

利用显性分子标记和 F_1 群体进行林木遗传连锁图谱的构建

尹佟明 黄敏仁

(南京林业大学森林遗传与基因工程开放实验室)

朱立煌

(中国科学院遗传所植物生物技术重点实验室)

构建遗传图谱是当前生物学研究领域中的一个热点,进行图谱构建,林木与作物既有共性,又有各自的特点。因为进行图谱构建一是要有合适的分离群体,二是要有能揭示亲本多态性的遗传标记。两者可利用的遗传标记是一样的,但在进行作图群体构建时却存在一定差异,林木的遗传组成高度杂合,大多数林木为长期异交的树种,一般都有自交不亲和和近交衰退现象,象作物那样利用近交系或其它的高世代群体进行遗传图谱构建是不太可能的。由于研究方法的不同,以前的林木育种工作者很少留意建立和保存谱系清楚的 F_2 和 BC_1 群体,由于林木生命周期很长,建立这样一些高世代的群体也需要很长的时间,而且除了杨树、柳树等一些可在温室内水培杂交的物种,大多数林木的控制授粉杂交操作也是很困难的。因此研究过程中,等待材料的问题成为目前林木遗传图谱构建的主要限制因子之一。如何利用现成的材料和低世代群体,对林木遗传图谱构建工作的广泛开展具有重要意义。下面就这一问题作一些理论上的探讨和分析。

1. 理论分析

1.1 标记在群体中的分离

由于林木的高度杂合性,许多基因位点在 F_1 即发生分离,所以 F_1 群体能够满足成为作图群体的条件,这样的群体一般是现成的或者可以在短时间内建成,而通过 PCR 反应获得 RAPD 标记,由于不需要印迹、杂交和显影等一系列繁琐步骤,显示出简便高效的特点,使我们有可能在短时间内获得大量的基因组信息,而且 RAPD 标记本身为一种显性标记,在遗传行为上存在显隐性现象,即带的有无 (+/-),虽然对比于共显性的 RFLP 标记, RAPD 标记会丢失部分信息,但其显隐性遗传方式却给我们带来了分析上的简化,因为我们可以通过合适

的交配进行“假”测交,使杂合的位点在后代的分离中产生 1:1 的分离比,假设图式如下:

(1) $AA \times aa$ 或 ∞ , 不分离

P_1	P_2	F_1				
—		—	—	—	—
AA	aa 或 ∞	Aa 或 Ao	Aa 或 Ao	Aa 或 Ao	Aa 或 Ao

(2) $Aa \times aa$ 或 ∞ , 1:1 分离

P_1	P_2	F_1				
—		—	—	—	—
AA	aa 或 ∞	Aa 或 Ao	aa 或 ao	Aa 或 Ao	aa 或 ao

其中 A 代表显性位点, a 代表与 A 对应的隐性

位点,○代表相应的零等位基因位点。

上图中有带出现的可能基因型为显性纯合或显隐性杂合,无带的情形为隐性纯合或无相应的等位基因(即零等位)。通过杂交后,显性纯合位点不发生分离,而杂合位点则发生上述分离,且期望分离比为 1:1,这一分离模式与花培双单倍体或回交群体的分离模式以及可采用的作图模型是相似的。而且可以同时双亲进行作图。在引物的筛选上我们只要通过适当利用一部分 F_1 个体,就能鉴别可利用的引物,即将在双亲中能检测到多态性的引物对 F_1 进行检测,如果 F_1 个体间存在多态性,则这一引物就可用于图谱的构建。但检测多少个 F_1 个体合适呢?如果所用 F_1 个体太多必然增加工作量和作图成本,根据期望概率,由下式确定所选个体数: $\eta = (1-1/2)^n\%$,其中 η 为信息损失量, n 为所用个体数。如果所用 F_1 个体数为 5,则在 F_1 中检测不到多态性的概率为 1/32,即可能有 3% 的信息损失量。所以我们认为利用 5 个以上的 F_1 个体进行多态性检测就可保证有 97% 以上的概率使合适的引物得以入选。这一方法集中了三交与测交的特点,利用现成的杂合度较高的个体作亲本,与另一亲缘关系相对较远的个体进行交配,产生多数测交分离位点,但双亲的亲缘关系也不易太远,否则即使杂交亲合,子代中可能非整倍体现象严重,对作图造成干扰。华盛顿州立大学在建立毛果杨×美洲黑杨的 F_2 群体过程中,利用 RFLP 分析和流式细胞计对 11 个 F_1 的分析表明,其中有 3 个个体是三倍体或非整倍体。因此杂交要考虑到避免多倍体或非整倍体的出现。

