Supplementary Information

Contents

1. Supplementary notes

Note S1 Modeling the additive genetic effects in $\mathbf{u}_a$ under models FN_1, FN_2 and RN

Under models FN_1 and FN_2, the variance-covariance matrix for the additive genetic effects in $\mathbf{u}_a$ was defined as:

$$Var\begin{bmatrix} \mathbf{u}_{a_d} \\ \mathbf{u}_{a_{i(K_1)}} \\ \mathbf{u}_{a_{i(K_2)}} \\ . \\ . \\ . \\ . \\ . \\ \mathbf{u}_{a_{i(K_8)}} \end{bmatrix} = \begin{bmatrix} \sigma^2_{a_d} & \sigma_{a_{di(K_1)}} & \sigma_{a_{di(K_2)}} & \cdots\cdots & \sigma_{a_{di(K_8)}} \\ \sigma_{a_{di(K_1)}} & \sigma^2_{a_{i(K_1)}} & \sigma_{a_{i(K_1)i(K_2)}} & \cdots\cdots & \sigma_{a_{i(K_1)i(K_8)}} \\ \sigma_{a_{di(K_2)}} & \sigma_{a_{i(K_1)i(K_2)}} & \sigma^2_{a_{i(K_2)}} & \cdots\cdots & \sigma_{a_{i(K_2)i(K_8)}} \\ . & . & . & \cdots\cdots & . \\ . & . & . & \cdots\cdots & . \\ . & . & . & \cdots\cdots & . \\ . & . & . & \cdots\cdots & . \\ . & . & . & \cdots\cdots & . \\ \sigma_{a_{di(K_8)}} & \sigma_{a_{i(K_1)i(K_8)}} & \sigma_{a_{i(K_2)i(K_8)}} & \cdots\cdots & \sigma^2_{a_{i(K_8)}} \end{bmatrix} \otimes \mathbf{A} \qquad [S\_1]$$

where $\sigma^2_{a_d}$ is the direct genetic variance, $\sigma^2_{a_{i(K_1)}}$ to $\sigma^2_{a_{i(K_8)}}$ are indirect genetic variances pertaining to individual neighbor effects (i.e. for neighboring positions from K_1 to K_8; Figure 2) on a focal tree, $\sigma_{a_{di(K_1)}}$ to $\sigma_{a_{di(K_8)}}$ are covariances between the direct genetic effect and the indirect genetic effects of K_1 to K_8, and the remaining components (i.e. from $\sigma_{a_{i(K_1)i(K_2)}}$ to $\sigma_{a_{i(K_7)i(K_8)}}$) are covariances among the indirect genetic effects of K_1 to K_8; $\mathbf{A}$ is the matrix of additive genetic relationship coefficients among all individuals in the pedigree (i.e. base population parents, first-generation parents and their field-tested progeny); and $\otimes$ denotes the Kronecker product operation. Under model RN, the variance-covariance matrix for effects in $\mathbf{u}_a$ was similar in form to that provided in Equation [S_1], but for the reduced set of neighbors.

As shown in Equation [S_1], the unstructured (US) form of the variance-covariance matrix for $\mathbf{u}_a$ has 45 (co)variance parameters to be estimated but the sampling variance of each parameter is likely to result in a non-positive definite estimate of the matrix, because variances for some neighbor positions may be negligible and correlations among indirect effects should be 1. Therefore, as an alternative parameterization for the US matrix, we applied a structured variance model of the form used in factor analysis. The variance-covariance matrix for effects in $\mathbf{u}_a$ was thus defined as $(\mathbf{\Lambda\Lambda}' + \mathbf{\Psi}) \otimes \mathbf{A}$, where $\mathbf{\Lambda}$ is the v x p matrix of loadings (v = number of sub-vectors or terms in $\mathbf{u}_a$; p = number of fitted factors), $\mathbf{\Psi}$ is the v x v diagonal matrix of "specific" (lack of fit) variances, and the superscript ' refers to the transpose operation. This is known as a factor analytic (FA) structure (Mardia *et al*., 1988). It is more parsimonious than the US form, having fewer parameters, and can be constrained to be positive (semi-) definite.

To obtain the structure we required, we set p = 1 (i.e. a factor analytic structure with one factor; FA1) and the specific variances for the neighbor effects to zero. The variance related to a given term in $\mathbf{u}_a$ can then be obtained as $\lambda_j^2 + \Psi_j$ - where λ_j and Ψ_j are, respectively, the loading and the specific variance associated with the j^{th} term in $\mathbf{u}_a$ (j = 1 ... v) - with Ψ_j being zero for j = 2 ... v. In addition, a matrix with direct-indirect and indirect-indirect genetic correlations (i.e. referring to the covariances in Equation [S_1]) can be obtained from $\mathbf{\Lambda_{(c)}\Lambda'_{(c)}} + \mathbf{\Psi_{(c)}}$, where $\mathbf{\Lambda_{(c)}} = \mathbf{D}^{-1}\mathbf{\Lambda}$ and $\mathbf{\Psi_{(c)}} = \mathbf{D}^{-2}\mathbf{\Psi}$ are matrices of loadings and specific variances (respectively) on a correlation scale, and $\mathbf{D}$ is a v x v diagonal matrix of standard deviations related to the terms in $\mathbf{u}_a$.

The algorithm used for fitting the FA1 model in the software ASReml standardizes the factor to have a variance equal to one, and thus in our case the factor corresponds to a standardized indirect effect. The loadings then scale this effect depending on the neighbor position relative to the focal tree.

Note S2 Modeling the residual effects in ξ and η under models FN_1, FN_2 and RN

To explore the need to account for non-stationarity of the spatially correlated residuals in ξ (e.g. due to global environmental trend and extraneous variation; Gilmour *et al*., 1997), the mixed model in Equation (1) (for the description of this equation, see *Materials and Methods*) also incorporated: cubic smoothing splines (Verbyla *et al*., 1999) and linear covariates in the row and/or column directions for MLD and DBH; and an interaction between the row and column linear covariates for DBH. By naturally incorporating distance to neighbors, a separable first-order autoregressive process (AR1 x AR1) models (co)variances associated with non-heritable indirect effects affecting a focal tree and its neighbors. Stringer (2006) and Stringer *et al*. (2011) proposed an equal-roots third-order autoregressive process (EAR3) for jointly modeling both local environmental trend and competition effects at the residual level, since the AR1 process cannot model both of these sources of residual variation at the same time. Previous analyses of 2- and 4-year DBH data, in which IGEs were modeled as a combined effect (Costa e Silva *et al*., 2013), found that the magnitude of an autocorrelation parameter related to competition in the EAR3 model decreased considerably relative to its standard error when the two (co)variance parameters associated with IGEs were included in the mixed model. This suggested that the residual EAR3 model was picking up (co)variation due to genetic competition when heritable competition effects were not explicitly included in the model. A similar tendency was also observed in the present study for 8-year DBH data, with the improvement of the EAR3 over the AR1 becoming non-significant at the 5% level after including the two (co)variance parameters related to IGEs. Given these results, we have modeled spatial variation in the current study for DBH simply as an AR1 x AR1 structure. After the inclusion of a spatial correlation structure in the model defined in Equation (1), restricted maximum likelihood (REML) estimates of the variances associated with row and column effects within replicates became non-significant at the

5% level (based on likelihood-ratio tests) for the traits and ages examined, and thus these terms were dropped from the mixed model in all data analyses.

