The interplay between modern genetic evaluation and breeding strategies

Luis A. Apiolaza

School of Forestry, University of Canterbury, Christchurch, New Zealand

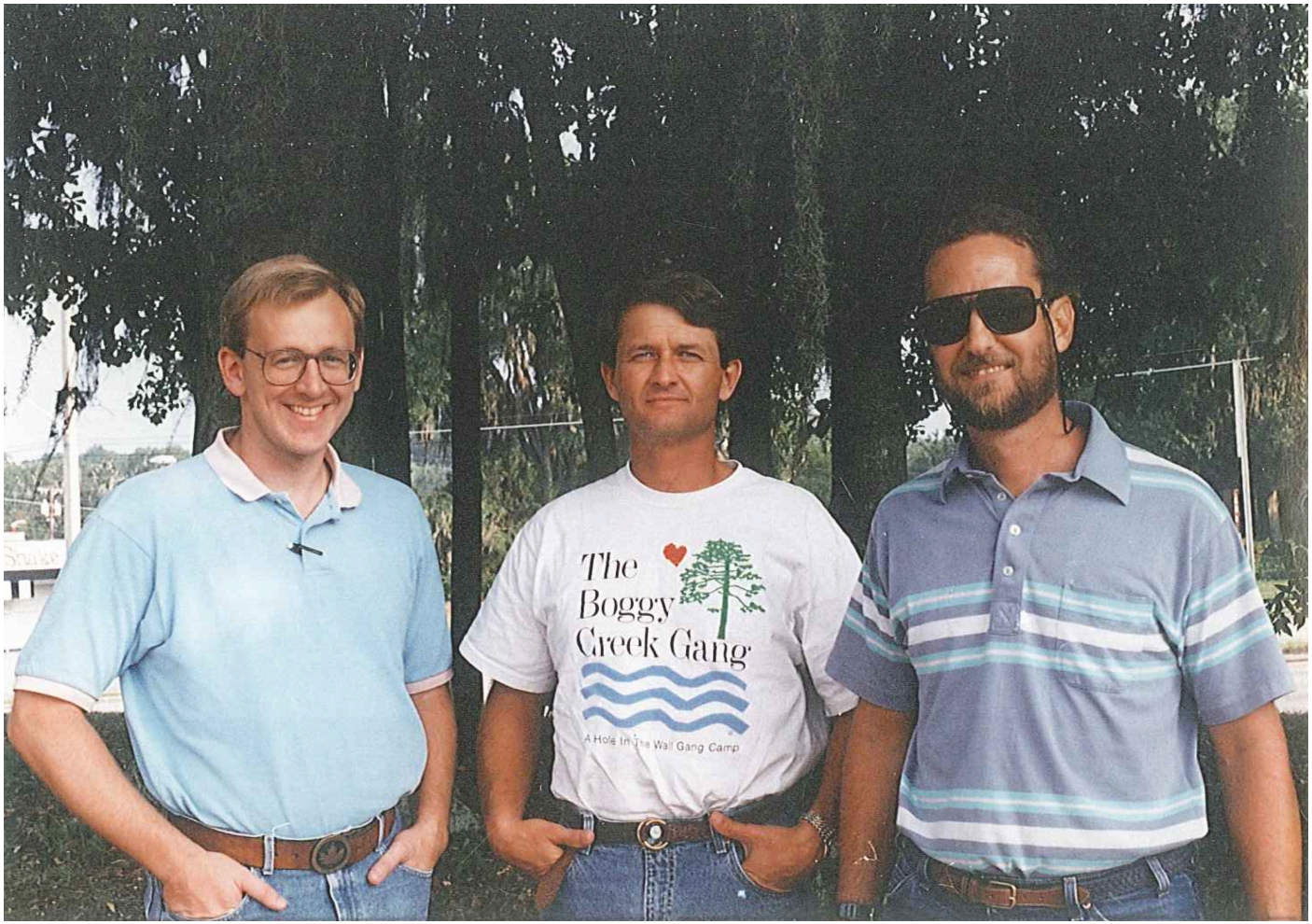
Luis.Apiolaza@canterbury.ac.nz - http://apiolaza.net

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

This presentation connects work funded by the NZ Foundation for Research Science and Technology, Radiata Pine Breeding Company (NZ), ProSeed (NZ), Forests New South Wales (AU), Weyerhaeuser (USA) and Forests and Wood Products Australia.

Many thanks to the conference organizers and sponsors for making my attendance possible.

Gainesville 1994



2001 2005 1994

Fifty years earlier…

Pinus radiata
circa 1944

Our objective as breeders

Achieving the maximum (operational) annual gain as possible

For a combination of traits that affect profit

If we do this well we can afford other objectives (e.g. conservation)

Back to Tim's presentation

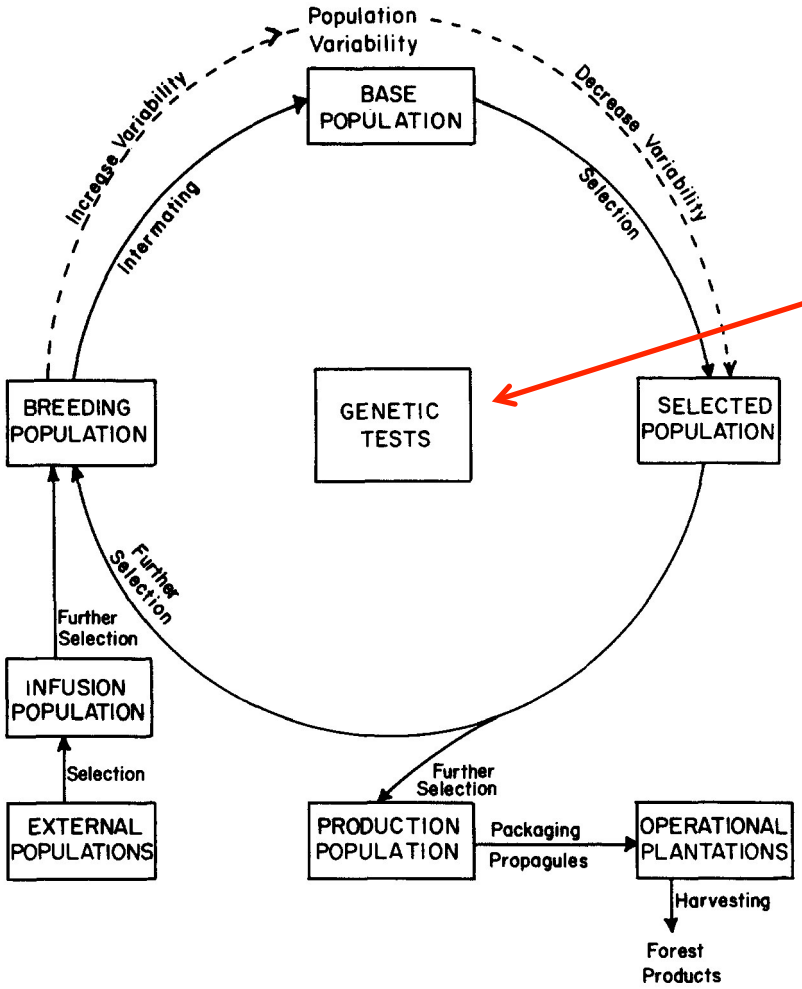


Fig. 1. Major components and activities of the breeding cycle of forest tree improvement programs. Each generation of breeding begins with the formation of a selected population. Each of three population types in the central part of the cycle (selected, breeding, base) are formed during a given generation in the sequence shown. The infusion and production populations may or may not be formed depending upon circumstances.

We'll focus on this component

White 1987 A conceptual framework for tree improvement programs. New Forests 4:325

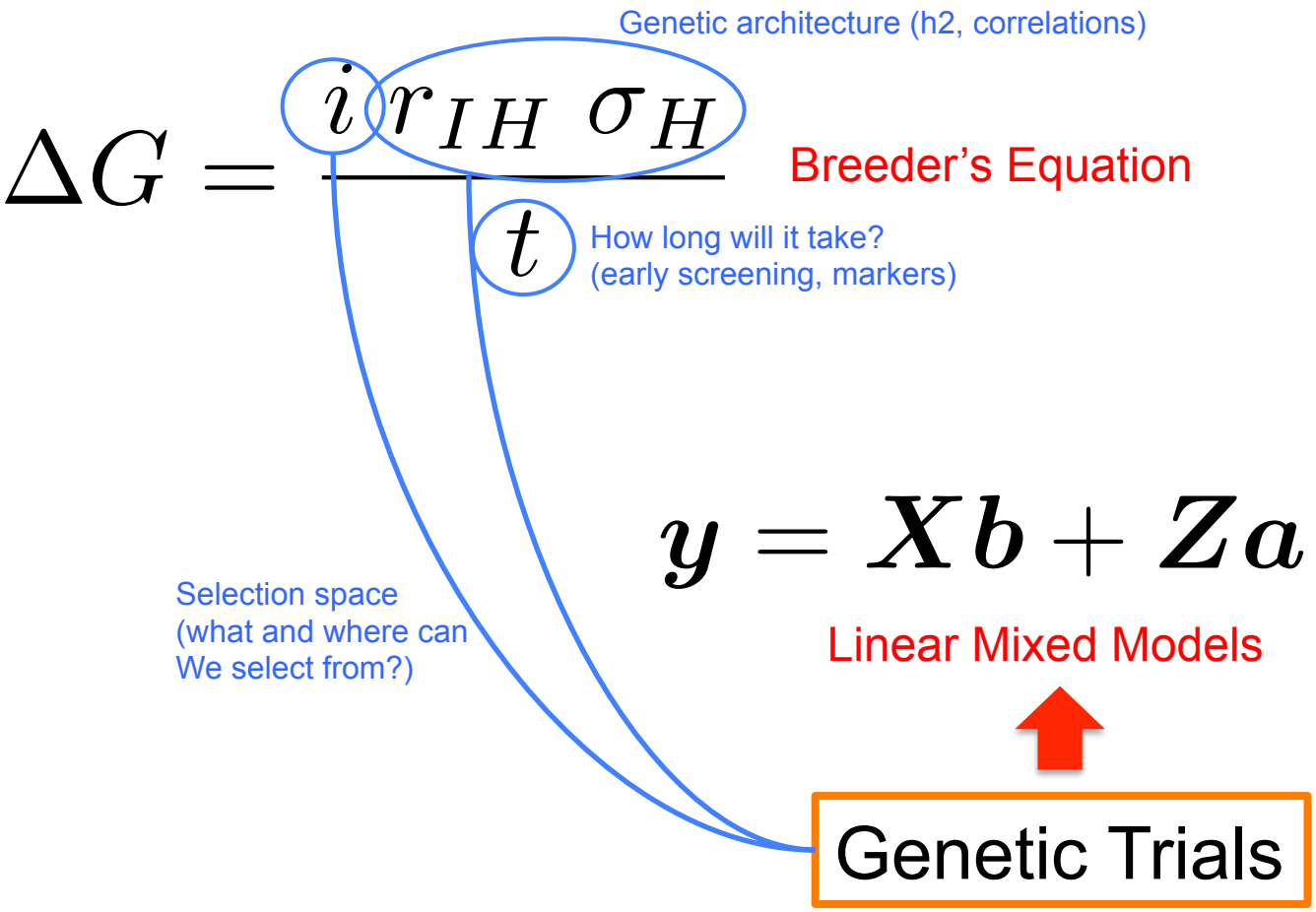
Traditional genetic test

200 families, 30 trees each

Early screening of wood properties (2011-2013)