1.2 连锁检验与重组频率

下面简要说明一下利用显性标记和这类群体进行标记间连锁检验和重组率计算的数学模型与方法。

由于隐性位点与零等位的情形在实验结果上表现是一致的,均为无带出现,所以统一计为 0。对于两位点 A,B,假设它们在双亲中的重组

率相等($P_1 = P_2 = P$),则这两个位点在子代中的分离情形如表 1 和表 2。

表 1 $\frac{AB}{OO} \times \frac{OO}{OO}$ 配子频率分布

	GF	$\frac{(1-P)}{2}$	$\frac{(1-P)}{2}$	$\frac{P}{2}$	$\frac{P}{2}$	
GF	GT	OO	OO	OO	OO	配子数
$\frac{(1-P)}{2}$	AB	AB	AB	AB	AB	a
$\frac{(1-P)}{2}$	OO	OO	OO	OO	OO	b
$\frac{P}{2}$	AO	AO	AO	AO	AO	c
$\frac{P}{2}$	OB	OB	OB	OB	OB	d

	AB	AO	OB	OO
完全连锁:	8	: 0	: 0	: 8
自由组合:	4	: 4	: 4	: 4

表 2 $\frac{AO}{OB} \times \frac{OO}{OO}$ 配子频率分布

	GF	$\frac{(1-P)}{2}$	$\frac{(1-P)}{2}$	$\frac{P}{2}$	$\frac{P}{2}$	
GF	GT	OO	OO	OO	OO	配子数
$\frac{(1-P)}{2}$	AO	AO	AO	AO	AO	a
$\frac{(1-P)}{2}$	OB	OB	OB	OB	OB	b
$\frac{P}{2}$	AB	AB	AB	AB	AB	c
$\frac{P}{2}$	OO	OO	OO	OO	OO	d

	AO	OB	AB	OO
完全连锁:	8	: 8	: 0	: 0
自由组合:	4	: 4	: 4	: 4

GF 为配子频率

GT 为配子类型

表中,如果 A、B 完全连锁($P = 0$),则双亲

表型在 F_1 子代中各占 50%, 如果自由分离 ($P = 0.5$), 则出现四种表型 (两个双亲型, 两个重组型), 其频率各为 25%。所以重组率可以通过与完全连锁和独立分配情形带型的差异进行计算。

利用 χ^2 检验, 可以检验表型 (带型) 分离是否严重偏离独立分配的期望分离比。如果我们同时考虑两个要检验的基因位点, 就可以将分离比的检验与连锁检验合并起来进行。这是实际分析过程中常用的方法。我们所讨论的连锁形式为 (1:1) — (1:1) 类型。可以进行如下的 χ^2 检验:

$$\chi^2_T = (a^2 + b^2 + c^2 + d^2) - n \quad df = 3$$

$$\chi^2_A = (a + b - c - d)^2 / n \quad df = 1$$

$$\chi^2_B = (a - b + c - d)^2 / n \quad df = 1$$

$$\chi^2_L = (a - b - c + d)^2 / n \quad df = 1$$

χ^2_T : 总的卡方值, 它可以分解成为三个部分

χ^2_A : 测验 A 基因座位的分离比例是否异常

χ^2_B : 测验 B 基因座位的分离比例是否异常

χ^2_L : 测验 A、B 基因座之间是否存在连锁。

如果 A 与 B 相连锁, 则重组频率可以通过解下面的极大似然方程进行估算

$$\frac{\delta \ln L(P)}{\delta P} = \sum_j Z_j \frac{1}{P_j} \frac{\delta P_j}{\delta P} = 0 \quad (1)$$

这里 P_j 为期望频率, Z_j 为观察到的表型数, 式中各项的计算如表 3

表 3 式 (1) 中各项的计算值

Z_j	配子类型	P_j	$\frac{\delta P_j}{\delta P}$	$\frac{1}{P_j} \frac{\delta P_j}{\delta P}$	$\frac{1}{P_j} \left(\frac{\delta P_j}{\delta P} \right)^2$
AB	$\frac{1-P}{2}$	$\frac{1}{2}$	$\frac{-1}{1-P}$	$\frac{1}{2(1-P)}$	a
AO	$\frac{P}{2}$	$\frac{1}{2P}$	$\frac{1}{2}$	$\frac{1}{2P}$	b
OB	$\frac{P}{2}$	$\frac{1}{2P}$	$\frac{1}{2}$	$\frac{1}{2P}$	c
OO	$\frac{1-P}{2}$	$\frac{-1}{2}$	$\frac{-1}{1-P}$	$\frac{1}{2(1-P)}$	d
合计	1	0	$i = \frac{1}{P(1-P)}$		n