By regressing on the immediate neighbor, the AR1 process in a given direction implies an exponential decay of spatial correlation with distance of a focal tree and more distant neighbors. Under model FN_1, we explicitly modeled direct and indirect (co)variances pertaining to non-heritable effects that may cause interactions between a focal tree and its immediate neighbors, and distributed independently of the correlated residuals in $\boldsymbol{\xi}$. In this sense, for the independent residuals in $\boldsymbol{\eta}$, we specified $\boldsymbol{\eta} = (\boldsymbol{\eta}'_d, \boldsymbol{\eta}'_{i(K_1)}, \boldsymbol{\eta}'_{i(K_2)}, ..., \boldsymbol{\eta}'_{i(K_8)})'$ with the related incidence matrix given by $\mathbf{Z} = (\mathbf{Z}_d, \mathbf{Z}_{i(K_1)}, \mathbf{Z}_{i(K_2)}, ..., \mathbf{Z}_{i(K_8)})$, where the subscripts *d* and *i* denote direct and indirect effects, respectively, the subscripts K_1 to K_8 are defined as before, and the transpose operation is indicated by the superscript ' . $\mathbf{Z}_d$ relates the phenotype of a focal tree to its own direct independent residual, and $\mathbf{Z}_{i(K_1)}$ to $\mathbf{Z}_{i(K_8)}$ relate the phenotype of a focal tree to the indirect independent residual effect of each of its immediate neighbors. Under this specification of model terms, the variance-covariance matrix for effects in $\boldsymbol{\eta}$ used a matrix similar in form to that defined in Equation [S_1] for effects in $\mathbf{u}_a$, except that it applied the Kronecker product with $\mathbf{I}_{n_{obs}}$ (i.e. an identity matrix of order equal to the number of tree observations) rather than with the matrix **A**. This residual variance-covariance matrix was then parameterized through a factor analytic structure with one factor (FA1), and using the constraints defined as for the effects in $\mathbf{u}_a$ (i.e. a constrained FA1 matrix with 10 parameters, comprising: one specific variance and one loading for the direct effect pertaining to the focal tree, and eight loadings for the indirect effects of its neighbors K_1 to K_8; see *Materials and Methods*). Under models FN_2 and RN, the residuals in $\boldsymbol{\eta}$ were assumed to represent direct non-heritable effects only, and thus $Var(\boldsymbol{\eta}) = \sigma^2_{ie}\mathbf{I}_{n_{obs}}$, where σ^2_{ie} is the variance associated with an independent residual term (i.e. a 'nugget' effect) in the mixed model.

Note S3 Model comparison using the AIC and BIC information criteria

Information criteria may be used to compare and select models. In this context, the aim is to choose a model (or models) that minimizes an estimate of a criterion, which is a function of a measure of the goodness of fit and a penalty for model complexity. Two such criteria have been derived by Akaike (1974) and Schwarz (1978) and, in a broad sense, both attempt to achieve a trade-off between descriptive accuracy and model complexity. The Akaike's information criterion (AIC) is an estimate of the relative, expected Kullback-Leibler (K-L) distance of a model, which is an information-theoretic measure of the mean difference between the model and the unknowable truth (Burnham and Anderson, 2002; Kuha, 2004; Richards, 2005, 2008). For large samples, the Schwarz's Bayesian information criterion (BIC) may be a reasonable approximation of the natural logarithm of the Bayes factor, which is a measure of the evidence given by the data in favor of one model over another; thus, the difference between the BIC estimates of two models may provide a sensible indicator to quantify the strength of evidence for or against a model (Kass and Raftery, 1995; Kass and Wasserman, 1995).

Although based on different theoretical frameworks, both of the AIC and BIC criteria have the form of a penalized likelihood (by adding a function of the number of model parameters), and can be used to compare either nested or non-nested models (Burnham and Anderson, 2002; Kuha, 2004). For a given model, and regarding particularly mixed model selection (e.g. Wolfinger, 1993), we calculated the AIC as:

$$\mathrm{AIC} = -2ll_R + 2n_{vp} \qquad [S\_2]$$

and the Schwarz's Bayesian information criterion (BIC) as:

$$\mathrm{BIC} = -2ll_R + n_{vp}\log(n_{obs} - n_{fe}) \qquad [S\_3]$$

where ll_R is the model's REML log-likelihood, n_{vp} is the number of (co)variance parameters, and n_{obs} - n_{fe} is the residual degrees of freedom (n_{obs} = number of observations; n_{fe} = number of fixed effects). Sugiura (1978) suggested a finite-sample correction for the AIC criterion, which is recommended when $n_{obs}/n_{vp} < 40$ (Burnham and Anderson, 2002; page 445). However, reflecting the large sample sizes used in the analyses of our field experiment, the finite-sample correction suggested by Sugiura (1978) had a minor effect on the AIC values, and thus it is not considered in the AIC reported in our study.

A model with a lower AIC or BIC value has a greater level of empirical support than a model with higher values of either of these criteria (Burnham and Anderson, 2002; Kuha, 2004). Hence, when comparing a pair of models *M1* and *M2*, the larger the positive difference (Δ) for either $\Delta AIC_{(M1,M2)} = AIC_{M1} - AIC_{M2}$ or $\Delta BIC_{(M1,M2)} = BIC_{M1} - BIC_{M2}$ the less plausible it is that *M1* is a good approximating model to explain the variation in the observed data when compared with *M2*. Yet, as indicated by Burnham and Anderson (2002) for the particular case of comparing nested models by using the AIC, when $0 \leq \Delta AIC_{(M1,M2)} \leq 2$ there is a substantial level of empirical support for *M1*, in the sense that there will be little information loss by using this model relative to *M2* (which has a lower AIC value). This guideline incorporates the uncertainty associated with the fact that the AIC is an estimate, and thus its sampling error means that the model with a lower AIC value may not necessarily be better in terms of expected K-L distance (Richards, 2005, 2008). Therefore, under this simple rule, increasing evidence that a model with a lower AIC value is a better approximating model (given the data) may be provided when $\Delta AIC_{(M1,M2)} > 2$, which indicates a decreasing level (being essentially none when $\Delta AIC_{(M1,M2)} > 10$; Burnham and Anderson, 2002; page 70) of empirical support for the model with a higher AIC value. Parsimony is an important consideration when the number of fixed and random effects is large (Müller *et al.*, 2013) and, if the probable values of parameters added to a model are not well enough established by the available data, then a simpler model may be preferred over a complex model with more parameters (Kuha, 2004). This is likely to be more relevant for models sharing similar levels of

empirical support based on the AIC (e.g. when $0 \leq \Delta AIC_{(M1,M2)} \leq 2$) as, despite penalizing a model for added complexity, this criterion may still over-fit the data; on the other hand, the BIC tends to over-penalize model complexity, and thus it may favor models with fewer parameters than does the AIC (see below).

