
**GENETIC ANALYSIS OF TREE
CLONAL TRIALS:**

**Combining Individual and Clonal
Information into ASReml**

Salvador A. Gezan

University of Florida

sgezan@ufl.edu

BACKGROUND

- Phenotypic value:

$$\mathbf{P} = \boldsymbol{\mu} + \mathbf{G} + \mathbf{E}$$

where,

$$\mathbf{G} = \mathbf{A} + \mathbf{D} + \mathbf{I}$$

hence,

$$\mathbf{P} = \boldsymbol{\mu} + \mathbf{A} + \mathbf{D} + \mathbf{I} + \mathbf{E}$$

A is the **additive** component, i.e. effect of the sum of the genes or breeding value.

D is the **dominance** deviation, i.e. interaction between alleles or within-locus interaction.

I is the **epistatic** deviation, i.e. non additive interactions and higher order interactions.

BACKGROUND

- Clonal Test** A field planting of several vegetatively propagated plants, with the objective of predicting the relative performance of different genotypes.
- Clone** A plant which is genetically identical to the parent plant. Produced asexually, from cuttings, stump, tissue culture.
- Species** Sugar cane, coffee, cassava, poplar, willow, eucalyptus, etc.

Why Clonal trials are relevant?

- Allow to separate all genetic components, i.e. estimate **additive (A)**, **dominant (D)** and **epistatic (I)** effects.
- Helps to understand underlie genetic biology and *guide* future breeding strategies.
- It is possible to *capture* a larger portion of genetic variability by ranking parents, crosses or clones, i.e. achieve greater genetic gains.

CLONAL TRIALS

DIFFICULTIES

- Most breeding programs have been based in trials using single plants (i.e. seedlings) rather than clones (e.g. cuttings) therefore there are several genetics concerns about the change:
 - Are prior selections, *based on seedling trials*, equivalent to selections *based on clonal trials*?
 - Are the relationships among the different traits of interest *similar* for both propagule types (seedlings and cuttings)?

OBJECTIVES

- 1) To evaluate statistically if is possible to combine the genetic effects (additive and dominance) estimated using seedlings and cuttings propagules and to understand their level of association.
- 2) Estimate correlation matrices for phenotypic and genetic effects between a series of size and crown traits in order to facilitate understanding of the their genetic association.

EXPERIMENTS

Series 1: Growth and canopy measurements
collected from sites:

C

N

TRAITS

Trial	C	N
Design Type	IB	IB
# Trees	4400	4480
# Reps	4	4
Rep Size	1100	1120
# IBlocks/rep	110	80
Block Size	10	14
# Families	61	61
# Clones	868	947
# Clones/Family	5-20	16-18
Prop. Missing	38.8%	11.8%

VOL: total stem volume
VCROWN: crown volume
MDIAMB: mean diameter of branches
(sample of 3 branches)
MANG: mean angle of branches
(sample of 3 branches)
RLC: ratio of live crown to total height

~ 20-30% seedlings.

STATISTICAL ANALYSIS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{b} + \mathbf{Z}_2 \mathbf{m} + \mathbf{Z}_3 \mathbf{f} + \mathbf{Z}_4 \mathbf{mf} + \mathbf{Z}_5 \mathbf{c} + \mathbf{e}$$

where,

\mathbf{y} vector of n measurements for a given trait,

μ overall mean,

$\boldsymbol{\beta}$ vector of fixed effects (overall mean and propagule type),

\mathbf{b} vector of incomplete blocks within replicate random effects, $\sim N(0, \sigma_b^2 \mathbf{I}_b)$,

\mathbf{m} vector of mother random effects, $\sim N(0, \mathbf{A} \otimes \mathbf{V}_f)$,

\mathbf{f} vector of father random effects, $\sim N(0, \mathbf{A} \otimes \mathbf{V}_f)$,

\mathbf{mf} vector of family random effects, $\sim N(0, \mathbf{I}_{mf} \otimes \mathbf{V}_{mf})$,

\mathbf{c} vector of clone within family random effects, $\sim N(0, \sigma_c^2 \mathbf{I}_c)$,

\mathbf{e} vector of residual errors, $\sim N(0, \mathbf{D})$,

\mathbf{V}_x matrices of size 2×2 used to specify the variance-covariance of the genetic effects x between seedling and cutting genetic effects.

\mathbf{D} diagonal matrix of size 2×2 used to specify heterogeneity of error variances, i.e. a different error variance component for each propagule type.

\otimes Kronecker product.