重组频率 P 的估计质量可由下式计算

$$-E \left(\frac{\delta^2 \ln(P)}{\delta P^2} \right) = n \sum_j \frac{1}{P_j} \left(\frac{\delta P_j}{\delta P} \right)^2 = n_i = I_p \quad (2)$$

这里 n 为样本量 (作图群体的个体数), 重组率 P 的方差为

$$V(P) = \frac{1}{I_p} \quad (3)$$

其标准差为

$$\sqrt{V(P)} = \sqrt{\frac{1}{I_p}} \quad (4)$$

极大似然估计值为重组频率 P 的最小方差无偏估计值。

其中表 3 中的期望频率 P_j 是通过特定表型的配子频率的乘积加和得到的。如

$$P_{\infty} = 2 \left(\frac{1-P}{2} \frac{1-P}{2} \right) + \left(\frac{P}{2} \frac{1-P}{2} \right) = \frac{1-P}{2} \quad (5)$$

假定 $P_1 = P_2 = P$, 由于其它各项均为 P_j 的导数, 将上述各项带入方程 1, 则方程 1 变为

$$\frac{-a}{1-P} + \frac{b}{P} + \frac{c}{P} + \frac{-d}{1-P} = 0 \quad (6)$$

对 P 求解得到

$$P = \frac{b+c}{a+b+c+d} = \frac{b+c}{n} \quad (7)$$

由式 2、3 得到

$$\bar{V}(P) = \frac{P(1-P)}{n} \quad (8)$$

对于表 2 的情形, 由于重组频率不大于 0.5, 所以我们容易识别重组子。

上面分析是两位点情形, 我们实际可检测到的这种位点是很多的, 通过多点分析和基因的直线排序, 利用 kossambi 函数完成重组率与图距间的转换, 就可以完成图谱构建。

Carlson 等人对花旗松和白云杉的双列杂交子代进行了 RAPD 标记分离研究, 证实大量测交位点的存在, 对于花旗松, 所利用的 10 个引物有 7 个检测到了在 0.05 显著水平上符合 1:1 分离的位点。Dario 等通过假测交的策略, 利用 RAPD 标记进行巨桉 × 尾叶桉的图谱构建。建成的图谱中, 巨桉具有 240 个标记, 分属

14 个连锁群, 覆盖的基因组长度约为 1550 cM, 尾叶桉图谱具有 251 个标记, 分属 11 个连锁群(桉树具 11 对染色体), 覆盖的基因组长度约为 1101 cM。两个图谱上的标记对基因组的覆盖率都达到了 95% 以上。最近又发表了长叶松×湿地松利用 F_1 群体和 RAPD 标记建成的图谱。

2. 讨论

已发表的林木作图群体有以下几种: ① F_2 群体, 如杨树、火炬松等; ② 回交群体, 如美洲栗; ③ F_1 群体, 如长叶松、湿地松; ④ 利用大配子体进行针叶树单株作图, 如火炬松、湿地松、花旗松、欧洲赤松、糖松等。

上述群体中, F_2 群体适用于杂合度低的树种。其特点是: 群体具有永久性质, 利于引物及探针筛选及数量性状的系统研究。可同时利用共分离分析及分离个体混合分析进行基因定位。不足之处是: 建群时间长, 有些树种由于近交不亲合或近交衰退现象, 难以建群。同时非整倍体出现机率增加。

回交群体与 F_2 群体相似, 群体具有永久性质。尤适用于利用显性标记作图, 分析简单。但与 F_2 群体相比, 降低了可检测的位点数。同时也受近交衰退的影响。

F_1 群体适合于杂合度较高的树种。其优点为具有永久性质, 可同时利用共分离分析和分离个体混合分析法进行基因定位, 能避免近交衰退的影响, 建群时间短。缺点是: 增加筛选引物和探针的工作量。

利用大配子体作图是大多数针叶树可以采用的群体。因为大配子体是纯合单倍体, 而且分析简便, 不需专门建群。但是暂时性群体, 作图与定位不能同时进行, 应用有一定局限性。

综合来看, F_1 群体对林木而言是一个比较理想的作图群体, 本文介绍的方法为这种群体在林木中的应用提供了一种更有效的交配设计。文中的方法利用显性标记, 结合三交与回交特点, 通过零等位增加测交分离位点数, 这样就会获得足够的位点构建较为饱和的遗传图谱, 这种群体可以利用现成的杂合度较高的个

