OA 41

Use of prediction error co-variances for determining connectedness among test series

Richard Kerr[1], Gregory Dutkowski[1], Jansson Gunnar[2], Johan Westin[2], Persson Torgny[2]
[1]PlantPlan Genetics, Australia; [2]Skogforst, Sweden

Establishing connectedness across multiple test series is an important component of forest tree breeding programs. It is well recognized that the genetic means of each series can change especially if series comprise different sub-populations (genetic groups). Simultane ously estimating genetic group and environmental effects is made more accurate if there are genetic connections between the series. Even if genetic group effects are absent genetic connectedness is important for improving the accuracy of a contrast between genotypes from different series.
The accuracy of a genetic value contrast is the correlation between the predicted and the true difference of genetic values and is derived using prediction error co-variances (PEC) and the additive relationships between the individuals in question. The PEC are obtained by inverting the coefficient matrix of the mixed model equations. The aim of our study was to determine the best strategy to connect multiple test series, where best is defined as improving the average accuracy of a genetic value contrast while minimizing resources needed to achieve connection. Connection is achieved by sharing common reference material. The source and amount of reference material defined each strategy. The percentage of reference material relative to test material ranged between 0% (total disconnection) to 100% (full connection). It was shown that use of over-replicated families as the source of reference material was the most efficient strategy. Test families for a series will have variabl e numbers of seedlings. Some will have sufficient numbers to plant as replicated treatments in other series. If few as 10% of the families in each series are tested across all series, the accuracy of comparing two genotypes is the same regardless of whether they are from the same series or not. Implementing such a strategy also improves the accuracy of individual genetic values, even for those progeny in families not over-replicated.

Genetic connectedness and accuracy of selection

Richard Kerr, Greg Dutkowski, Gunnar Jansson, Torgny Persson & Johan Westin

Multiple trial series are commonplace in forest tree breeding

- Not possible to test all material at once, especially in the initial generation
- Breeding of sub-populations (e.g. Norway spruce in Sweden)

BLUP analysis incorporating data from all series will lead to greater gains

- More accurate selection
- Greater selection intensity

A traditional method for linking trial series is to use common check-lots (non-test material)

The outcome of this study was to demonstrate that use of common test material is more optimal

- leads to faster genetic gain
- is just as cost effective

Searle (1987) Linear models for unbalanced data

"... if not all comparisons between effects fitted in the linear model are estimable, then the data are said to be disconnected ".

This statement only applies to fixed effect models

In BLUP, breeding values are fitted as random effects

- Comparisons between individuals are always meaningful
- Disconnectedness is never really an issue
- The issue is better stated as

"**how to optimally link trials so that comparisons between individuals are accurate and precise as possible**"

Laloë (1993) – the coefficient of determination (CD) of a linear contrast - $\mathbf{x'\hat{u}}$

For example contrasting individual 1 with 3

$$\underset{\mathbf{x'}}{\begin{pmatrix} 1 & 0 & -1 & 0 \end{pmatrix}} \overset{\hat{\mathbf{u}}}{\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \end{pmatrix}}$$

The square root of the CD is referred to as accuracy

We derive accuracies by inverting the coefficient matrix (C) of the mixed model equations (MME)

Fixed trial site effects (s) — Random within-site block effects (b) — Solution vector

$$\begin{pmatrix} X'X & X'W & 0 & X'Z \\ W'X & W'W + I * 1/\gamma_1 & 0 & W'Z \\ 0 & 0 & Q'A^{-1}Q * 1/\gamma_3 + I * 1/\gamma_2 & -A^{-1}Q * 1/\gamma_3 \\ Z'X & Z'W & -Q'A^{-1} * 1/\gamma_3 & Z'Z + A^{-1} * 1/\gamma_3 \end{pmatrix} \begin{pmatrix} s \\ b \\ g \\ u^* \end{pmatrix}$$

Random provenance effects (g) — Random additive genetic effects (u)

Invert C to get PEV and ACC of contrast

$$\mathbf{C}^{-1} = \begin{pmatrix} C^{ss} & C^{sb} & C^{sg} & C^{su} \\ C^{bs} & C^{bb} & C^{bg} & C^{bu} \\ C^{gs} & C^{gb} & C^{gg} & C^{gu} \\ C^{us} & C^{ub} & C^{ug} & C^{uu} \end{pmatrix}$$

$$PEV(\mathbf{x}) = \mathbf{x}'\mathbf{C}^{uu}\mathbf{x}\sigma_e^2$$

Elements in the sub-matrix $\mathbf{C}^{uu}$ are used to compute prediction error variances (PEV) and accuracies (ACC)

$$ACC(\mathbf{x}) = \sqrt{1 - \frac{\mathbf{x}'\mathbf{C}^{uu}\mathbf{x}\sigma_e^2}{\mathbf{x}'\mathbf{A}\mathbf{x}\sigma_a^2 + \mathbf{x}'\mathbf{QIQ}'\mathbf{x}\sigma_g^2}}$$