STATISTICAL ANALYSIS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

V_x matrices structures

UN unstructured matrix, a matrix that considers three *different* (co)variance components (i.e. it allows to model heterogeneity of variances and covariance among effects),

$$\begin{bmatrix} a & c \\ c & b \end{bmatrix}$$

CS compound symmetry matrix, matrix that is restricted to have the *same variance* for both propagule types but incorporates a non-zero covariance among them, and

$$\begin{bmatrix} a & c \\ c & a \end{bmatrix}$$

DIAG, homogeneous diagonal matrix with *equal variances* for both propagule types but *independence* among effects (i.e. it assumes zero-covariance among effects).

$$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$$

$$c = \rho \times \sqrt{(a \times b)}$$

Here ρ measures the similarity (ranking) in the estimation of genetic effects between propagule types.

STATISTICAL ANALYSIS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

Link between causal components and genetic components.

$$V_A = 4 \sigma_m^2 \quad (\text{when } \sigma_m^2 = \sigma_f^2)$$

$$V_D = 4 \sigma_{mf}^2$$

$$V_I = \sigma_c^2 - (\sigma_m^2 + \sigma_f^2) - 3 \sigma_{mf}^2$$

$$V_G = V_A + V_D + V_I$$

$$V_P = \sigma_m^2 + \sigma_f^2 + \sigma_{mf}^2 + \sigma_c^2 + \sigma^2$$

$$H^2 = V_G / V_P \quad h^2 = V_A / V_P \quad d^2 = V_D / V_P$$

STATISTICAL ANALYSIS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

- Several combinations of variance structures for the parental and family effects were considered and fitted for every traits of interest

Model	Parent Structure (a)	Family Structure (d)
M1	UN	UN
M2	CS	UN
M3	CS	CS
M4	DIAG	DIAG

- All analysis done with ASReml v. 3 (code available upon request).
- Comparisons between models was done using the Akaike information criteria (AIC) and the likelihood ratio test (LRT).

STATISTICAL ANALYSIS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

- Traits considered: MDIAMB, MANG, VOL, VCROWN, RLC
- Each of the genetic effects (mother, father, family, etc.) is extended to a 5x1 vector of effects incorporating *correlations* among effects across all traits, e.g.

i^{th} -mother will have an effect for each of the 5 traits as:

$$\mathit{mother}_i = \begin{bmatrix} m_{i,1} \\ m_{i,2} \\ m_{i,3} \\ m_{i,4} \\ m_{i,5} \end{bmatrix} \quad \text{Var}(\mathit{mother}_i) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \sigma_{14} & \sigma_{15} \\ & \sigma_2^2 & \sigma_{23} & \sigma_{24} & \sigma_{25} \\ & & \sigma_3^2 & \sigma_{34} & \sigma_{35} \\ & & & \sigma_4^2 & \sigma_{45} \\ & & & & \sigma_5^2 \end{bmatrix}$$

- All analysis done with ASReml v. 3 (code available upon request).

STATISTICAL ANALYSIS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

TYPE-A CORRELATIONS

Correlation between traits (pleitrophy).

- Property of genes of influencing more than one trait.
- It could be negative or positive (-1 to 1).
- Informs about the biological relationships among traits.
- Assists in the influence of one trait over the other by selecting 'good' individual.

$$rg_{A(P)} = \frac{\text{Cov}(P_1, P_2)}{\sqrt{\text{Var}(P_1) \times \text{Var}(P_2)}}$$

$$rg_{A(G)} = \frac{\text{Cov}(g_1, g_2)}{\sqrt{\text{Var}(g_1) \times \text{Var}(g_2)}}$$

STATISTICAL ANALYSIS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{Z}_1 \mathbf{b} + \mathbf{Z}_2 \mathbf{m} + \mathbf{Z}_3 \mathbf{f} + \mathbf{Z}_4 \mathbf{mf} + \mathbf{Z}_5 \mathbf{c} + \mathbf{e}$$

where,

\mathbf{y} is a matrix of measurements of size $n \times 5$,

$\boldsymbol{\mu}$ is the vector of overall means,

$\boldsymbol{\beta}$ is a matrix of fixed effects (overall mean and propagule type),

\mathbf{b} is a matrix of incomplete blocks within replicate random effects, $\sim N(0, \mathbf{D}_b)$,

\mathbf{m} is a matrix of mother random effects, $\sim N(0, \mathbf{A} \otimes \mathbf{V}_f)$,

\mathbf{f} is a matrix of father random effects, $\sim N(0, \mathbf{A} \otimes \mathbf{V}_f)$,

\mathbf{mf} is a vector of family random effects, $\sim N(0, \mathbf{I}_{mf} \otimes \mathbf{D}_{mf})$,

\mathbf{c} is a vector of clone within family random effects, $\sim N(0, \sigma^2_c \mathbf{D}_c)$,

\mathbf{e} is a vector of residual errors, $\sim N(0, \mathbf{I}_t \otimes \mathbf{V}_{e(p)})$, $p = \{s, c\}$

$\mathbf{D}_b, \mathbf{D}_{mf}, \mathbf{D}_c$ diagonal matrices allowing for a different variance component per trait for the block, family and clonal effects.