体与相应的较远亲缘关系的个体通过 1 代杂交得到。这样的群体对许多树种都有现成的材料或可以在短期内建成。因此针对林木的特点这是一种简单有效的作图群体构建方法。但 RAPD 这样一种分子标记不能使所有的分离位点都能得以利用, 从而影响图谱的饱和度, 而且这一方法增加了引物筛选的工作量, 因为它需要通过一定数量的 F_1 个体对引物进行筛选。我们实际可利用的信息还包括双亲中都为杂合的位点, 其分离比为 3:1, 所以用显性的分子标记作图时, 我们全部可利用的信息包括如下几种分离与连锁检验模型: (1:1) - (1:1); (1:1) - (3:1); (3:1) - (3:1)。但双亲中均为杂合的位点在亲本中检测不到多态位点, 其在 F_1 中能够检测出多态性的效率 $\eta = 100\% - (1 - 1/4)^n \%$, 当 $n = 5$ 时, 只有 70% 的检出率, 工作量增加相对较大, 而且分析较为复杂。在图谱的饱和度要求不太高的情况下, 可以不考虑这类情况。所以在图谱的快速构建过程中, 尤其是对框架图的构建, 这是一种极为有效的方法。而且我们可以通过合适的选配和增加所用引物的数量, 使图谱达到较高的饱和度。这一方法还存在一个标记本身的缺点, 由于我们还不完全了解 RAPD 反应的机理, 在实验过程中我们一定要严格控制实验条件, 保证结果的稳定性。RAPD 相对的不稳定性成为其最主要的限制因子。但这些不稳定因素都是可以通过控制实验条件而加以避免的, 而且 RAPD 标记的简单与快速是其它标记所不能比拟的。在基因位点研究方面, F_1 群体不适合于通过选型交配研究目的基因, 因为所选的具有极端表现的性状在子一代中, 一般不发生分离。在 QTLs (Quantitative trait loci) 研究时可通过共分离分析 (Cosegregating analysis) 与混合个体分离分析 (Bulked segregants analysis) 结合进行。

参考文献

- [1] 徐云碧, 朱立煌 (1993), 分子数量遗传学, 浙江农业出版社。
- [2] Carlson J, et, al. (1991) Theor Appl Genet 83: 194—

- 200.
- [3] Bradshaw HD, et, al. (1994) Theor Appl Genet 89: 167—178.
- [4] Dendrome.
- [5] Nelson C D, et, al. (1993) Theor Appl Genet 87: 145—151.
- [6] Binelli G, et, al. (1994) Theor Appl Genet 88:283—288.
- [7] Tulsieram L K, et, al. (1992) Biotechnology 10:686—690.
- [8] Williams JGK, et, al. (1990) Nucleid Acids Research 18:6531—5635.
- [9] Ritter E, et, al. (1990) Genetics 125:645—654.
- [10] Benet H, et, al. (1995) Theor Appl Genet 90:1068—1073.
- [11] Grattapaglia D, et, al. (1994) Genetics 137:1121—1137.
- [12] Devey ME, et, al. (1994) Theor Appl Genet 88:273—278.
- [13] Kubisiak TL, et, al. (1995) Theor Appl Genet 90: 1119—1127.
- [14] Carlson JE, et, al. (1991) Theor Appl Genet 83:194—200.
- [15] Plomion C, et, al. (1995) Theor Appl Genet 90: 1028—1034.

Construction of Genetic Linkage Map in Forest Tree Using Dominant Molecular Markers and F_1 Pedigree

Yin Tongming Huan Minren Zhu Lihuang

(Institute of Genetics, Academic Sinica, Beijing 100101)

Abstract

Most linkage maps in plants have been obtained from segregating pedigree derived from crosses between inbred lines. Such pedigrees or even three-generation pedigrees are generally not available in trees and are difficult to obtain due to a significant genetic load and time constraints. To incorporate molecular marker assisted strategies into forest tree breeding, it is imperative to explore alternative approaches for the construction of linkage maps that make use of pedigrees already existing and commonly generated in tree breeding programs. Forest tree has a high heterozygosity, so the F_1 progeny is a segregating pedigree. When analysed with dominant markers, such as RAPD markers, many single dose dominant markers will be heterozygous in one parent, null in the other and therefore segregate 1:1 in their F_1 progeny following a testcross configuration. This essay propose the combined use of dominant markers and F_1 progeny as a strategy for the construction of genetic linkage map of both parents(full sib) or of the maternal tree(half sib) in forest trees as well as in any highly heterozygous sexually reproducing living organism. The optimizing experiment protocol is suggested and the theory basis of this strategy is discussed in detail.