While comparing directly the log-likelihood (*LogL*) of pairs of nested models can be pursued through likelihood-ratio tests, model comparisons based on differences among either AIC or BIC values incorporate a penalty for model complexity, such that empirical support in favor of a given model relative to another will reflect a better balance between a good fit (as indicated by the *LogL*) and parsimony. The penalty for model complexity offsets partly the large-sample behavior of significance tests (such as likelihood-ratio tests), where simple models are increasingly likely to be rejected when the number of observations (n_{obs}) is large (as in our field experiment) (Kuha, 2004). Indeed, Burnham and Anderson (2002) indicated that, when compared with the AIC criterion, likelihood-ratio tests may tend to favor more complex models. In particular, Burnham and Anderson (2002; Table 6.20, page 338) also showed that, in cases where the difference between the null and alternative models in the number of fitted parameters is ≥ 8, the likelihood-ratio testing method resulted in increasingly strong support of the models with many parameters and strong rejection of the simple null model. In this sense, for our model comparisons FN_2/FN_1 and RN/FN_2 that involve many parameters (for example, 9 and 6 parameters in the FN_2/FN_1 and RN/FN_2 comparisons, respectively; see footnote *d* of Table 1), likelihood-ratio tests may be too liberal. Thus, rather than applying likelihood-ratio tests, we used information criteria as a more conservative approach to compare models FN_1, FN_2 and RN, as described in the *Materials and Methods* section.

By taking n_{obs} into account in its calculation, BIC uses a more conservative penalty for model complexity than does AIC, implying that a greater improvement in *LogL* may be required to offset the increased number of parameters added to a model, and thus to enhance the relative performance of more complex models. Thus, due to the larger weight in the penalty for model

complexity when $n_{obs} > e^2$ (as ascertained by comparing the Equations [S_2] and [S_3] presented above), BIC will tend to favor simpler models (i.e. with fewer parameters) than does the AIC. The increasing function of n_{obs} in the penalty term of BIC is needed for asymptotically consistency (Burnham and Anderson, 2002). BIC is a consistent model selector, in the sense that the probability of identifying the true data-generating model (assuming that it exists and is included within the candidate set) approaches unity as sample size increases; however, consistent criteria such as BIC may be particularly useful if, within a fixed family of models, there actually exists a simple (low-dimensional) target "true" model, given the data (Burnham and Anderson, 2002; Kuha, 2004; Vrieze, 2012).

Unlike the case of BIC, the penalty term in the AIC is not an increasing function of n_{obs}; thus, AIC is not a consistent model selector, as it has always some probability of selecting overly parameterized models (Kuha, 2004; Richards, 2005, 2008; Vrieze, 2012). As also mentioned by Forster and Sober (1994), a feature of the Akaike's theorem is that the weight placed on model complexity relative to the *LogL* term declines with increasing sample size because, with a large amount of data, the estimate of how close a model is to the truth will be determined mainly by goodness-of-fit. There are studies indicating that the AIC may favor overly complex models (Link and Barker, 2006), even asymptotically (Kass and Raftery, 1995). However, when the dimensionality of the model that is closest to reality increases with n_{obs}, the AIC is expected to be an asymptotically efficient model selector, such that it may select models with a better predictive performance (e.g. with a lower mean squared prediction error) than does BIC as sample size increases (Kuha, 2004; Vrieze, 2012). If it is assumed that the processes generating the observed data are conceivably best characterized by a model with a complex dimensionality (i.e. being infinite or increasing with sample size), then the AIC may be preferred to BIC for model selection; in such a condition, the assumptions forming the basis of the BIC properties may not hold (e.g. it is unlikely that the "true" model is included in the candidate set, in which case BIC

may not be a consistent model selector), and thus AIC may outperform BIC in terms of enabling to choose a model that best approximates reality (Burnham and Anderson, 2002; Vrieze, 2012).

As indicated by simulation studies, it is possible that there will not be always a single criterion having the best performance in terms of model selection, as this may depend on the nature of the "true" data-generating model, the given set of candidate models being compared, the number of observations, and the theoretical foundations of the criteria (Burnham and Anderson, 2002; Kuha, 2004; Vrieze, 2012; Müller *et al*., 2013). Hence, considering more than one information criterion for model comparison has been advocated to provide a useful guidance for model selection, while also enabling to identify models that are favored by different criteria (Kuha, 2004; Müller *et al*., 2013). Although the AIC and BIC are based on distinct motivations (e.g. model predictive performance expected for new data, in the case of AIC; identification of the "true" data-generating model, in the case of BIC), as well as on different assumptions underlying their properties (e.g. asymptotic efficiency for AIC; asymptotic consistency for BIC), they are well-founded information criteria for model comparison and, in a broad sense, both aim at identifying good approximating models that are able to provide an adequate description of the observed data while minimizing the number of fitted parameters (Burnham and Anderson, 2002; Kuha, 2004). Besides the differences between the AIC and BIC in the theoretical foundations for measuring optimal performance in model comparison, it is also likely that model selection based on either of these criteria will not capture the full complexity inherent to the representation of reality. Neither the AIC or BIC may be able to effectively minimize all the measures of divergence aiming to represent how close a model is to the truth; however, the choice of these loss functions can determine the performance of the information criteria, and thus ideally the loss function(s) that is (are) relevant in a given context should be identified to decide on whether the AIC or BIC would be more adequate (Vrieze, 2012). Yet, useful information for model selection may be obtained by using the AIC and BIC together (as we have done in our study), particularly in terms of trying to find models favored by both criteria (Kuha, 2004). In this sense, when the AIC and BIC provide

similar indications in regard to a preferred model, there will be a stronger support on the strength of the choice for a good approximating model, given the observed data.

Note S4 Comparing the RN and FN_2 models

The FA1 loadings for the same-row neighbors K_1 and K_2 were dominant in the FN_2 model when assessed relative to their standard errors (see Table S1 in *Supplementary tables*), and so were included in the initial RN model; dropping either one of these neighbors led in general to a substantially worse fit of the RN model.

When applying the AIC to compare models, the initial K_1 and K_2 positions were accepted as the main influential neighbors if their inclusion in the RN model resulted in $\Delta AIC_{(RN,FN\_2)} \leq 2$; otherwise, a forward selection process, adding other neighbor positions in order of relative magnitude of the FA1 loadings (i.e. according to the information given in Table S1), was pursued until $\Delta AIC_{(RN,FN\_2)} \leq 2$. The use of $\Delta AIC_{(RN,FN\_2)} \leq 2$ as a stopping rule was a conservative approach intended to minimize the chance of retaining too many parameters (and thus achieving a more parsimonious model) in the FA1 structure as a result of including successively more neighbor positions in the RN model, and considering that the AIC may favor overly complex models (see *Note S3* above). In this sense, the AIC will indicate the largest acceptable RN model given the observed data.

When model comparison was based on the BIC, the initial K_1 and K_2 positions were accepted as the main influential neighbors if adding the next most important neighbor to the RN model did not lead to an improvement in this criterion; otherwise, such a sequential inclusion of neighbor positions in the RN model was followed until no further improvement was observed in the BIC (i.e. when $\Delta BIC_{(RN,FN\_2)}$ started to become less negative). Thus, considering that the BIC tends to over-penalize model complexity, the smallest acceptable RN model given the observed data will be identified by BIC.

