency</th>
<th>Increase relative to phenotypic selection (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBH</td>
<td>B.F. Grant</td>
<td>0.79</td>
<td>0.75</td>
<td>1.90</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Cuthbert</td>
<td>0.75</td>
<td>0.73</td>
<td>1.95</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>Nassau</td>
<td>0.85</td>
<td>0.65</td>
<td>1.53</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Palatka</td>
<td>0.81</td>
<td>0.67</td>
<td>1.65</td>
<td>65</td>
</tr>
<tr>
<td>HT</td>
<td>B.F. Grant</td>
<td>0.74</td>
<td>0.77</td>
<td>2.08</td>
<td>108</td>
</tr>
<tr>
<td></td>
<td>Cuthbert</td>
<td>0.68</td>
<td>0.74</td>
<td>2.18</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>Nassau</td>
<td>0.80</td>
<td>0.64</td>
<td>1.60</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Palatka</td>
<td>0.85</td>
<td>0.67</td>
<td>1.58</td>
<td>58</td>
</tr>
</tbody>
</table>
Will this work (next generation) in forestry?

Of course, it worked with...
Unfortunately, breeding trees is not the same as breeding chicken...
What are the threats to the application of genomics/genome-wide selection in forestry?

- **Genotype by environment interaction**
  Prediction models apply to narrow regions

- **Low multiplier**
  Many suppliers, horizontal market

- **High cost of genotyping**
  Scale and affordability

- **Low heritabilities**
  Better training populations needed

- **Slow breeding**
  Slow translation of science to application
What are the threats to the application of genomics/genome-wide selection in forestry?

- **Genotype by environment interaction**
  Prediction models apply to narrow regions
- **Low multiplier**
  Many suppliers, horizontal market
- **High cost of genotyping**
  Scale and affordability
- **Low heritabilities**
  Better training populations needed
- **Slow breeding**
  Slow translation of science to application
### Genotype by environment interaction

<table>
<thead>
<tr>
<th>DBH</th>
<th>Nassau</th>
<th>Palatka</th>
<th>BFGrant</th>
<th>Cuttberth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nassau</td>
<td>0.71</td>
<td>0.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatka</td>
<td>0.48</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BFGrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuttberth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH</td>
<td>Nassau</td>
<td>Palatka</td>
<td>BFGrant</td>
<td>Cuttberth</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>---------</td>
<td>-----------</td>
</tr>
<tr>
<td>Nassau</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palatka</td>
<td></td>
<td></td>
<td><strong>0.11</strong></td>
<td></td>
</tr>
<tr>
<td>BFGrant</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cuttberth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Genotype by environment interaction

Map showing distances between locations:
- Nassau to Palatka: 550 Km
- Palatka to BFGrant: 0.11
- BFGrant to Cuttberth: 0.11
- Cuttberth to Nassau: 550 Km
What are the threats to the application of genomics/genome-wide selection in forestry?

- **Genotype by environment interaction**
  Prediction models apply to narrow regions

- **Low multiplier**
  Many suppliers, horizontal market

- **High cost of genotyping**
  Scale and affordability

- **Low heritabilities**
  Better training populations needed

- **Slow breeding**
  Slow translation of science to application
- **Pedigree Elite Stock**
  - 35 - 40 Purelines
  - 400 thousand
  - Yr. 0

- **Grade Grand Parent**
  - 500 thousand
  - Yr. 1

- **Grand Parent Stock**
  - 12 million
  - Yr. 2

- **Parent Stock**
  - 300 million
  - Yr. 3

- **Broilers**
  - 400 billion
  - Yr. 4
What are the threats to the application of genomics/genome-wide selection in forestry?

- **Genotype by environment interaction**
  Prediction models apply to narrow regions

- **Low multiplier**
  Many suppliers, horizontal market

- **High cost of genotyping**
  Scale and affordability

- **Low heritabilities**
  Better training populations needed

- **Slow breeding**
  Slow translation of science to application
High cost of genotyping – imputation or other solutions

Fig. 2. Prediction accuracy across families and within a family against the number of markers used in the prediction set with 200 individuals; 2000 training individuals had the true high-density genotypes (HD); LD, low-density genotypes imputed to high density; letters denote significant difference within the scope of prediction at $p \leq 0.01$ according to the Tukey's multiple comparison test.
What are the threats to the application of genomics/genome-wide selection in forestry?

- **Genotype by environment interaction**
  Prediction models apply to narrow regions

- **Low multiplier**
  Many suppliers, horizontal market

- **High cost of genotyping**
  Scale and affordability

- **Low heritabilities**
  Better training populations needed

- **Slow breeding**
  Slow translation of science to application
<table>
<thead>
<tr>
<th>Species</th>
<th>Cycle</th>
<th>Current Populations</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slash</td>
<td>3</td>
<td>Main - 300 Elite – 60</td>
<td>3rd cycle test planted 2011-12</td>
</tr>
<tr>
<td>Longleaf</td>
<td>1</td>
<td>Wild selections – 1000</td>
<td>Not active</td>
</tr>
<tr>
<td>Loblolly</td>
<td>2</td>
<td>Main - 150</td>
<td>2nd cycle test planted 2012-13</td>
</tr>
<tr>
<td>Sand</td>
<td>1</td>
<td>Wild selections - 143</td>
<td>Not active</td>
</tr>
<tr>
<td>Hybrids</td>
<td>1</td>
<td>60 – F1</td>
<td>Non active</td>
</tr>
</tbody>
</table>

Funded by the joint USDA/DOE Plant Genomics for Bioenergy program (2013-67009-21200)
500 Selns
topgrafted
8 Sublines
CP Tests

Time (yrs)

Company 1

Company 2

Company 3

Company 4

Company 5

Company 6

Company 7
Summary

- Historical data in other systems suggests that GWS is applicable to forestry
- Preliminary data in conifers and other tree species support that expectation
- However, the uniqueness of tree breeding will impose limitations that may slow the application of GWS
- Which scenario will we find?
Summary

• Historical data in other systems suggests that GWS is applicable to forestry
• Preliminary data in conifers and other tree species support that expectation
• However, the uniqueness of tree breeding will impose limitations that may slow the application of GWS
• Which scenario will we find?
Acknowledgements

Chris Dervinis
Rodrigo Santos (UF - PMCB)
Janeo E. de Almeida Filho (UFV)
João F.R. Guimarães (UFV)
Gabi Nunes Silva (UFV)
Patrício Munoz (Hort. Sciences)
Esteban Rios (Agronomy)
Marcio Resende (RAPiD now UF)

Salvador Gezan
Gary Peter
Greg Powell
Coop members

This project was funded by the joint USDA/DOE Plant Genomics for Bioenergy program (2013-67009-21200)