x'Cx just sums diagonal elements and subtracts the off-diagonals

$$\begin{pmatrix} 1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} C^{11} & C^{12} & C^{13} & C^{14} \\ C^{21} & C^{22} & C^{23} & C^{24} \\ C^{31} & C^{32} & C^{33} & C^{34} \\ C^{41} & C^{42} & C^{43} & C^{44} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}$$

$$\mathbf{x}'\mathbf{C}\mathbf{x} = C^{11} + C^{33} - 2C^{13}$$

Off-diagonals are much harder to compute than diagonals

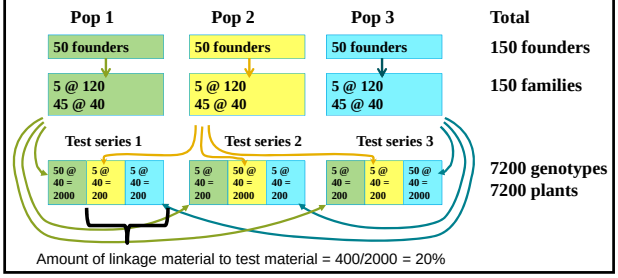
Baseline strategy – no linkage

Pop 1	Pop 2	Pop 3	Total
50 founders	50 founders	50 founders	150 founders
50 @ 40	50 @ 40	50 @ 40	150 families
Test series 1	Test series 2	Test series 3	
50 @ 40 = 2000	50 @ 40 = 2000	50 @ 40 = 2000	6000 genotypes 6000 plants

"50 @ 40" means 50 families each with 40 progeny
DPM is used (i.e. 1 with 2, 2 with 3, etc

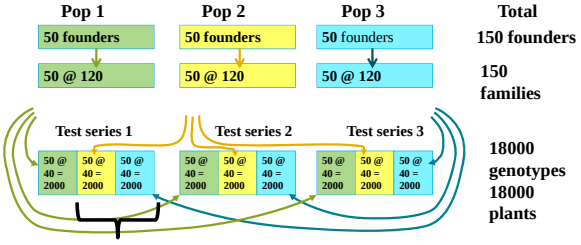
Linkage via excess progeny from common test families

In this example 5 or 10% of families have excess progeny that can be used in other test series



Amount of linkage material to test material = 400/2000 = 20%

Full linkage – every family equally replicated in all test series

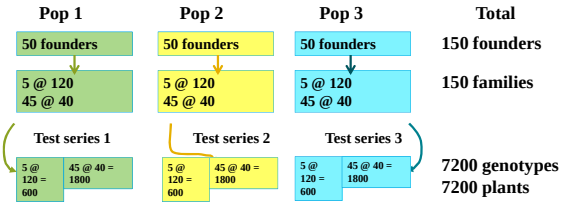


Amount of linkage material to test material = 4000/2000 = 200%

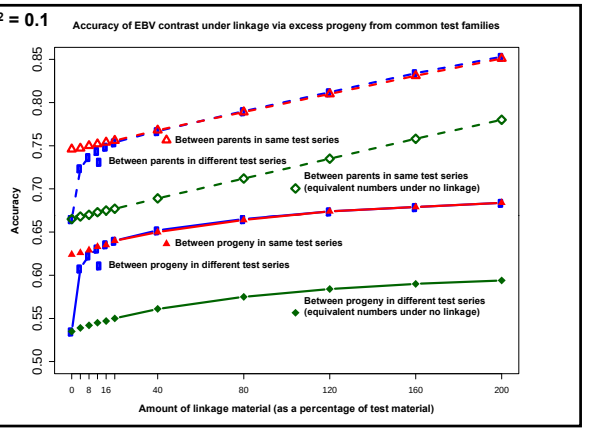
Amount of linkage material relative to test material (%)	Number of genotypes tested per test series	Total number of genotypes tested
0 (baseline case)	2000	6000
2	2040	6120
4	2080	6240
8	2160	6480
12	2240	6720
16	2320	6960
20	2400	7200
40	2800	8400

Tester schemes (for comparison purposes)

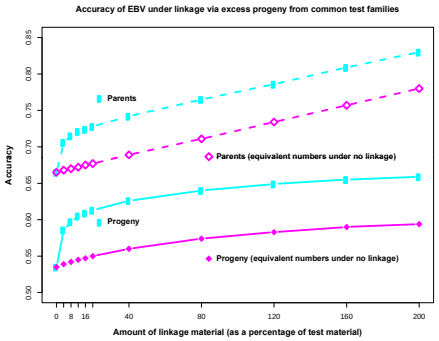
Each test series has a percentage of families that contain more progeny tested.