90 families + 10 clones, 30 trees each

In a nutshell (and coming out of the closet)

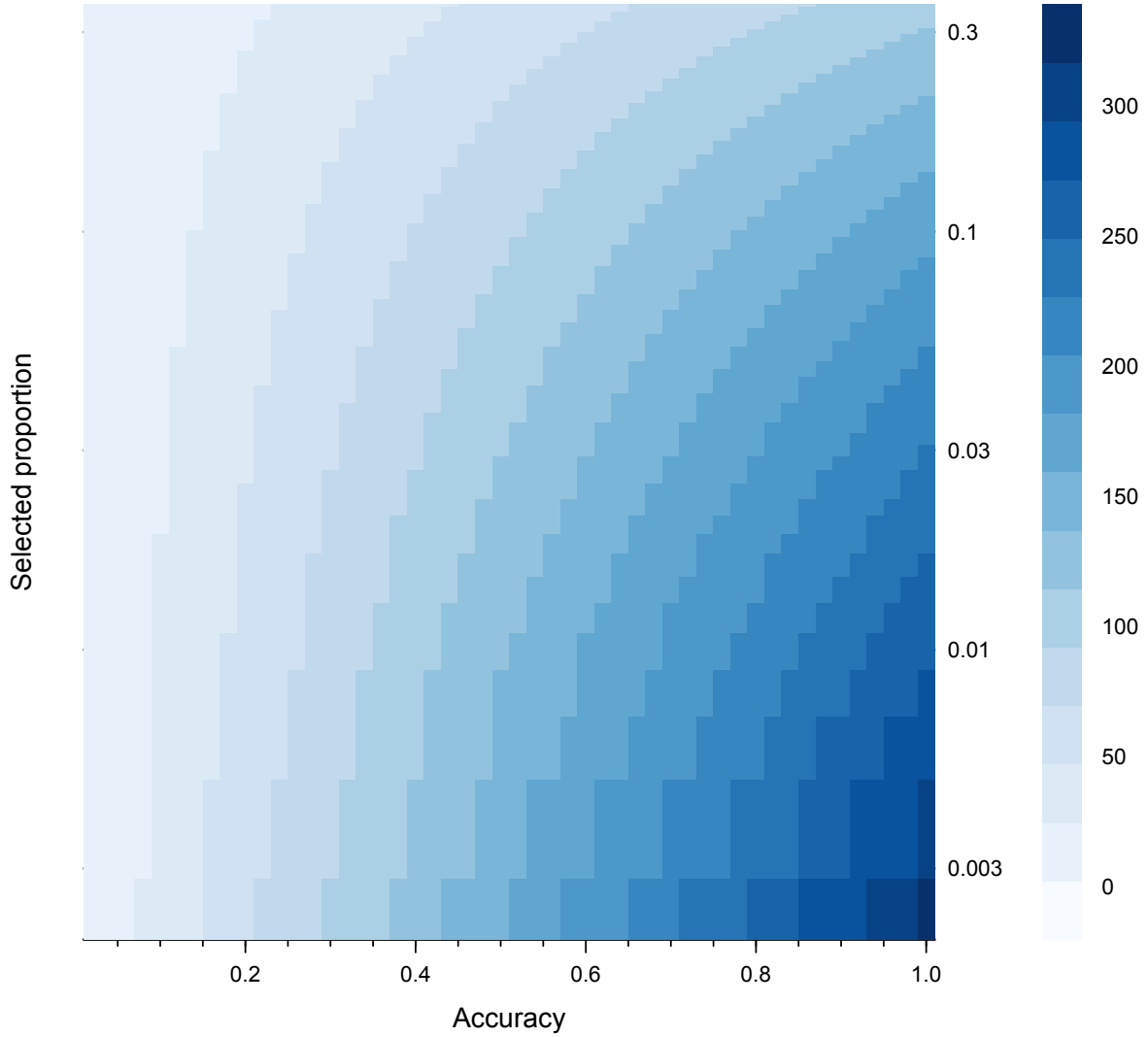


Henderson 1950, 1975abc, 1977, 1984 and a few others.

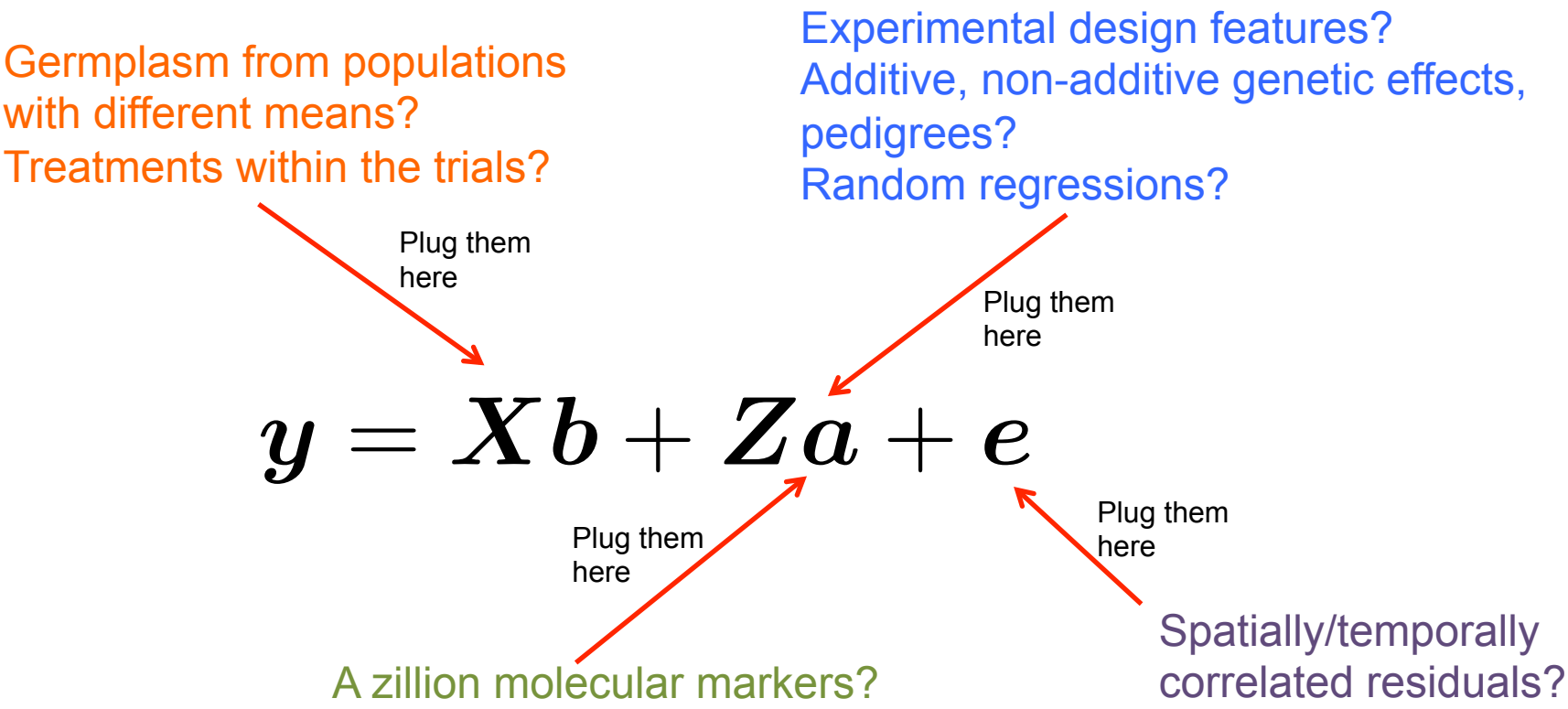
Modern, Ha!

A quick look at the effect of parameters

Genetic gain assuming a constant objective standard deviation (100) and time (1)



Generalized Linear Mixed Models



Frequentist or Bayesian Church? Ecumenical? Use REML, MCMC, INLA or other acronym to fit the model.

Multivariate? Stack up the vectors and matrices and borrow some patience to fit the model.

Generalized Linear Mixed Models

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}$$

Breeders' main interest: estimated genetic parameters & the genetic worth of individuals so we can:

Deploy the best trees

Turnover generations

while maximizing genetic gain

We'll have a look at three examples

GxE: genotype-site matching for maximum value.

Wood quality: avoid further commoditization.

Genomic selection: we want to believe that this time is right.

There is a constant tension between what's **possible** in an evaluation and what's **desirable**.

What we call **modern** is some times very old.

GENOTYPE X ENVIRONMENT

There are two naïve extremes concerning GxE:
There is none | Every site is highly interacting

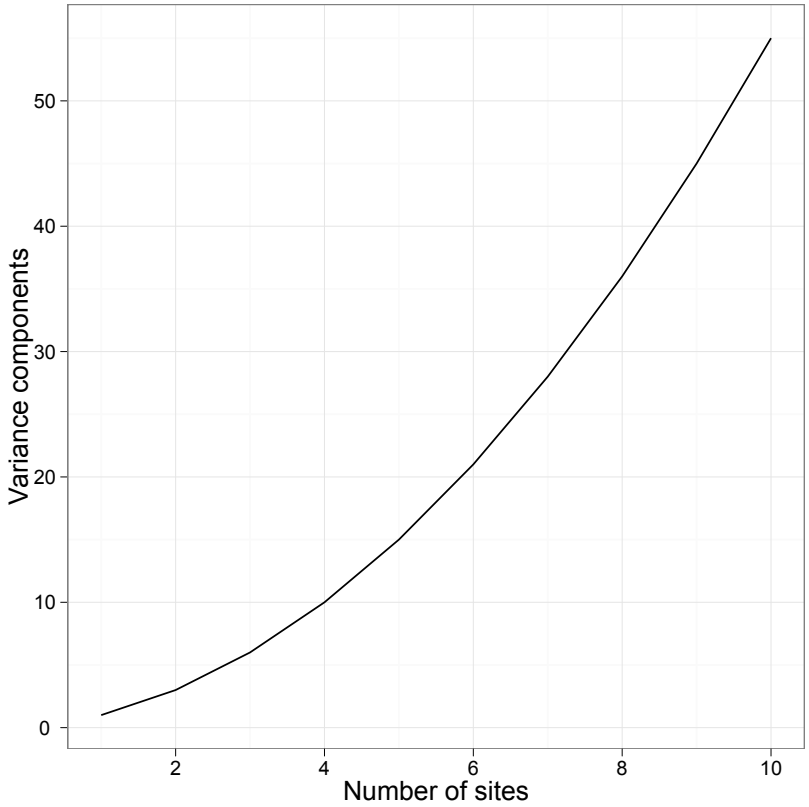
One way to explore this problem is to use a multivariate version of the linear mixed model, considering the genetic worth of each genotype in each site as a different trait.