Note S5 Modeling a combined indirect effect under univariate data analysis

When modeling IGEs as a combined effect, we specified $\mathbf{u}_a = (\mathbf{u}'_{a_d}, \mathbf{u}'_{a_i})'$ with the related incidence matrix given by $\mathbf{Z}_a = (\mathbf{Z}_{a_d}, \mathbf{Z}_{a_i})$, where the subscripts *d* and *i* are defined as before. The matrix $\mathbf{Z}_{a_d}$ relates the phenotype of a focal tree to its own direct additive genetic effect (i.e. akin to the models FN_1, FN_2 and RN). The matrix $\mathbf{Z}_{a_i}$ was constructed by summing the matrices $\mathbf{Z}_{a_{i(K1)}}$ to $\mathbf{Z}_{a_{i(K8)}}$ pertaining to the immediate neighbors of a focal tree, and containing intensity of interaction factors calculated specifically for each neighborhood. These are weighting factors that account for the differential intensity of interaction effects that neighbors may exert on the phenotype of a focal tree, as a result of missing trees in the neighborhood (e.g. due to mortality and/or edge position of focal trees) and/or differences in inter-tree distance (Cappa and Cantet, 2008; Costa e Silva and Kerr, 2013). Therefore, the two approaches - that is, modeling IGEs using a FA1 model and as a combined effect - differ in that, in the former, the relative magnitude of the indirect effect with respect to a particular neighbor position is estimated in the model whereas, in the latter, it is a function of the relative distances among trees but allowing for missing individuals.

Cappa and Cantet (2008) developed formulae to calculate intensity of interaction factors by assuming inter-row spacing to be equal to inter-column spacing. However, these formulae cannot be used for the *Eucalyptus globulus* trial we have studied here due to the different row and column spacing (i.e. 2.125 and 5.0 m within and between planting rows, respectively). Consequently, Costa e Silva and Kerr (2013; see that article's Supplementary Material) developed formulae for a general application (i.e. inter-row spacing not necessarily equal to inter-column spacing), and thus enabling intensity of interaction factors to be calculated for same-row, same-column and diagonal neighbors in our studied trial (see also Costa e Silva *et al*., 2013).

The derivation of the intensity of interaction factors enables uniformity of scaling of indirect effects within and across neighborhoods. Without the scaling of indirect genetic effects using the intensity of interaction factors, the total contribution (i.e. summed across all immediate neighbors of a focal tree) of the indirect additive genetic variance ($\sigma^2_{a_i}$) to the phenotypic variance would equal $n_{ngh}\sigma^2_{a_i}$ (n_{ngh} = number of immediate neighbors of a focal tree), assuming no inbreeding and genetically unrelated neighbors. However, through the scaling achieved by using the intensity of interaction factors, $\sigma^2_{a_i}$ corresponds directly to the total contribution of indirect genetic effects to the phenotypic variance, assuming again absence of inbreeding and unrelated neighbors (for further details, see Supporting Information in Costa e Silva et *al*., 2013).

The genetic (co)variance parameters estimated by modeling IGEs as a combined effect were compared between using all eight immediate neighbors and a reduced set comprising the most influential neighbors as identified under model RN, using recalculated intensity of interaction factors. The variance-covariance matrix specified for the additive genetic effects in $\mathbf{u}_a$ had the form:

$$Var\begin{bmatrix}\mathbf{u}_{a_d}\\ \mathbf{u}_{a_i}\end{bmatrix} = \begin{bmatrix}\sigma^2_{a_d} & \sigma_{a_{di}}\\ \sigma_{a_{di}} & \sigma^2_{a_i}\end{bmatrix} \otimes \mathbf{A} \qquad [S\_4]$$

where $\sigma^2_{a_i}$ corresponds to the total contribution of the indirect genetic variance to the phenotypic variance, $\sigma_{a_{di}}$ is the direct-indirect genetic covariance, and $\sigma^2_{a_d}$ and $\mathbf{A}$ are defined as before (i.e. as in Equation [S_1]).

A variance-covariance matrix explicitly incorporating direct and indirect (co)variances pertaining to residual effects in $\boldsymbol{\eta}$ (i.e. akin to model FN_1, but using: a 2 x 2 matrix similar in form to that given in Equation [S_4]; and intensity of interaction factors in a incidence matrix constructed as $\mathbf{Z}_{a_i}$, but relating the phenotype of a focal tree to the indirect non-heritable effects

of the immediate neighbors) was not applied in a final model. This decision was made on the basis of the results obtained in preliminary analyses modeling a combined indirect effect for terms in $\mathbf{u}_a$ and $\boldsymbol{\eta}$: for the traits and ages examined, REML estimates for the (co)variance parameters related to indirect non-heritable effects were never significant. Table S3 in *Supplementary tables* presents these results, which were obtained from analyses using the neighbors identified as contributing most to IGEs, and the neighbors appearing to be dominant at the residual level (i.e. based on the magnitude of the FA1 loadings relative to their standard errors, as estimated under the FN_1 model; see Table S2). In addition, these analyses indicated positive and high estimates for the autocorrelation parameters ϕ_{row} and ϕ_{col} associated with local environmental trend. All of these results suggested that, at the population level, indirect heritable effects and local environmental trend seemed to be more important than indirect non-heritable (residual) effects for the given data. Thus, a final model used the following definition of the variance-covariance structure for the residual effects in **e** (which was similar to that specified for models FN_2 and RN):

$$Var(\mathbf{e}) = Var(\boldsymbol{\xi} + \boldsymbol{\eta}) = \sigma^2_{ce} \left[\boldsymbol{\Sigma}_{row}(\phi_{row}) \otimes \boldsymbol{\Sigma}_{col}(\phi_{col})\right] + \sigma^2_{ie}\mathbf{I}_{n_{obs}} \qquad [S\_5]$$

where σ^2_{ie} and $\mathbf{I}_{n_{obs}}$ are defined as in *Note S2*, and the remaining parameters as in *Materials and Methods*.

Note S6 Variance-covariance matrices for random terms defined under the bivariate model

Under the bivariate linear mixed model defined in Equation (2), the variance-covariance matrix for the additive genetic effects was specified as:

$$Var\begin{bmatrix}\mathbf{u}_{d_1}\\ \mathbf{u}_{i_1}\\ \mathbf{u}_{d_2}\\ \mathbf{u}_{i_2}\end{bmatrix}=\begin{bmatrix}\sigma^2_{a_{d_1}} & \sigma_{a_{d_1},a_{i_1}} & \sigma_{a_{d_1},a_{d_2}} & \sigma_{a_{d_1},a_{i_2}}\\ \sigma_{a_{d_1},a_{i_1}} & \sigma^2_{a_{i_1}} & \sigma_{a_{d_2},a_{i_1}} & \sigma_{a_{i_1},a_{i_2}}\\ \sigma_{a_{d_1},a_{d_2}} & \sigma_{a_{d_2},a_{i_1}} & \sigma^2_{a_{d_2}} & \sigma_{a_{d_2},a_{i_2}}\\ \sigma_{a_{d_1},a_{i_2}} & \sigma_{a_{i_1},a_{i_2}} & \sigma_{a_{d_2},a_{i_2}} & \sigma^2_{a_{i_2}}\end{bmatrix}\otimes\mathbf{A} \qquad [S\_6]$$

where, for 1 and 2 representing a pair of traits or ages, $\sigma_{a_{d_1},a_{d_2}}$ is the covariance between the direct genetic effects of 1 and 2, $\sigma_{a_{d_1},a_{i_2}}$ is the covariance between the direct genetic effect of 1 and the indirect genetic effect of 2, $\sigma_{a_{d_2},a_{i_1}}$ is the covariance between the direct genetic effect of 2 and the indirect genetic effect of 1, and $\sigma_{a_{i_1},a_{i_2}}$ is the covariance between the indirect genetic effects of 1 and 2; for a given trait or age, $\sigma^2_{a_{d_1}}$ and $\sigma^2_{a_{d_2}}$ are direct genetic variances, $\sigma^2_{a_{i_1}}$ and $\sigma^2_{a_{i_2}}$ are indirect genetic variances, and $\sigma_{a_{d_1},a_{i_1}}$ and $\sigma_{a_{d_2},a_{i_2}}$ are direct-indirect genetic covariances. For the l^{th} term in $\sum_{l=1}^{t}\mathbf{Z}_l\mathbf{u}_l$, the variance-covariance matrix was defined as:

$$Var\begin{bmatrix}\mathbf{u}_{l_1}\\ \mathbf{u}_{l_2}\end{bmatrix}=\begin{bmatrix}\sigma^2_{l_1} & \sigma_{l_1,l_2}\\ \sigma_{l_1,l_2} & \sigma^2_{l_2}\end{bmatrix}\otimes\mathbf{I}_{n_l} \qquad [S\_7]$$

where $\sigma^2_{l_1}$, $\sigma^2_{l_2}$ and σ_{l_1,l_2} are (co)variances for effects of the l^{th} term, and $\mathbf{I}_{n_l}$ is an identity matrix of order n_l (i.e. the number of effects of the l^{th} term). For a spline term, σ_{l_1,l_2} was always assumed to be zero. For each of the other terms in $\sum_{l=1}^{t}\mathbf{Z}_l\mathbf{u}_l$, the σ_{l_1,l_2} component was fitted only

when the REML estimates of $\sigma^2_{l_1}$ and $\sigma^2_{l_2}$ were both found to be significant at the 5% level in the analyses of single traits or ages, otherwise it was assumed to be zero. For two traits or ages using a spatial correlation structure of the same form, the covariance between effects in $\boldsymbol{\xi}$ can be accounted for by using the following variance-covariance matrix (Costa e Silva and Graudal, 2008):

$$Var\begin{bmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{bmatrix} = \boldsymbol{\Sigma}_{row}(\phi_{row}) \otimes \boldsymbol{\Sigma}_{col}(\phi_{col}) \otimes \begin{bmatrix}\sigma^2_{ce_1} & \sigma_{ce_1,ce_2}\\ \sigma_{ce_1,ce_2} & \sigma^2_{ce_2}\end{bmatrix} \qquad [S\_8]$$

where $\sigma^2_{ce_1}$, $\sigma^2_{ce_2}$ and σ_{ce_1,ce_2} are (co)variances for the spatially correlated residuals; $\boldsymbol{\Sigma}_{row}$, ϕ_{row}, $\boldsymbol{\Sigma}_{col}$ and ϕ_{col} are defined as before. The specification of the variance-covariance matrix in Equation [S_8] assumes common autocorrelation parameters (ϕ_{row} and ϕ_{col}) for 1 and 2 in a given direction. For the independent residuals in $\boldsymbol{\eta}$, the variance-covariance matrix was a 2 x 2 matrix similar in form to that given in Equation [S_7], but with the order of the identity matrix in the Kronecker product (⊗) being equal to the number of tree observations (i.e. $\mathbf{I}_{n_{obs}}$).

Note S7 Correlation between the total breeding values of two traits or ages

The correlation between the total breeding values ($r_{TBV_{(1,2)}}$) of two traits or ages was computed as:

$$r_{TBV_{(1,2)}} = \frac{Cov(TBV_1, TBV_2)}{\sqrt{Var(TBV_1)\,Var(TBV_2)}} \qquad [S\_9]$$

where *TBV* denotes total breeding value, $Var(TBV_1)$ and $Var(TBV_2)$ are total heritable variances for traits or ages 1 and 2, respectively, and $Cov(TBV_1, TBV_2)$ is the covariance between the total breeding values of traits or ages 1 and 2. The *TBV* is the heritable impact of an individual's own

genes on the population's mean trait value (Bijma *et al*., 2007), and the variance among individuals in *TBV* (i.e. *Var*(*TBV*)) determines the population's potential to respond to selection for a trait affected by IGEs at a given age (Bijma, 2011). *Var*(*TBV*) was calculated by Equation (16) given by Costa e Silva and Kerr (2013) and, following Equation (14) also provided by these authors to estimate *TBV*, *Cov*(TBV_1, TBV_2) in Equation [S_9] was computed as:

$$\sigma_{a_{d_1},a_{d_2}} + (\overline{n_{row_2}}\,\overline{f_{row_2}} + \overline{n_{col_2}}\,\overline{f_{col_2}} + \overline{n_{diag_2}}\,\overline{f_{diag_2}})\sigma_{a_{d_1},a_{i_2}} + \qquad [S\_10]$$

$$(\overline{n_{row_1}}\,\overline{f_{row_1}} + \overline{n_{col_1}}\,\overline{f_{col_1}} + \overline{n_{diag_1}}\,\overline{f_{diag_1}})\sigma_{a_{d_2},a_{i_1}} +$$

$$(\overline{n_{row_1}}\,\overline{f_{row_1}} + \overline{n_{col_1}}\,\overline{f_{col_1}} + \overline{n_{diag_1}}\,\overline{f_{diag_1}})(\overline{n_{row_2}}\,\overline{f_{row_2}} + \overline{n_{col_2}}\,\overline{f_{col_2}} + \overline{n_{diag_2}}\,\overline{f_{diag_2}})\sigma_{a_{i_1},a_{i_2}}$$

where, for traits or ages 1 and 2, $\overline{n_{row}}\,\overline{f_{row}}$, $\overline{n_{col}}\,\overline{f_{col}}$ and $\overline{n_{diag}}\,\overline{f_{diag}}$ denote products of means, taken across all focal trees at a given trial, for the number of their immediate neighbors (n_{row}, n_{col} and n_{diag}) and corresponding intensity of interaction factors (f_{row}, f_{col} and f_{diag}) in the row, column and diagonal directions of the trial layout, respectively; $\sigma_{a_{d_1},a_{d_2}}$, $\sigma_{a_{d_1},a_{i_2}}$, $\sigma_{a_{d_2},a_{i_1}}$ and $\sigma_{a_{i_1},a_{i_2}}$ are defined as in Equation [S_6] (see *Note S6*). Equation [S_10] refers explicitly to modeling IGEs as a combined effect using all eight immediate neighbors of a focal tree. For an analysis comprising a reduced neighborhood set, Equation [S_10] will use only the most influential neighbors (i.e. as identified under model RN) with the corresponding recalculated intensity of interaction factors.

2. Supplementary tables

Table S1 Estimates of the FA1 loadings obtained under the FN_2 model for the indirect effects at the heritable level. [a, b, c, d, e]

Neighbors	MLD (2 years)	DBH (2 years)	DBH (4 years)	DBH (8 years)
K_1	1.093	1.816	5.404	8.287
	(4.35)	(6.17)	(12.65)	(13.51)
K_2	0.668	2.390	4.794	10.937
	(2.75)	(7.50)	(10.95)	(15.51)
K_3	0.194	0.079	0.762	1.175
	(0.96)	(0.32)	(2.15)	(1.90)
K_4	0.190	0.764	1.851	2.549
	(0.77)	(2.55)	(4.50)	(4.12)
K_5	0.086	0.452	0.274	0.036
	(0.42)	(1.49)	(0.65)	(0.06)
K_6	**0.297**	0.502	0.955	1.937
	(1.24)	(1.85)	(2.47)	(3.21)
K_7	0.471	**0.107**	0.122	1.253
	(2.10)	**(0.31)**	(0.28)	(1.94)
K_8	0.492	**0.153**	0.627	**0.148**
	(1.93)	**(0.59)**	(1.63)	**(0.24)**

[a] The indirect effects at the heritable (i.e. additive genetic) level pertain to each of the eight immediate neighbors of a focal tree (i.e. K_1 to K_8).