$\mathbf{V}_f, \mathbf{V}_{e(s)}, \mathbf{V}_{e(c)}$ unstructured (UN) matrices of size 5×5 that define the variance-covariance between traits for parental effects and residual errors.

RESULTS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

Trial	Comparison	df	Trait					Model	Parent Structure (a)	Family Structure (d)
			VOL	VCROWN	RLC	MDIAMB	MANG			
C	M1 vs M2	1	0.265	0.393	0.933	0.777	0.647			
	M1 vs M3	2	0.368	0.791	0.865	0.679	0.9			
	M1 vs M4	2	0.078	0.124	0.339	0.076	0.002			
	M3 vs M4	1	0.079	0.054	0.171	0.037	< 0.001			
N	M1 vs M2	1	0.92	0.655	0.495	0.841	0.752	M1	UN	UN
	M1 vs M3	2	0.634	0.803	0.9	0.914	0.235	M2	CS	UN
	M1 vs M4	2	< 0.001	< 0.001	0.004	0.018	< 0.001	M3	CS	CS
	M3 vs M4	1	< 0.001	< 0.001	< 0.001	0.005	< 0.001	M4	DIAG	DIAG

- According to AIC the best model is **M3**.
- The inclusion of a correlation among propagule effects, for both additive and dominance effects *is relevant* (as model M4 is shows a significant loss of information).
- Hence, additive and dominance effects can be estimated combining information from both types of propagules that results in an improved model (i.e. use M3).

RESULTS

1. COMPARISON OF GENETIC VALUES FOR PROPAGULE TYPES

Trial	Term	Trait (M3)				
		VOL	VCROWN	RLC	MDIAMB	MANG
C	Propagule Type (p-value)	< 0.001	0.944	1.000	< 0.001	< 0.001
	Additive Effect (ρ_a)	0.959	0.865	0.751	0.999	0.999
	Dominant Effect (ρ_d)	0.616	0.602	-0.169	0.999	0.999
N	Propagule Type (p-value)	< 0.001	< 0.001	0.596	< 0.001	< 0.001
	Additive Effect (ρ_a)	0.988	0.999	0.949	0.999	0.999
	Dominant Effect (ρ_d)	0.771	0.999	0.999	-0.838	0.999

- *P*-values from testing the difference among the means of propagule type indicate *significant differences* on both trials for VOL, MDIAMB and MANG, and no differences for RLC.
- Most estimated correlations are close to 1 with a positive association for both parental and family effects.

RESULTS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

GENETIC CORRELATIONS (G = A+D+I)

	Trial C					Trial N					
	MDIAMB	MANG	VOL	VCROWN	RLC	MDIAMB	MANG	VOL	VCROWN	RLC	
MDIAMB	0.257	0.173	0.443	0.465	0.304	0.147	0.020	0.400	0.427	0.322	MDIAMB
MANG		0.341	-0.172	-0.116	0.222		0.338	-0.375	-0.358	-0.098	MANG
VOL			0.340	0.962	0.413			0.372	0.971	0.403	VOL
VCROWN				0.312	0.564				0.335	0.603	VCROWN
RLC					0.178					0.271	RLC

$$H^2 = \frac{Vg}{Vt} = \frac{2\sigma_{mad}^2 + \sigma_{fami}^2 + \sigma_{clon}^2}{\sigma_{ib}^2 + 2\sigma_{mad}^2 + \sigma_{fami}^2 + \sigma_{clon}^2 + \sigma^2}$$

Diagonal – Broad-sense heritabilities (H^2)

Upper diagonal - Correlations

- High *genetic correlation* between VOL and VCROWN.
- Moderate correlations found among other relevant traits (e.g. VOL and MANG).
- Almost null correlation between MDIAMB and MANG.
- Moderate levels of broad-sense heritability.

RESULTS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

ADDITIVE CORRELATIONS (A)

	Trial C					Trial N					
	MDIAMB	MANG	VOL	VCROWN	RLC	MDIAMB	MANG	VOL	VCROWN	RLC	
MDIAMB	0.098	0.080	0.258	0.215	-0.181	0.115	-0.027	0.509	0.505	0.205	MDIAMB
MANG		0.202	-0.323	-0.234	0.199		0.296	-0.521	-0.474	-0.060	MANG
VOL			0.167	0.977	0.242			0.334	0.970	0.337	VOL
VCROWN				0.147	0.441				0.302	0.553	VCROWN
RLC					0.100					0.214	RLC

$$h_a^2 = \frac{Va}{Vt} = \frac{4\sigma_{mad}^2}{\sigma_{ib}^2 + 2\sigma_{mad}^2 + \sigma_{fami}^2 + \sigma_{clon}^2 + \sigma^2}$$

Diagonal – Narrow–sense heritabilities (h_a^2)

Upper diagonal - Correlations

- Tendencies are similar to total genetic correlations.
- Still high *genetic correlations* between VOL and VCROWN.
- Increased correlation in VOL and MANG.
- Interesting levels of heritability for MANG, VOL and VCROWN.