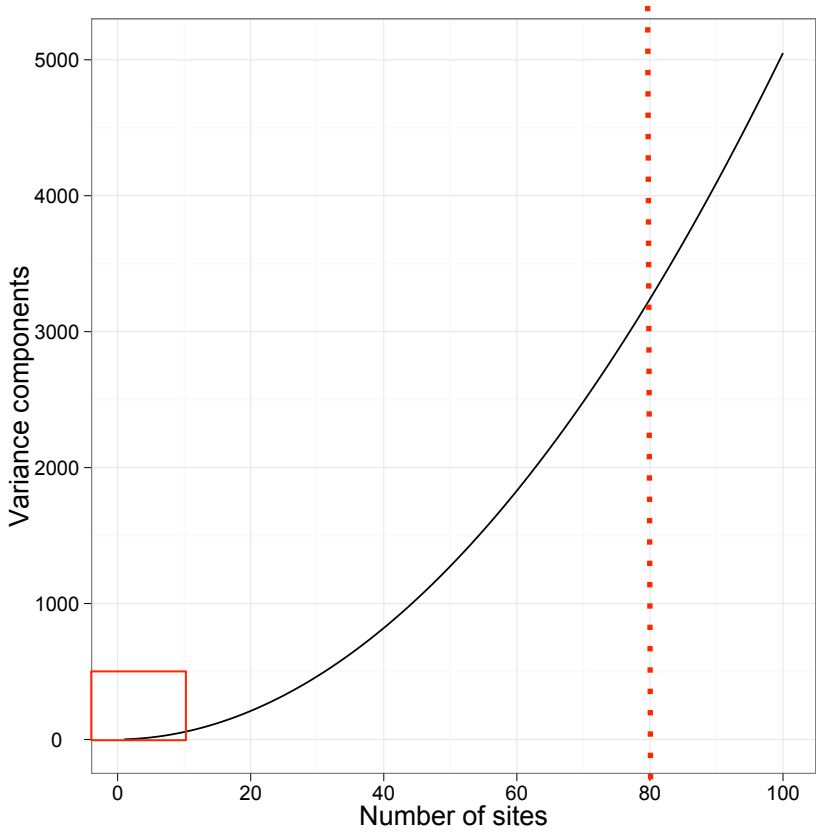
$$\boldsymbol{G} = \begin{bmatrix} \sigma^2_{11} & r_{12} & r_{13} \\ r_{12} & \sigma^2_{22} & r_{23} \\ r_{13} & r_{23} & \sigma^2_{33} \end{bmatrix}$$

Falconer 1952 The problem of environment and selection. The American Naturalist 86:293.

Modern, Ha!

Number of **G** parameters to estimate

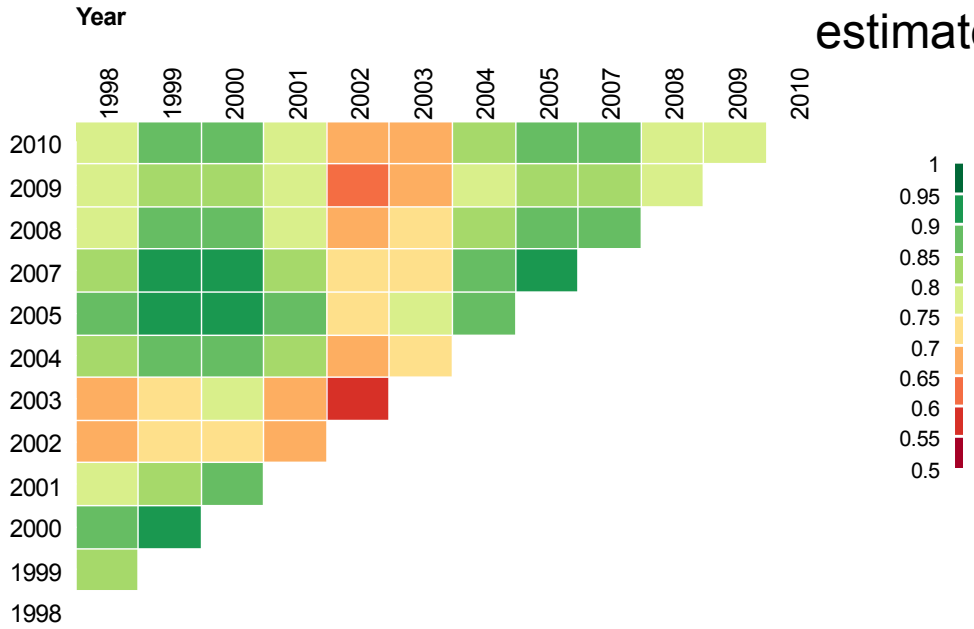




Sites involved in RPBC's evaluation for growth

Example: multi-environment evaluation in other crops

Factor analytic (order 1) covariance model to achieve convergence and estimate 24 instead of 55 parameters



Smith, Cullis & Thompson 2005 The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. J Agric Sci 143: 449.

Paget, Alspach, Genet & Apiolaza 2013 Genetic variance models for the evaluation of resistance to powdery scab (*Spongospora subterranea* f. sp. *subterranea*) from long–term potato breeding trials. *Submitted*.

Even this approach is not good enough for many large trials

Most trees do not have progeny, so we can use a **Reduced** Animal Model*

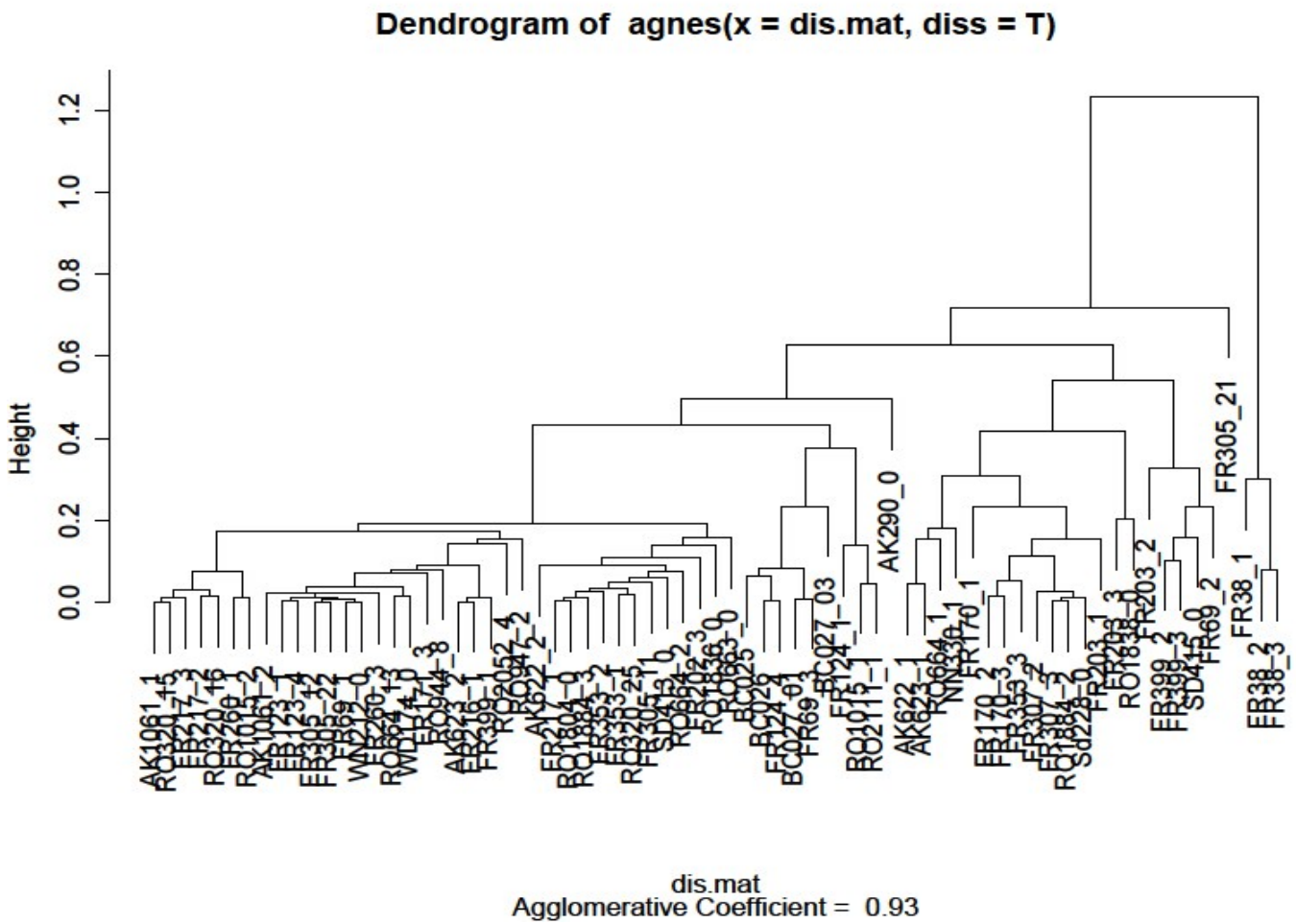
+ a Factor Analytic Model

Combination developed by B.R. Cullis in 2011

*Quaas & Pollack 1980 Mixed model methodology for farm and ranch beef cattle testing programs. J. Anim. Sci. 51:1277.