[b] For a given neighbor, the ratio of the magnitude of the loading relative to its standard error is presented in parenthesis. The only values that were consistently high relate to neighbors K_1 and K_2.

[c] The loadings and their estimate/standard error ratios are presented in absolute value. For a given trait and age, the loadings marked in bold have an opposite sign compared to the loadings obtained for the other neighbors.

[d] We used a FA structure that fixed the specific variances associated with the individual neighbor effects to be zero, so that the correlations between the direct genetic effect and the indirect genetic effects of K_1 to K_8 are the same except for sign, and the correlations among the indirect genetic effects of K_1 to K_8 are all one or minus one. For MLD, the correlations between the direct effect and the indirect effects were all positive, except for the correlation involving the relationship with the neighbor in which the loading is marked in bold. For DBH at a given age, the correlations between the direct effect and the indirect effects were all negative, except for the correlation estimate(s) involving the relationship(s) with the neighbor(s) in which the loading is marked in bold.

[e] For MLD, the parameter estimates refer to the (arcsine) transformed and rescaled observations (see footnote *e* of Table 1).

Table S2 Estimates of the FA1 loadings obtained under the FN_1 model for the indirect effects at the non-heritable level. [a, b, c, d, e]

Neighbors	MLD (2 years)	DBH (2 years)	DBH (4 years)	DBH (8 years)
K_1	0.813	0.525	**1.168**	**3.649**
	(0.82)	(0.63)	**(1.04)**	**(1.97)**
K_2	1.357	0.767	2.765	4.659
	(1.22)	(0.96)	(2.30)	(2.85)
K_3	0.040	0.405	**2.606**	**3.608**
	(0.06)	(0.48)	**(3.15)**	**(2.54)**
K_4	0.047	**0.155**	2.966	3.785
	(0.07)	**(0.19)**	(3.45)	(2.78)
K_5	0.218	1.266	0.132	**0.456**
	(0.43)	(1.66)	(0.18)	**(0.40)**
K_6	1.405	0.193	1.998	3.511
	(2.26)	(0.31)	(3.09)	(3.40)
K_7	0.141	**1.384**	**0.321**	1.160
	(0.21)	**(1.81)**	**(0.46)**	(1.07)
K_8	1.309	0.037	**0.822**	**1.191**
	(2.11)	(0.06)	**(0.98)**	**(0.90)**

[a] The indirect effects at the non-heritable (i.e. independent residual) level pertain to each of the eight immediate neighbors of a focal tree (i.e. K_1 to K_8).

[b] For a given neighbor, the ratio of the magnitude of the loading relative to its standard error is presented in parenthesis.

[c] The loadings and their estimate/standard error ratios are presented in absolute value. For a given trait and age, the loadings marked in bold have an opposite sign compared to the loadings obtained for the other neighbors.

[d] We used a FA structure that fixed the specific variances associated with the individual neighbor effects to be zero, so that the correlations between the direct genetic effect and the indirect genetic effects of K_1 to K_8 are the same except for sign, and the correlations among the indirect genetic effects of K_1 to K_8 are all one or minus one. For MLD, the correlations between the direct effect and the indirect effects were all positive. For DBH at a given age, the correlations between the direct effect and the indirect effects were all negative, except for the correlation estimate(s) involving the relationship(s) with the neighbor(s) in which the loading is marked in bold.

[e] For MLD, the parameter estimates refer to the (arcsine) transformed and rescaled observations (see footnote *e* of Table 1).

Table S3 REML estimates for (co)variance parameters (± standard errors) related to direct and indirect effects at the heritable (i.e. additive genetic) and non-heritable (i.e. independent residual) levels, obtained by modeling indirect effects as a combined effect, and using reduced sets of immediate neighbors. [a, b, c, d]

Trait (age)	$\hat{\sigma}^2_{a_d}$	$\hat{\sigma}_{a_{di}}$	$\hat{\sigma}^2_{a_i}$	$\hat{r}_{a_{di}}$	$\hat{\sigma}^2_{\eta_d}$	$\hat{\sigma}_{\eta_{di}}$	$\hat{\sigma}^2_{\eta_i}$	ΔLogL
MLD (2 years)	82.96 ± 14.45 (5.7)	9.26 ± 1.95 (4.7)	1.49 ± 0.62 (2.4)	0.83 ± 0.13 (6.4)	99.23 ± 8.0 (12.4)	-0.27 ± 1.12 (-0.2)	-1.46 ± 3.17 (-0.5)	0.13
DBH (2 years)	90.02 ± 15.28 (5.9)	-22.01 ± 2.92 (-7.5)	6.72 ± 1.48 (4.5)	-0.90 ± 0.06 (-15.0)	135.81 ± 9.25 (14.7)	0.85 ± 1.52 (0.6)	-4.0 ± 4.17 (-1.0)	0.66
DBH (4 years)	219.24 ± 29.47 (7.4)	-84.35 ± 7.98 (-10.6)	36.25 ± 5.25 (6.9)	-0.95 ± 0.03 (-31.7)	251.50 ± 19.96 (12.6)	-1.04 ± 3.71 (-0.3)	-0.42 ± 8.40 (-0.1)	0.04
DBH (8 years)	644.99 ± 81.54 (7.9)	-298.86 ± 28.85 (-10.4)	153.59 ± 19.74 (7.8)	-0.95 ± 0.03 (-31.7)	646.68 ± 42.24 (15.3)	6.30 ± 21.72 (0.3)	-41.23 ± 31.23 (-1.3)	1.29

[a] The tabulated (co)variance parameter estimates refer to: $\hat{\sigma}^2_{a_d}$ = direct additive genetic variance; $\hat{\sigma}_{a_{di}}$ = direct-indirect additive genetic covariance; $\hat{\sigma}^2_{a_i}$ = indirect additive genetic variance; $\hat{r}_{a_{di}}$ = direct-indirect additive genetic correlation; $\hat{\sigma}^2_{\eta_d}$ = direct residual variance; $\hat{\sigma}_{\eta_{di}}$ = direct-indirect residual covariance; $\hat{\sigma}^2_{\eta_i}$ = indirect residual variance. The ratio of the parameter estimate relative to its standard error is given in parenthesis. For MLD, the parameter estimates refer to the (arcsine) transformed and rescaled observations (see footnote *e* of Table 1).

[b] Modeling indirect effects as a combined effect used: for genetic effects, the neighbor positions identified as contributing most to IGEs (i.e. see "Neighbors kept in RN" in Table 1); and for residual effects, the neighbors appearing to be dominant at the residual level (i.e. based on the magnitude of the FA1 loadings relative to their standard errors, as estimated under the FN_1 model; see Table S2). In the latter case, the neighbor positions considered in the reduced set of immediate neighbors of a focal tree were: K_6 and K_8 for MLD; K_5 and K_7 for DBH at age 2 years; K_3, K_4 and K_6 for DBH at age 4 years; and K_2, K_3, K_4 and K_6 for DBH at age 8 years.