$h^2 = 0.1$

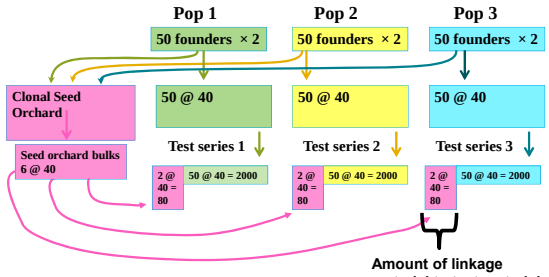


$h^2 = 0.1$



- EBV accuracies increase when test series are linked
- Increase in accuracy even for individuals not belonging to common test families
- Pink line is analogous to running a series of within test series analyses

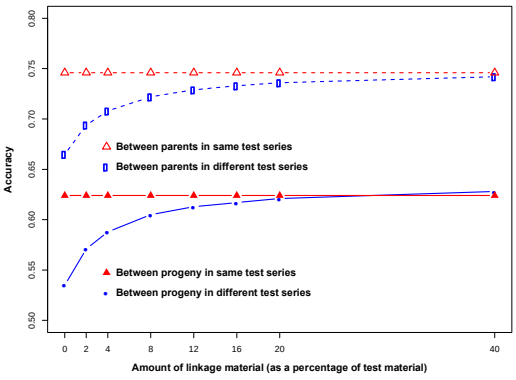
Linkage via bulked orchard seed



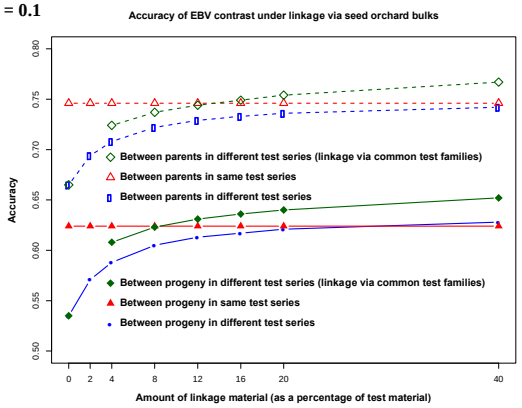
Amount of linkage material to test material = 80/2000 = 4%

Note: each seed orchard bulk has the same female and male parent aggregate

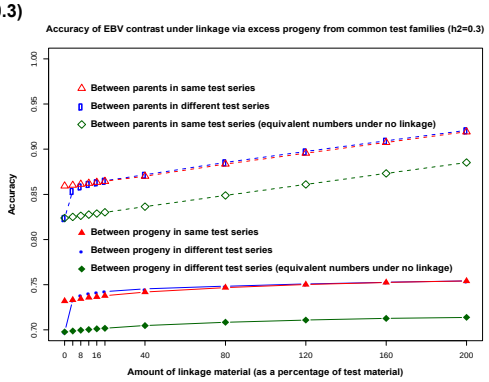
$h^2 = 0.1$



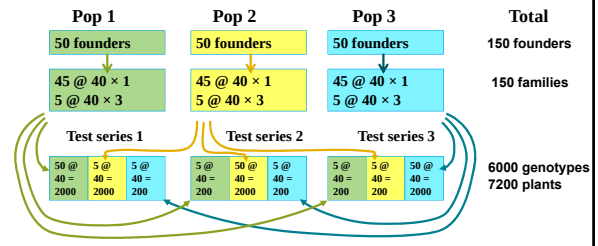
$h^2 = 0.1$



Less linkage material is needed for moderately heritable trait ($h^2 = 0.3$)

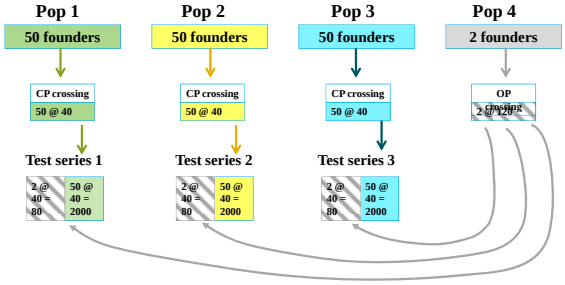


Linkage via cloned progeny from common test families



5 @ 40 × 3 means 5 families each with 40 progeny replicated 3 times

Linkage via open-pollinated seed



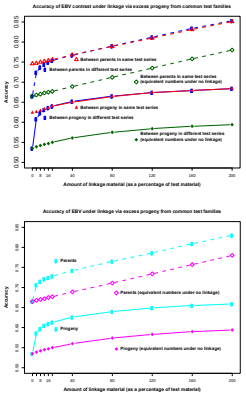
Note: OP progeny have genetic group as male parent

Conclusions

- Linkage via common families, with excess seedlings or with replicated seedlings, was clearly more optimal than any other strategy tested
- The point at which accuracy between individuals of different test series is the same as accuracy between individuals in same test series is a good indicator of the optimum degree of linkage

Conclusions

- Pattern for parents matches closely that for progeny
- Therefore consider only parental accuracy
- Reduce the C matrix such that progeny equations are absorbed
- Makes inverting C much faster
- EBV accuracy tapers off at the point where accuracy of EBV contrasts (between and within) are equal
- Perhaps consider only EBV accuracy



Issues

- What percentage of linkage material should be routinely used?
  - Breeding programs have as their objective the simultaneous improvement of a range of traits with differing heritability
- What are the implications of GxE?
  - E.g growth measured in each test series considered a different trait