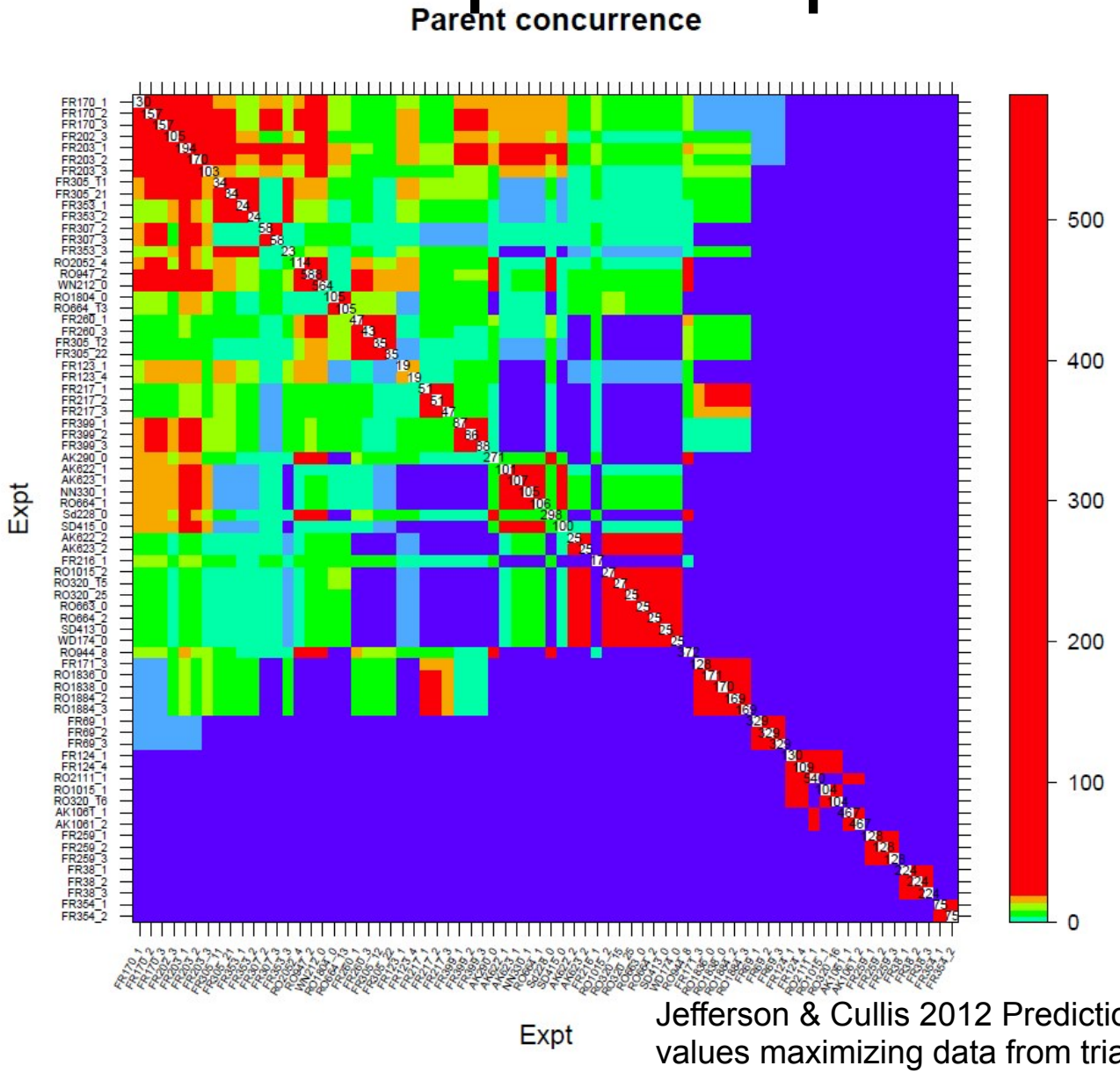
Modern, Ha!

We can classify sites a posteriori based on the correlation matrix



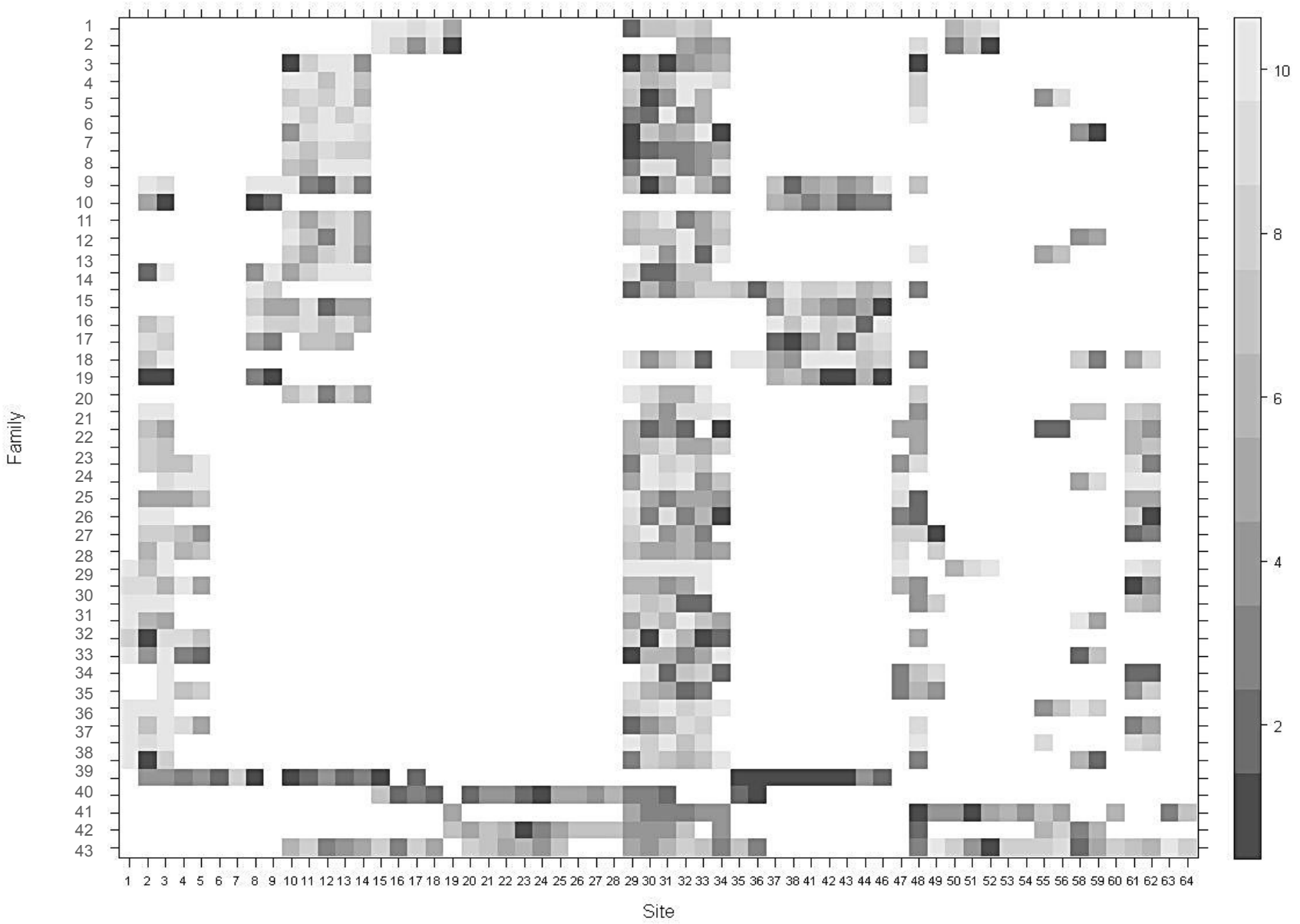
Jefferson & Cullis 2012 Prediction of breeding values maximizing data from trials over 76 sites.

Heat map example 1

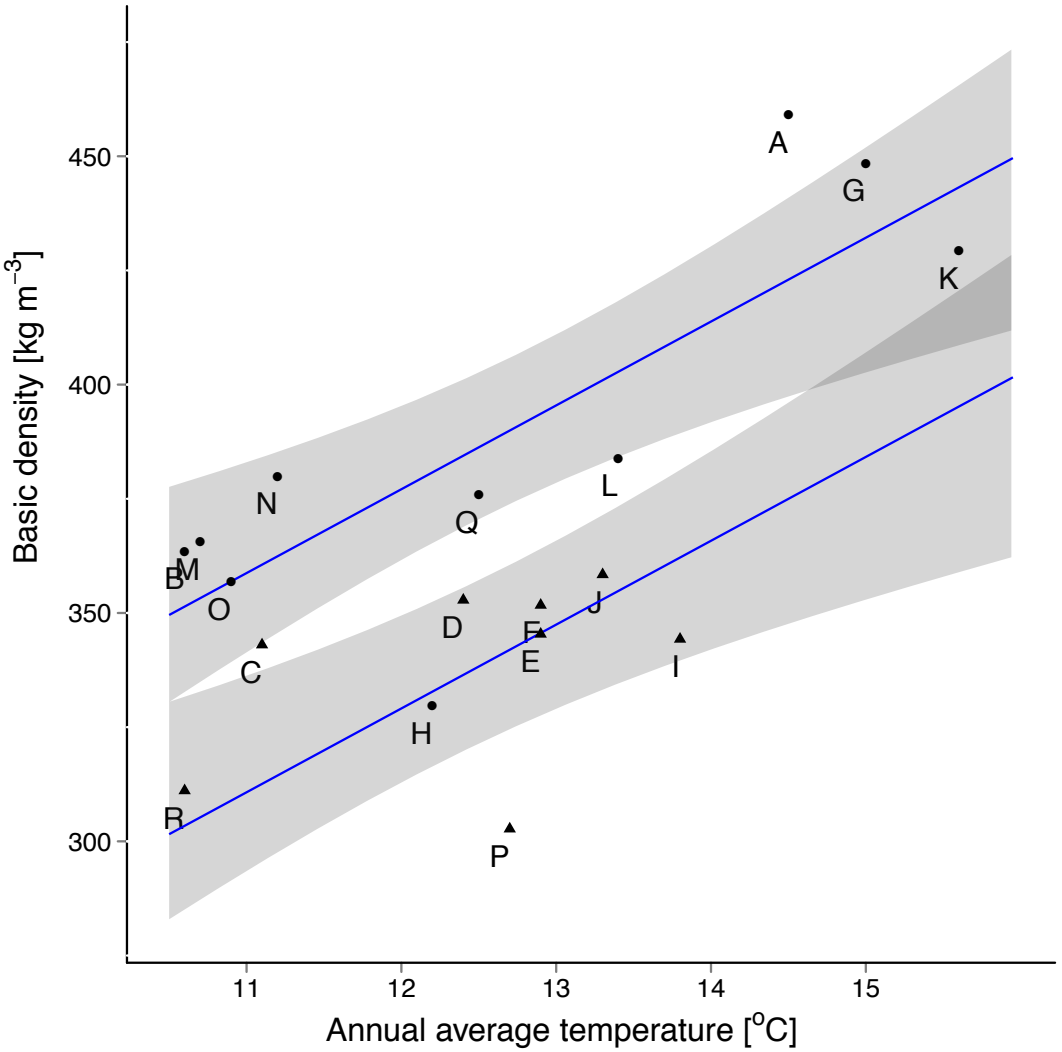


Jefferson & Cullis 2012 Prediction of breeding values maximizing data from trials over 76 sites.