[c] ΔLogL = $ll_{R(M1)} - ll_{R(M2)}$, where ll_R is the model's REML log-likelihood, *M1* is the model considering indirect effects at both heritable and non-heritable levels (i.e. as presented in the table above), and *M2* is the model considering indirect effects at the heritable level only (i.e. see the columns pertaining to "Reduced set of immediate neighbors" of Table 2). As indicated by the magnitude of ΔLog, a two-tailed likelihood-ratio test of the overall significance of the indirect effects at the non-heritable level (and involving a joint test of significance for $\hat{\sigma}_{\eta_{di}}$ and $\hat{\sigma}^2_{\eta_i}$) leads to non-significant results ($P > 0.05$; 2 df) for all traits and ages.

[d] The $\hat{\sigma}^2_{\eta_i}$ estimate was allowed to be negative. Negative estimates of variance components may occur when the inherent variability is low and/or there is an expected negative intraclass correlation (Oliveira *et al*., 2016). A negative intraclass correlation may reflect the presence of competition among group members for a fixed resource, with the resulting within-unit (or within-group) negative correlation being captured by a negative variance component estimate (Kenny *et al*,. 2002; Pryseley *et al*,. 2011; Oliveira *et al*., 2016). In these circumstances, constraining an estimate of a variance component to be positive may result in misleading conclusions (Kenny *et al*., 2002; Molenberghs and Verbeke, 2011; Pryseley *et al*., 2011; Oliveira *et al*., 2016). For DBH at a given age, modeling the variance-covariance matrix for the independent residuals in $\boldsymbol{\eta}$ resulted in estimates of FA1 loadings that had opposite signs for the neighbor positions used in the reduced set (see Table S2); this may be translated as negative correlations among the indirect non-heritable effects themselves, which could have been captured in the negative $\hat{\sigma}^2_{\eta_i}$ estimates obtained for DBH. This dissimilarity among indirect non-heritable effects was not apparent for MLD (i.e. no negative dependence seemed to occur for the neighbor positions used in the reduced set, as the corresponding FA1 loadings did not have opposite signs; see Table S2), and thus the negative $\hat{\sigma}^2_{\eta_i}$ estimates obtained for MLD could be mainly attributed to a low inherent variability for indirect non-heritable effects in the studied trial.

Table S4 Estimates over time for the total heritable variance ($\widehat{Var}(TBV)$), as well as for the contributions to $\widehat{Var}(TBV)$ due to direct and indirect genetic effects, based on additive genetic (co)variance components estimated from a univariate model in which IGEs were modeled as a combined effect, using either full (FS) or reduced (RS) sets of immediate neighbors of a focal tree. Absolute and relative (i.e. Estimate/ $\hat{\sigma}_p^2$) values of the parameter estimates are presented, and their respective standard errors are provided in parenthesis. [a, b, c, d]

		MLD (2 years)		DBH (2 years)		DBH (4 years)		DBH (8 years)	
		Estimate	Estimate/ $\hat{\sigma}_p^2$	Estimate	Estimate/ $\hat{\sigma}_p^2$	Estimate	Estimate/ $\hat{\sigma}_p^2$	Estimate	Estimate/ $\hat{\sigma}_p^2$
Total heritable variance ($\widehat{Var}(TBV)$)									
	FS	141.07	0.748	21.12	0.086	18.34	0.035	38.07	0.026
		(± 22.51)	(± 0.094)	(± 9.52)	(± 0.037)	(± 14.77)	(± 0.028)	(± 31.87)	(± 0.022)
	RS	115.58	0.608	41.45	0.168	40.54	0.077	80.98	0.056
		(± 18.43)	(± 0.074)	(± 11.27)	(± 0.042)	(± 15.07)	(± 0.027)	(± 32.11)	(± 0.021)
Contribution to $\widehat{Var}(TBV)$ *due to direct genetic effects*									
$\hat{\sigma}_{a_d}^2$	FS	80.17	0.425	91.27	0.373	218.12	0.415	708.05	0.483
		(± 14.0)	(± 0.058)	(± 16.18)	(± 0.056)	(± 29.60)	(± 0.047)	(± 72.62)	(± 0.039)
	RS	84.93	0.447	90.32	0.367	218.29	0.413	643.50	0.442
		(± 14.62)	(± 0.059)	(± 15.26)	(± 0.045)	(± 28.68)	(± 0.045)	(± 68.86)	(± 0.038)
Contributions to $\widehat{Var}(TBV)$ *due to indirect genetic effects*									
$2(\overline{n_{row}}\,\overline{f_{row}} + \overline{n_{col}}\,\overline{f_{col}} + \overline{n_{diag}}\,\overline{f_{diag}})\,\hat{\sigma}_{a_{di}}$	FS	49.52	0.263	-108.94	-0.445	-402.60	-0.765	-1394.8	-0.951
		(± 9.97)	(± 0.047)	(± 16.79)	(± 0.061)	(± 41.68)	(± 0.067)	(± 103.1)	(± 0.055)
	RS	27.47	0.144	-62.25	-0.253	-273.60	-0.517	-955.98	-0.657
		(± 5.53)	(± 0.026)	(± 8.19)	(± 0.029)	(± 24.17)	(± 0.038)	(± 62.34)	(± 0.032)
$(\overline{n_{row}}\,\overline{f_{row}} + \overline{n_{col}}\,\overline{f_{col}} + \overline{n_{diag}}\,\overline{f_{diag}})^2\ \hat{\sigma}_{a_i}^2$	FS	11.38	0.060	38.79	0.158	202.82	0.385	724.82	0.494
		(± 4.24)	(± 0.022)	(± 9.57)	(± 0.039)	(± 32.58)	(± 0.061)	(± 87.80)	(± 0.059)
	RS	3.18	0.017	13.38	0.054	95.85	0.181	393.46	0.271
		(± 1.26)	(± 0.007)	(± 2.94)	(± 0.012)	(± 13.46)	(± 0.026)	(± 43.56)	(± 0.030)

[a] The reduced set (RS) comprises the neighbor positions that were found to contribute most to IGEs (i.e. see "Neighbors kept in RN" in Table 1).

[b] Results obtained from modeling IGEs as a combined effect by using all eight immediate neighbors of a focal tree were previously reported by Costa e Silva *et al*. (2013) for MLD at age 2 years, and DBH at ages 2 and 4 years. However, the results provided above for the full set of neighbor positions differ slightly to those presented in Costa e Silva *et al*. (2013), as some parents have been subsequently re-allocated to the factorial to improve the number of families per parent and geographic focus of the parents in the diallel (see *Materials and Methods*).

[c] For MLD, the parameter estimates refer to the (arcsine) transformed and rescaled observations (for more details, see footnote *e* of Table 1).

[d] For $\hat{\sigma}_{a_d}^2$, the absolute values of the estimates are also shown in Table 2.