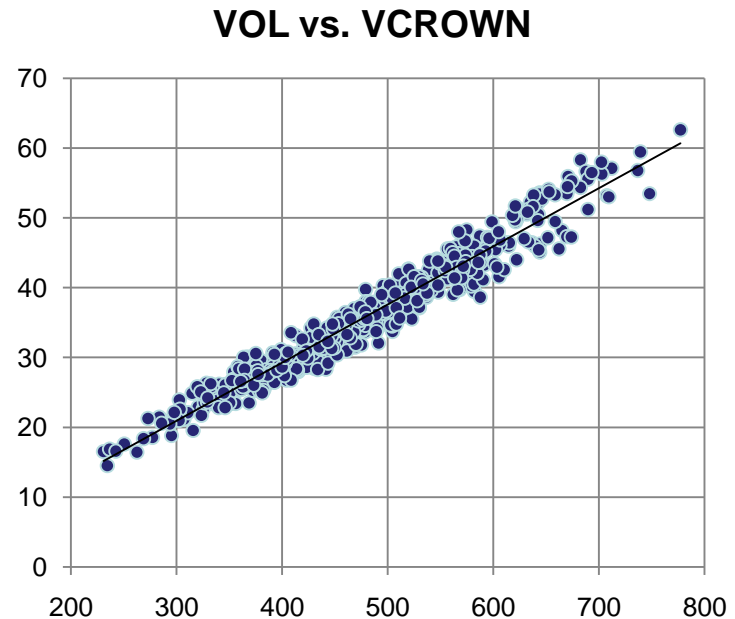
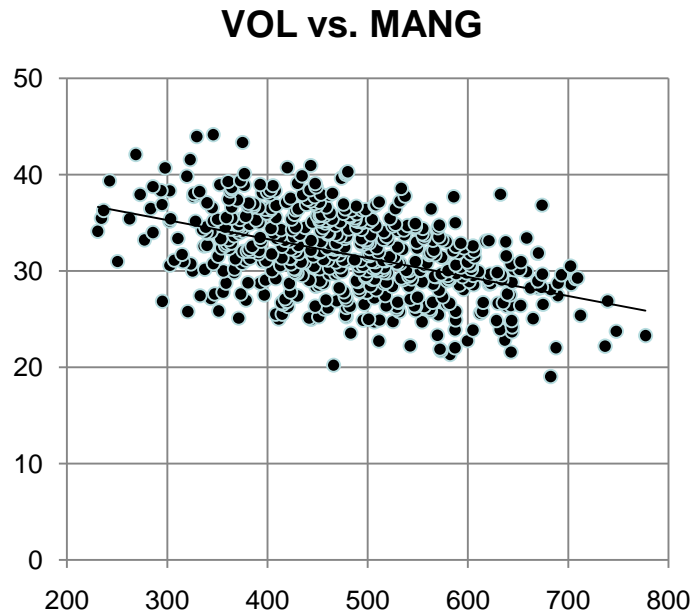
RESULTS

2. MULTIVARIATE ANALYSIS COMBINING PROPAGULE TYPES

FUTURE IMPLICATIONS OF MULTIVARIATE GENETIC ANALYSIS

Indirect Selection $\Delta G_{A1} = i_2 \times h_1 \times h_2 \times rg_{A(a)} \times \sigma_{P1}$

Search for correlation breakers



CONCLUSIONS

- Additive and dominant effects obtained from the different types of propagules are **not statistically different** (most correlations are high and positive).
- In addition, similar heritabilities are obtained from both propagule types.
- Information from both propagule types can be successfully combined to obtain a **unique ranking** of parent or crosses.
- There are some useful associations (i.e. genetic correlations) between different size and crown traits.
- A correlation of around 0.97 was found for both additive and genetic effects between crown volume and stem volume.
- **Increased precision** on variance components (and heritability estimates) was obtained from using a multivariate analysis rather than an univariate analysis.
- Interesting genotypes can be found by using **selection on multiple traits**.

AKNOWLEDGMENTS

- **Forest Biology Research Cooperative**
- **Cooperative Forest Genetics Research Project**
- **School of Forest Resources and Conservation, University of Florida**