Genotype performance across environments



Some times we have simple explanations for GxE; most times we don't



Rings 6-10

Rings 1-5

Site means based on 23,000+ increment cores in 18 sites

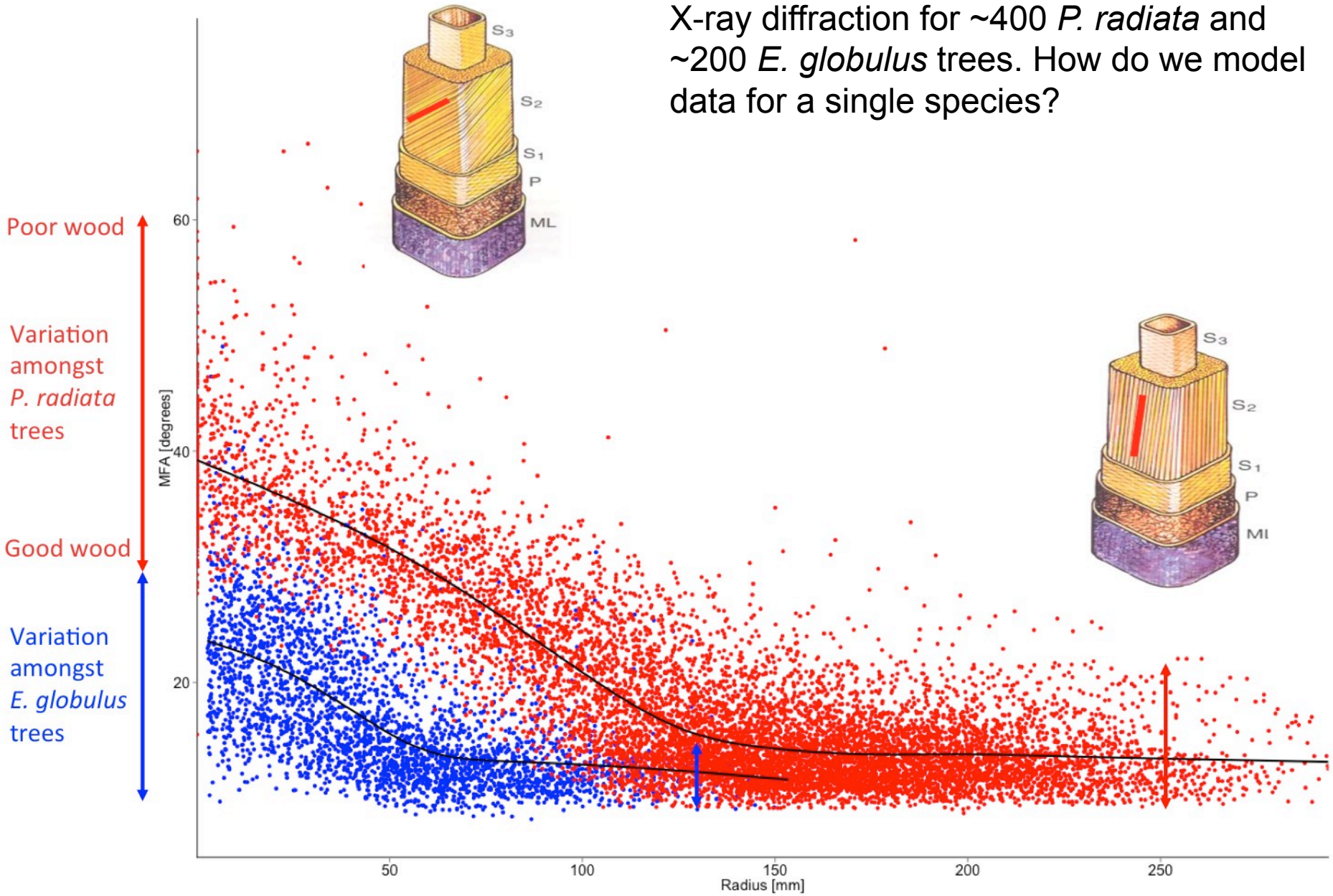
$R^2 = 0.75$

Apiolaza 2012 Basic density of radiata pine in New Zealand: genetic and environmental factors. TGG 8: 87-96

Aim: avoid further commoditization of radiata pine wood by ‘fixing’ poor corewood, which should reduce rotation length for the production of solid wood products.

WOOD QUALITY TRAITS: RESOLUTION

X-ray diffraction for ~400 *P. radiata* and ~200 *E. globulus* trees. How do we model data for a single species?



Learning from phenotypic data

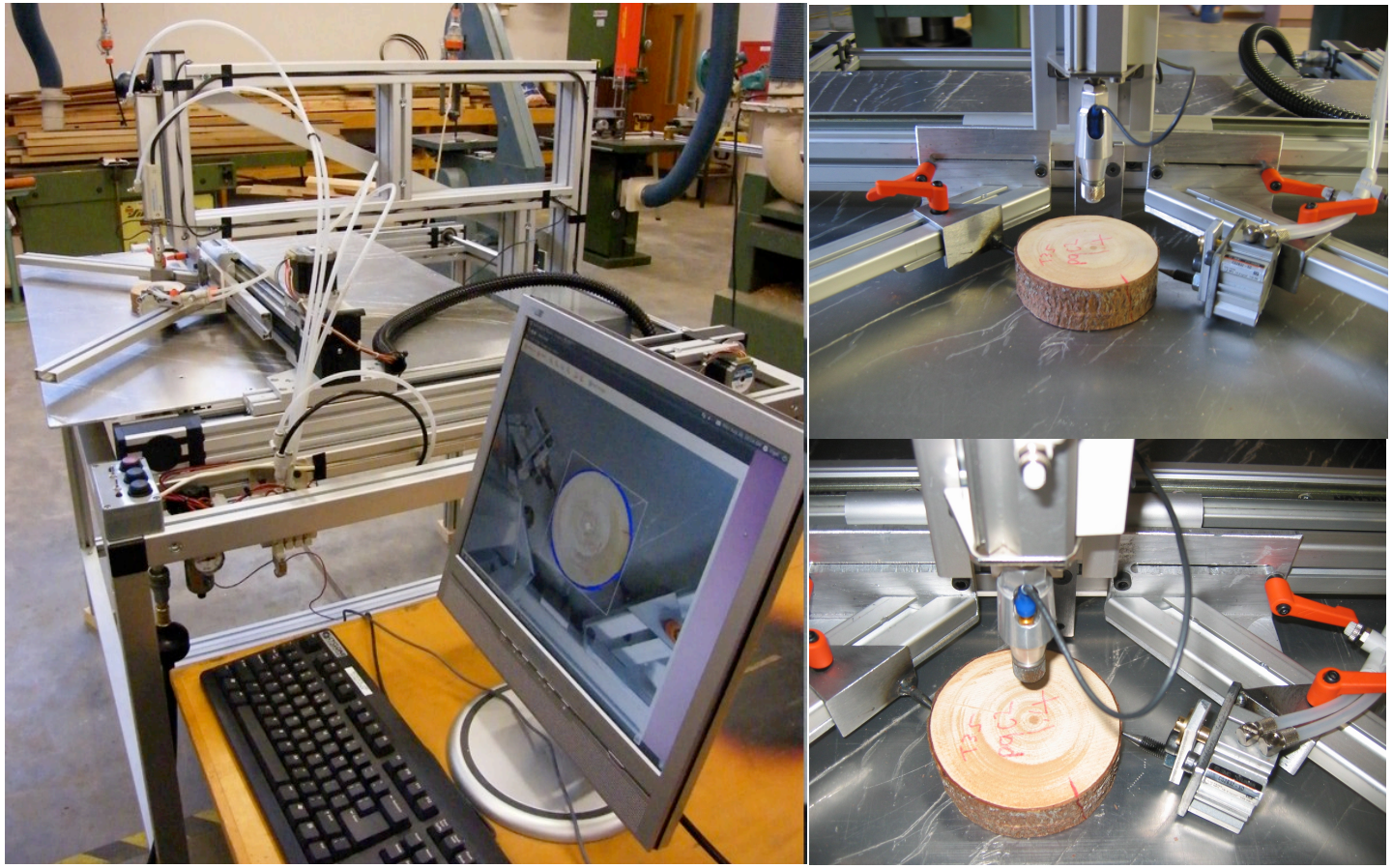
Before embarking in data analysis lucubrations: What can we see/learn from data?

Do we need high stiffness & dimensional stability (low MFA)? Use hardwoods

When do we have maximum variability? Early in the life of trees.

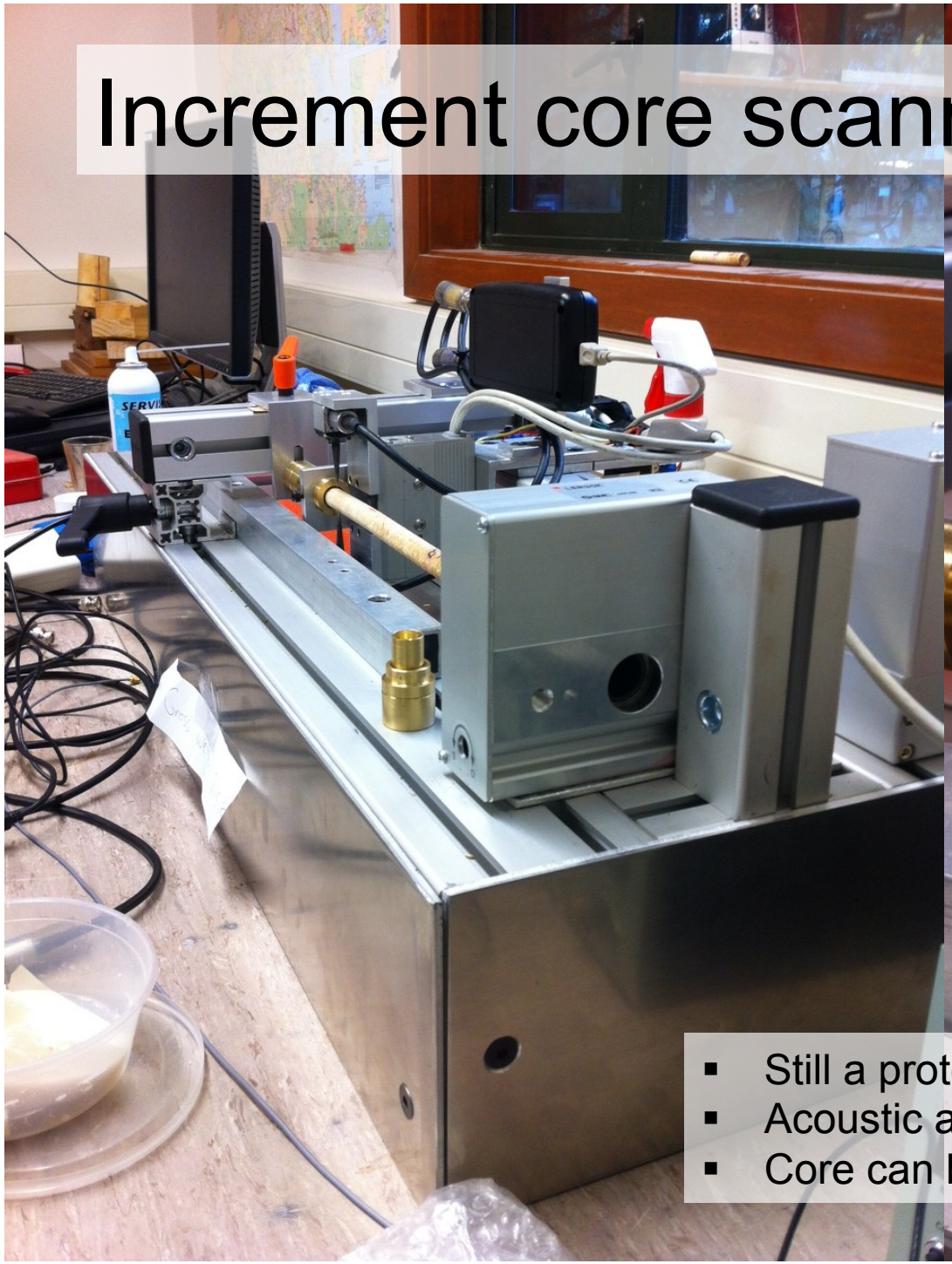
Some tools can provide large numbers of data points per individual. Do we use **all** of them, a **subset** or a **function** of them?