3. Supplementary figures

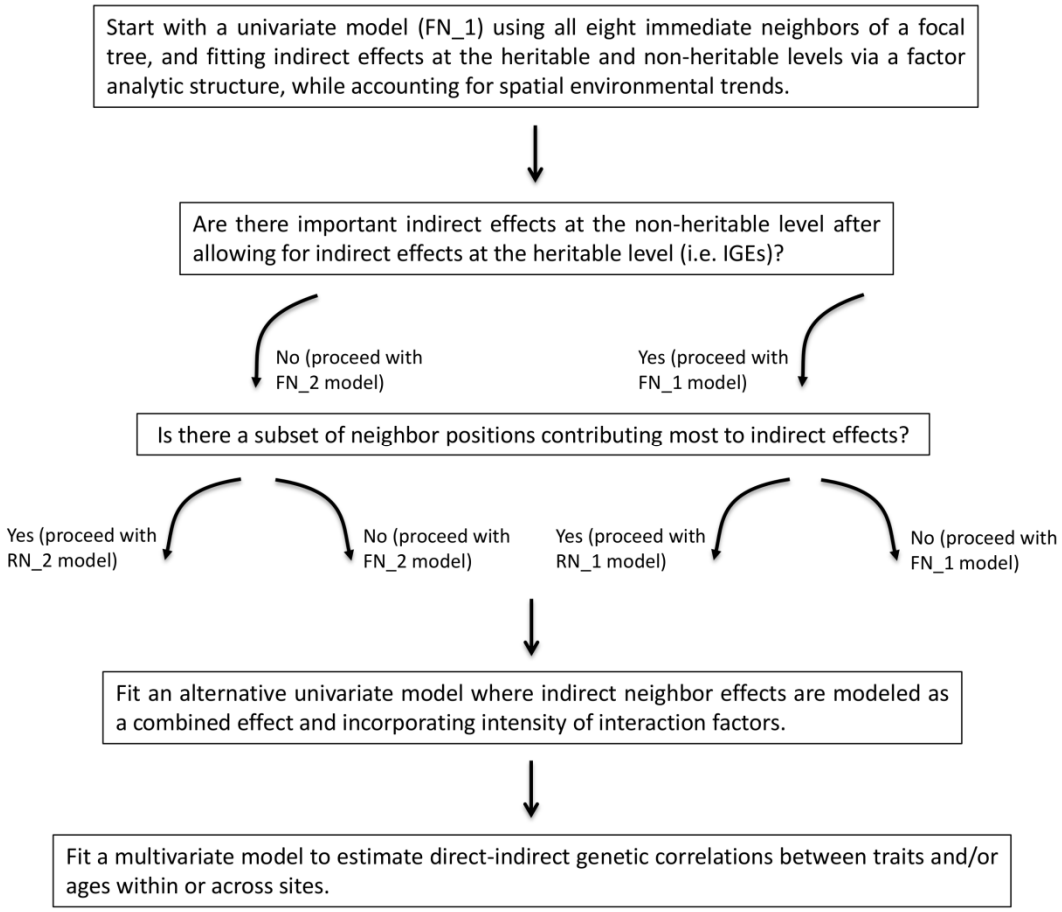

Figure S1 Proposed pathway for analyses to model indirect effects at the heritable and non-heritable levels in forest genetic trials. The models correspond to: FN_1, modeling indirect effects at both the heritable and non-heritable levels, using all eight immediate neighbors of a focal tree; FN_2: modeling indirect effects at the heritable level only, using all eight immediate neighbors of a focal tree; RN_1 modeling indirect effects at both the heritable and non-heritable levels, using subsets of the neighbor positions; and RN_2 modeling indirect effects at the heritable level only, using subsets of the neighbor positions. Note that "RN_2" corresponds to the acronym "RN" used in the sequence of analyses followed to model indirect effects for the MLD and DBH data measured in the *Eucalyptus globulus* trial (see Figure 1).

4. Supplementary references

Akaike H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716-723.

Bijma P (2011). A general definition of the heritable variation that determines the potential of a population to respond to selection. *Genetics* **189**: 1347-1359.

Bijma P, Muir WA, Van Arendonk JAM (2007). Multilevel selection 1: Quantitative genetics of inheritance and response to selection. *Genetics* **175**: 277-288.

Burnham KP, Anderson DR (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag: New York, USA.

Cappa EP, Cantet RJC (2008). Direct and competition additive effects in tree breeding: Bayesian estimation from an individual tree mixed model. *Silvae Genetica* **57**: 45-56.

Costa e Silva J, Graudal L (2008). Evaluation of an international series of *Pinus kesiya* provenance trials for growth and wood quality traits. *Forest Ecology and Management* **255**: 3477-3488.

Costa e Silva J, Kerr RJ (2013). Accounting for competition in genetic analysis, with particular emphasis on forest genetic trials. *Tree Genetics and Genomes* **9**: 1-17.

Costa e Silva J, Potts BM, Bijma P, Kerr RJ, Pilbeam DJ (2013). Genetic control of interactions amongst individuals: Contrasting outcomes of indirect genetic effects arising from neighbour disease infection and competition in a forest tree. *New Phytologist* **197**: 631-641.

Forster M, Sober E (1994). How to tell when simpler, more unified, or less *ad hoc* theories will provide more accurate predictions. *British Journal for the Philosophy of Science* **45**: 1-35.

Gilmour AR, Cullis BR, Verbyla AP (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics* **2**: 269-293.

Kass RE, Raftery AE (1995). Bayes factors. *Journal of the American Statistical Association* **90**: 773-795.

Kass RE, Wasserman L (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association* **90**: 928-934.

Kenny DA, Mannetti L, Pierro A, Livi S, Kashy DA (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology* **83**: 126-137.

Kuha J (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research* **33**: 188-229.

Link WA, Barker RJ (2006). Model weights and the foundations of multimodel inference. *Ecology* **87**: 2626-2635.

Mardia KV, Kent JT, Bibby JM (1988). *Multivariate Analysis*. Academic Press: London, UK.

Molenberghs G, Verbeke G (2011). A note on a hierarchical interpretation for negative variance components. *Statistical Modelling* **11**: 389-408.

Müller S, Scealy JL, Welsh AH (2013). Model selection in linear mixed models. *Statistical Science* **28**: 135-167.

Oliveira IRC, Molenberghs G, Verbeke G, Demétrio CGB, Dias CTS (2016). Negative variance components for non-negative hierarchical data with correlation, over-, and/or underdispersion. *Journal of Applied Statistics*, DOI: 10.1080/02664763.2016.1191624.

Pryseley A, Tchonlafi C, Verbeke G, Molenberghs G (2011). Estimating negative variance components from Gaussian and non-Gaussian data: A mixed models approach. Computational *Statistics and Data Analysis* **55**: 1071-1085.

Richards SA (2005). Testing ecological theory using the information-theoretic approach: examples and cautionary results. *Ecology* **86**: 2805-2814.

Richards SA (2008). Dealing with overdispersed count data in applied ecology. *Journal of Applied Ecology* **45**: 218-227.

Schwarz G (1978). Estimating the dimension of a model. *Annals of Statistics* **6**: 461-464.

Stringer JK (2006). *Joint Modelling of Spatial Variability and Interplot Competition to Improve the Efficiency of Plant Improvement*. PhD Thesis, University of Queensland, Brisbane.

Stringer JK, Cullis BR, Thompson R (2011). Joint modeling of spatial variability and within-row interplot competition to increase the efficiency of plant improvement. *Journal of Agricultural, Biological and Environmental Statistics* **16**: 269-281.

Sugiura N (1978). Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods* **A7**: 13-26.

Verbyla AP, Cullis BR, Kenward MG, Welham SJ (1999). The analysis of designed experiments and longitudinal data using smoothing splines. *Journal of the Royal Statistical Society Series C (Applied Statistics)* **48**: 269-300.

Vrieze SI (2012). Model selection and psychological theory: a discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods* **17**: 228-243.

Wolfinger R (1993). Covariance structure selection in general mixed models. *Communications in Statistics - Simulation and Computation* **22**: 1079-1106.