Ultrasonic automated x-y disc scanner



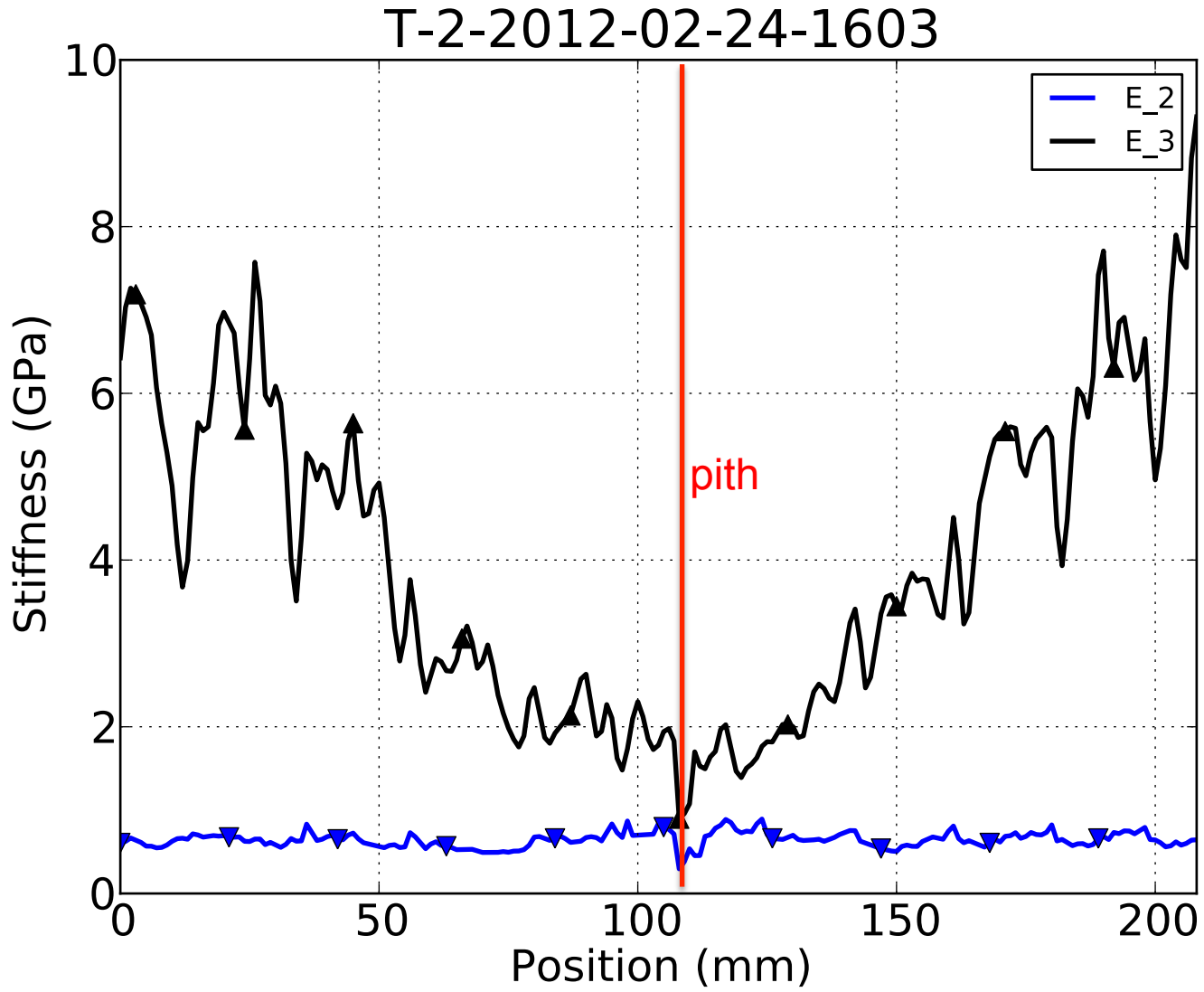
As soon as we showed our new machine to foresters and breeders they said ‘but we don’t want to use disks, **we want to use cores**!’

Increment core scanner



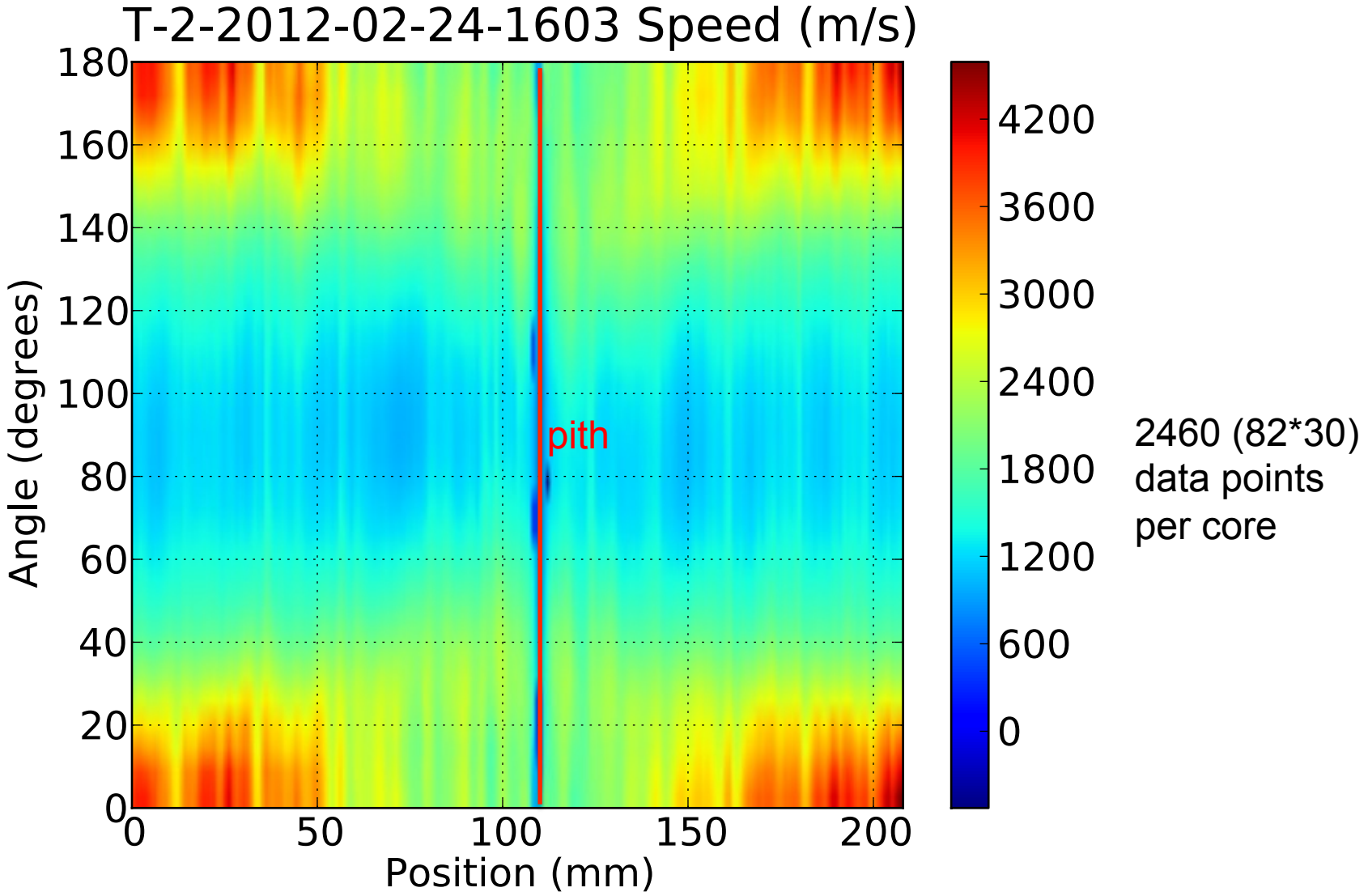
- Still a prototype.
- Acoustic assessments along the core.
- Core can be rotated every 6 degrees.

Acoustic velocity along the core

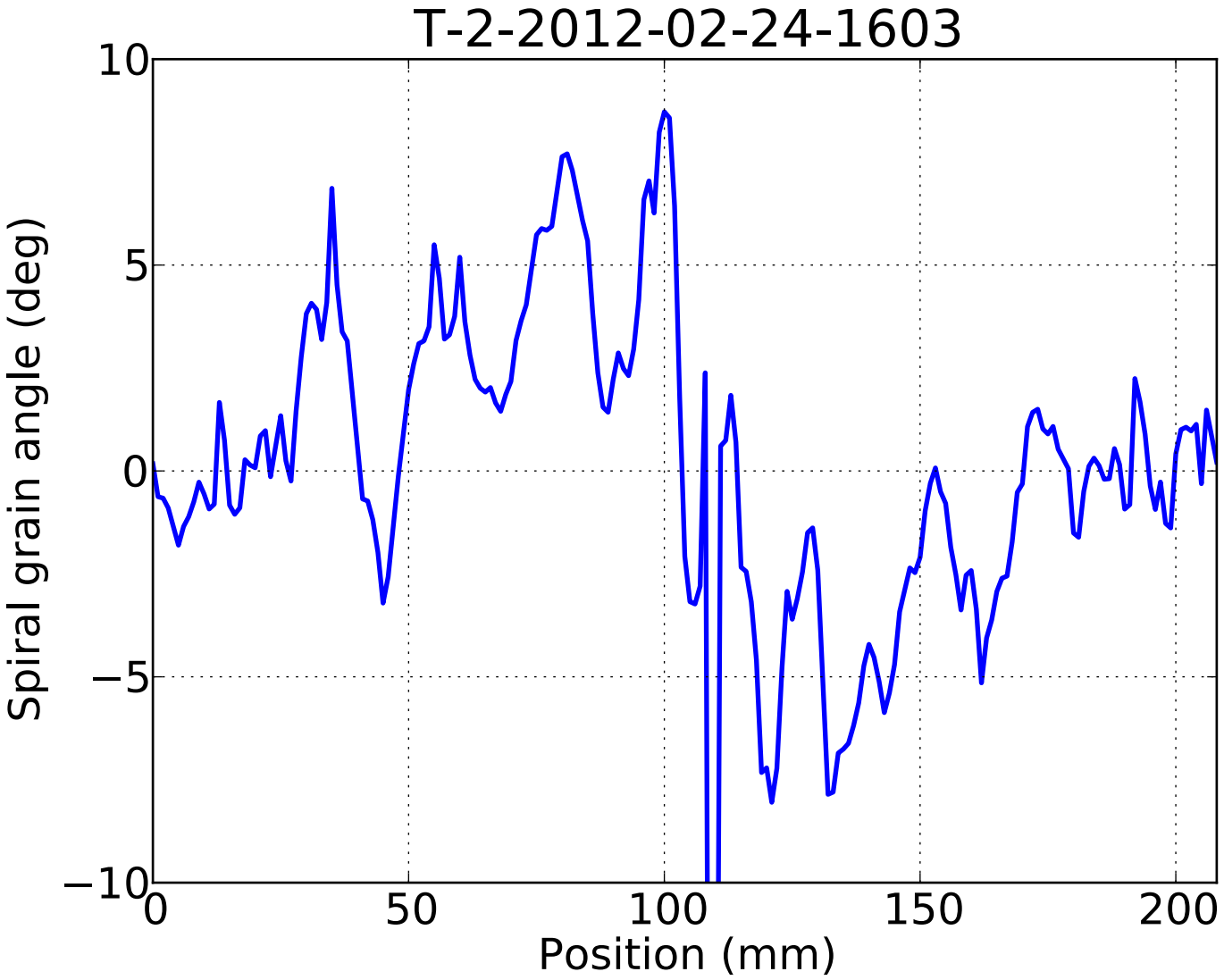


82 data points per core

Acoustic velocity along & around core



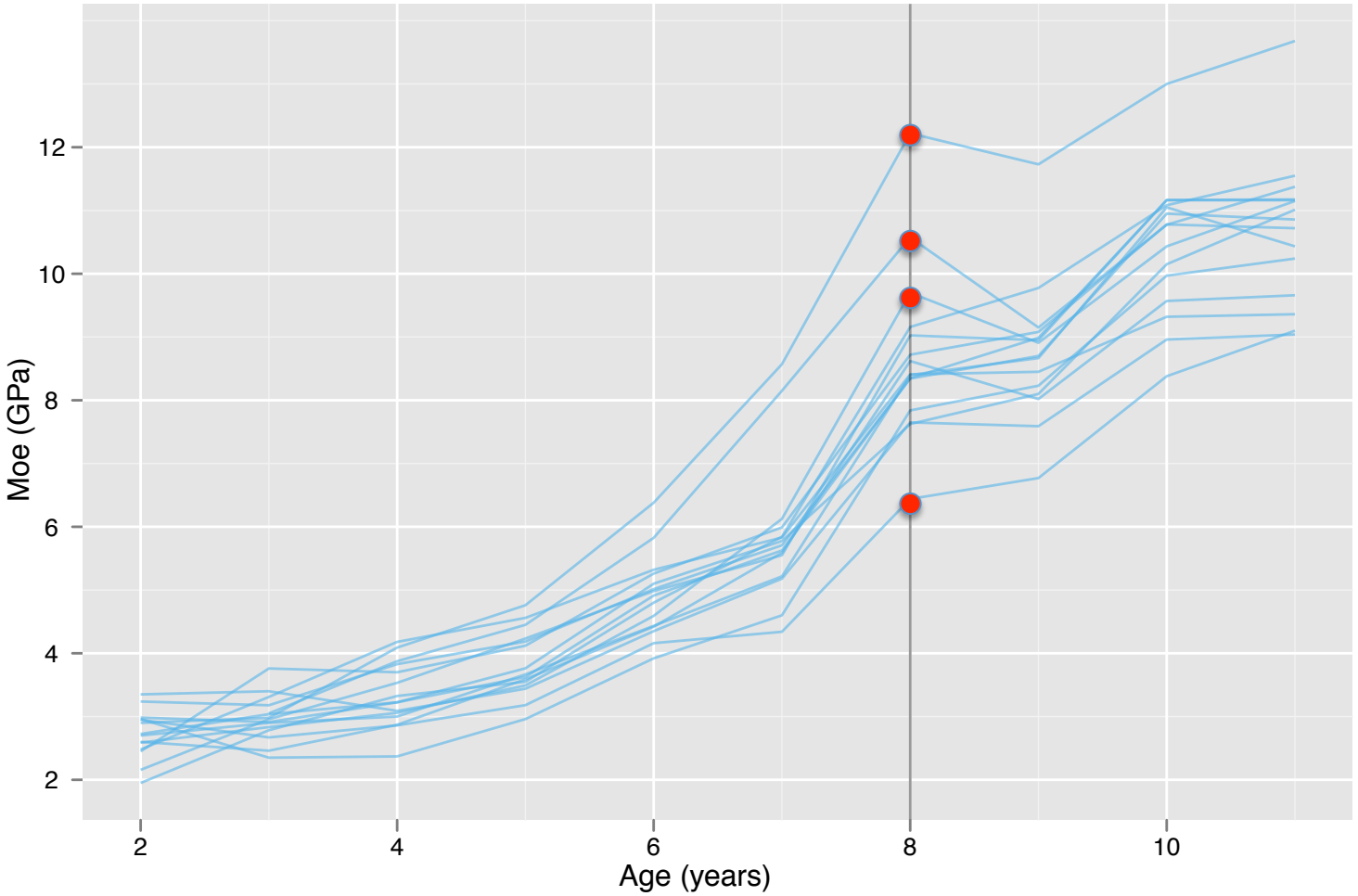
To recover spiral grain along the core



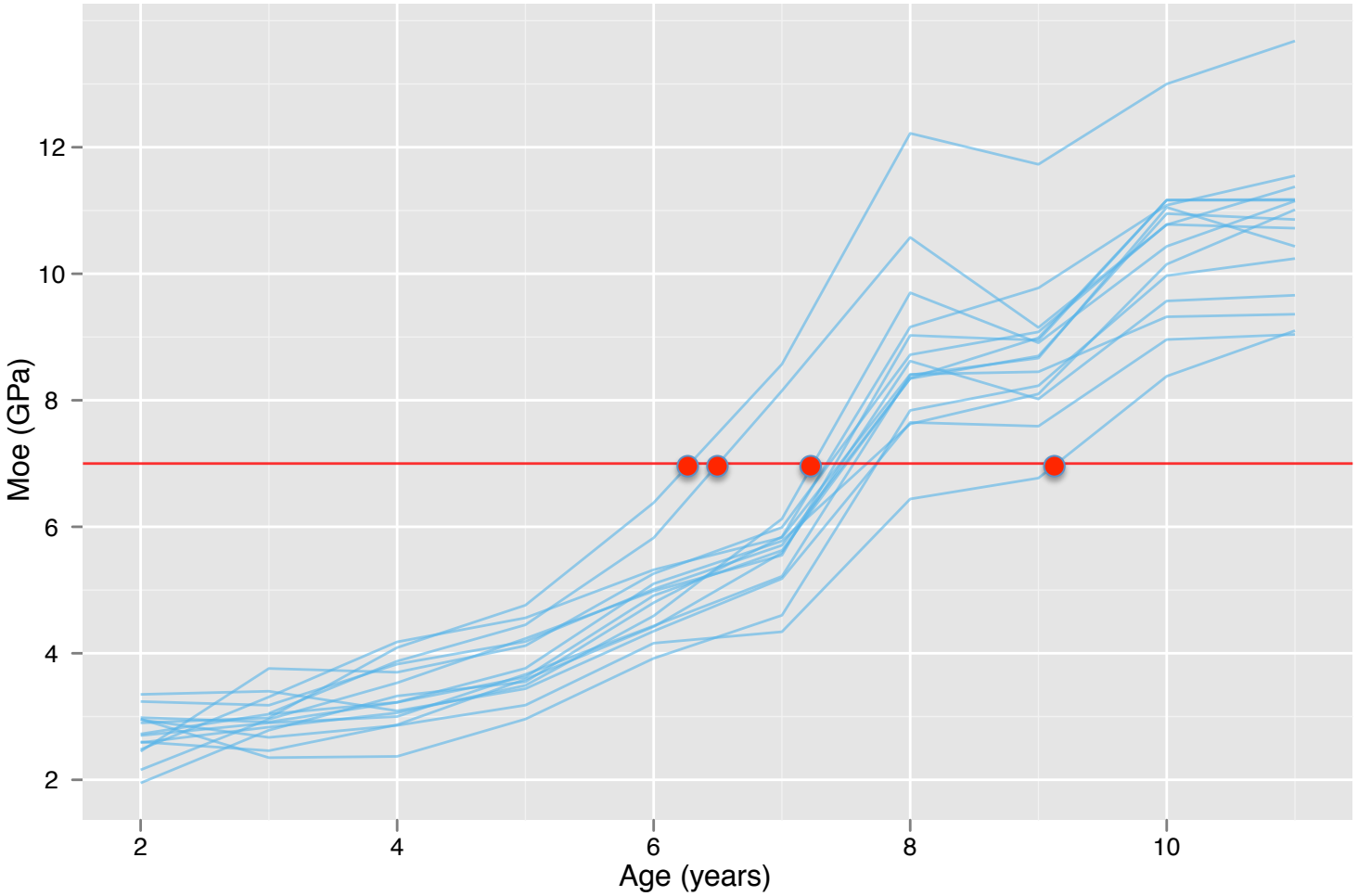
Processing the signal differences when rotating the core we can estimate spiral grain.

82 data points per core

Reframing the selection process: from maximum stiffness to meeting early thresholds



Reframing the selection process: from maximum stiffness to meeting early thresholds



RPBC's
7 GPa

If we want to predict time to threshold then we can select earlier, perhaps at age 2?

Apiolaza (2009) Annals of Forest Science 66:601

Early screening of wood properties (2011-2013)

90 families + 10 clones, 30 trees each

Amberley Seed Orchard

Screening for wood quality the parents of one of the largest orchards in the Southern Hemisphere

ALL OF US WANT TO BE GENOMIC, YEAH RIGHT

"I don't know if this is such a wise thing to do, George."

Probably most breeders feel closer to this cartoon by Larson.

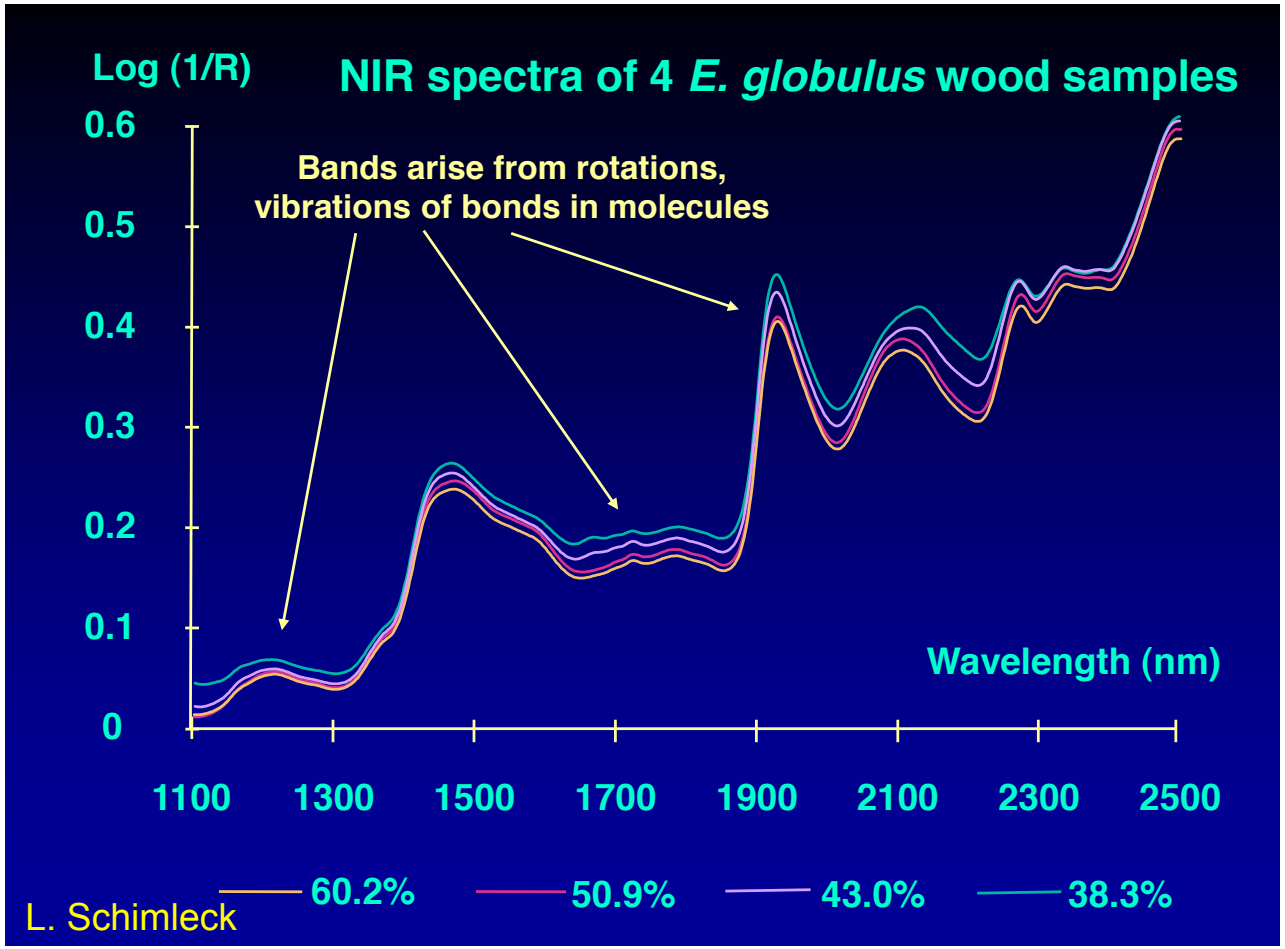
Scary part 1: We are expected to select trees using thousands of predictors (e.g. SNP).

However, some times we do use approaches with some similarities.

Something we could be doing already

response = intercept + pred1 + pred2 + pred3 + … + pred10000

and then we select based on predictions from this model



We have sort of used markers…

New Forests
May 2002, Volume 23, Issue 3, pp 177-191

Gene flow between introduced and native *Eucalyptus* species

Robert C. Barbour, Brad M. Potts, René E. Vaillancourt, Wayn

» Download PDF (378 KB)

Abstract

The first evidence of *in situ* F_1 hybridisation between an introc
Eucalyptus nitens, and a native eucalypt species is presented
from a mature *E. nitens* trial and from the adjacent native spe
island of Tasmania. Nearly 70 000 seedlings were grown to a
distinguished from pure species seedlings on the basis of mo
isozyme allele. Hybridisation was observed between *E. nitens*
E. viminalis were found. This pattern of hybridisation was con
between the *E. ovata* and *E. nitens*. *Eucalyptus nitens* progen
homogeneous level of hybridisation, averaging 0.15% per tre
hybrids obtained from the adjacent *E. ovata* trees varied from
progeny arising from such hybridisation will survive and grow
and introgression of the exotic genes into the native populatic

Thursday, February 7, 2013 Program Agenda (Day 4)

07:30 – 08:30	Continental Breakfast – *Ballroom Corridor*	
	Plenary Session 4-I – *Florida Salon D*	
08:30 – 09:15	**Molecular markers for dissecting trait architecture and for selection**, John Davis, University of Florida, USA [1005]	
09:15 – 10:00	**Genetic mapping rust resistance genes in pine: Implications for selection and breeding**, C. Dana Nelson, USDA Forest Service, USA [1011]	
10:00 – 10:25	Refreshment Break – *Ballroom Corridor*	
	Concurrent Session 4A – *Florida Salon D*	Concurrent Session 4B – *Omni Ballroom*
10:25 – 10:50	**Circumventing graft incompatibility in *Pinus maximinoi* by air-layering and needle fascicle propagation**, Nhora Isaza, Smurfit Kappa Carton de Colombia, COLOMBIA [1485]	**Performance of loblolly pine (*Pinus taeda*) varieties at age six years: Genotype by environment interaction, age-age correlations and predicted genetic gain**, Patrick Cumbie, Arborgen, USA [1465]
10:50 – 11:15	**Pitch canker fungus inoculation screening and early field-growth of clonally propagated and DNA finger-printed *Pinus patula* x *Pinus tecunumanii* hybrid polymix families**, Andre Nel, Sappi, SOUTH AFRICA [1469]	**Interaction between loblolly pine varieties and silvicultural intensity: Effects on the 4-year growth and leaf-level physiology**, Marco Yanez, Virginia Tech, USA [1474]
11:15 – 11:40	**Integrating association and expression analyses to discover genes regulating oleoresin production in loblolly pine**, Jared W. Westbrook, University of Florida, USA [1470]	**Growth and wood properties of loblolly pine clonal varieties**, Finto Antony, University of Georgia, USA [1489]
11:40 – 12:05	**Detection of candidate genes for fusiform rust resistance using genomic selection and association mapping**, Tania Quesada, University of Florida, USA [1505]	**The performance of loblolly pine varietals developed from advanced generation parents in the southeastern United States and tested at multiple sites in Brazil and Argentina**, Michael Cunningham, Arborgen, USA [1502]

Scary part: moving target

NATURE | LETTER

Finding the sources of missing heritability in a yeast cross

Joshua S. Bloom, Ian M. Ehrenreich, Wesley T. Loo, Thúy-Lan Võ Lite & Leonid Kruglyak

Affiliations | Contributions | Corresponding author

Nature (2013) | doi:10.1038/nature11867

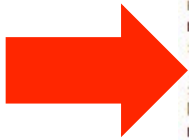
Received 27 June 2012 | Accepted 14 December 2012 | Published online 03 February 2013

Larger populations, denser sets of markers, better models will ‘find’ the trait

Here we use a large cross between two yeast strains to accurately estimate different sources of heritable variation for 46 quantitative traits, and to detect underlying loci with high statistical power. We find that the detected loci explain nearly the entire additive contribution to heritable variation for the traits studied. We also show that the contribution to heritability of gene–gene interactions varies among traits, from near zero to approximately 50 per cent. Detected two-locus interactions explain only a minority of this contribution. These results substantially advance our understanding of the missing heritability problem and have important implications for future studies of complex and quantitative traits.

Not so scary part: analysis are doable

This code will train a model using Bayes A for any number of markers



```
nmarkers = 2000;    # number of markers
startMarker = 1981; # set to 1 to use all
numiter  = 2000;    # number of iterations
vara     = 1.0/20.0;

# input data
data     = matrix(scan("trainData.out0"),ncol=nmarkers+2,byrow=TRUE);
nrecords = dim(data)[1];

beg = Sys.time()

# x has the mean followed by the markers
x = cbind(1,data[,startMarker:nmarkers]);
y = data[,nmarkers+1];
a =  data[,nmarkers+2];
# inital values

nmarkers = nmarkers - startMarker + 1;
mean2pq = 0.5;                          # just an approximation
scalea  = 0.5*vara/(nmarkers*mean2pq);  # 0.5 = (v-2)/v for v=4

size = dim(x)[2];
b = array(0.0,size);
meanb = b;
b[1] = mean(y);
var  = array(0.0,size);

# adjust y
 ycorr = y - x%*%b;

# MCMC sampling
for (iter in 1:numiter){
  # sample vare
  vare = ( t(ycorr)%*%ycorr )/rchisq(1,nrecords + 3);

  # sample intercept
  ycorr = ycorr + x[,1]*b[1];
  rhs = sum(ycorr)/vare;
  invLhs = 1.0/(nrecords/vare);
  mean = rhs*invLhs;
  b[1] = rnorm(1,mean,sqrt(invLhs));
  ycorr = ycorr - x[,1]*b[1];
  meanb[1] = meanb[1] + b[1];

  # sample variance for each locus
  for (locus in 2:size){
    var[locus] = (scalea*4+b[locus]*b[locus])/rchisq(1,4.0+1)
  }

# sample effect for each locus
  for (locus in 2:size){
    # unadjust y for this locus
    ycorr = ycorr + x[,locus]*b[locus];
    rhs = t(x[,locus])%*%ycorr/vare;
    lhs = t(x[,locus])%*%x[,locus]/vare + 1.0/var[locus];
    invLhs = 1.0/lhs;
    mean = invLhs*rhs;
    b[locus]= rnorm(1,mean,sqrt(invLhs));
    #adjust y for the new value of this locus
    ycorr = ycorr - x[,locus]*b[locus];
    meanb[locus] = meanb[locus] + b[locus];
  }
}

Sys.time() - beg

meanb = meanb/numiter;
aHat  = x %*% meanb;
```

or using ASReml-R

wgaim: Whole Genome Average Interval Mapping for QTL detection using mixed models

This package integrates sophisticated mixed modelling methods with a whole genome approach to detecting significant QTL in linkage maps.

Version:	1.3-0
Depends:	R (≥ 2.0.0), qtl, lattice
Published:	2012-09-11
Author:	Julian Taylor, Simon Diffey, Ari Verbyla and Brian Cullis.
Maintainer:	Julian Taylor <julian.taylor at adelaide.edu.au>
License:	GPL (≥ 2)
SystemRequirements:	asreml-R 3.x
Citation:	wgaim citation info
In views:	Genetics
CRAN checks:	wgaim results

Markers & GxE

It makes sense to use clones replicated across environments to train the models

e.g. Resende, Muñoz Del Valle, Acosta, Resende, Grattapaglia & Kirst 2012 Stability of Genomic Selection prediction models across ages and environments.

Can we move in forestry from a piecemeal approach to run a program (and redesign a program) based on genomic selection?

Sales pitch has some limits

In briefly reviewing a small fraction of the prodigious efforts to map G-P, we emphasize the extreme entanglement of the effects of numerous genes and of environmental influences on phenotype. Beyond this, organisms alter their environments, which reciprocally affect the organisms' own phenotypes, as well as those of surrounding organisms. Consequently, complete knowledge of a genome's loci and existing and potential allelic variants cannot, in principle, account for the phenotypic variation of multicellular organisms, except under exceedingly restrictive, unrealistically simplified genetic and environmental conditions.

Travisano & Shaw 2012. Lost in the map. Evolution 67(2): 305-314.

In summary I

The development of new assessments (either phenotypes or markers) will exponentially increase available information & the size of our problems.

We'll reach points when solving the problem becomes unfeasible. Options: more complex algorithms and/or redefining the problem.

GxE: alternative models can cope with massive multivariate approaches.

Solid wood properties: redefine the problem.

In summary II

Markers are an odd one: IMHO a small-scale intervention in the program has little use.

Large-scale intervention makes much more sense, again IMHO, but it's also quite risky.

Training across sites (to account for GxE) may turn up to be quite expensive unless one can rely on good clonal coverage across sites.

Acknowledgements: John Walker, Ryogo Nakada, Clemens Altaner & Paul McLean (wood quality), Brian Cullis, Tim McDonald & Mark Paget (GxE).