

参试子代样本数对数量性状 遗传分析结果的影响*

齐 明

摘要 以七年生的杉木 9×9 半双列交配试验林为研究材料,从全部的试验数据资料 40 个重复区组中,随机地抽取部分资料(即 8, 16, 24 及 32 个区组重复),进行方差分析和配合力分析,分别从随机模型和固定模型的角度出发,系统地考察了参试子代样本数大小对杉木主要经济性状遗传分析结果的影响。结果表明:不论是利用随机模型,还是利用固定模型进行分析,参试子代样本数大小对遗传分析结果有着显著的影响,并且这种影响随着研究性状的变化而有所不同。这一研究结果说明:在林木子代试验及其统计分析中,每个处理内必须要有一个合适的参试子代样本数,其试验分析结果才是正确可靠的。

关键词 参试子代样本数 随机模型 固定模型 遗传参数

遗传参数对于育种工作者制订遗传改良方案,选育优良品系具有十分重要的作用。我国自 70 年代展开林木良种选育工作以来,数量遗传学作为一种有力工具已被广泛地应用到林木选育种实践中,促进了优良品系的育成。但是根据我国林木遗传学研究积累的信息资料看,遗传参数的估计结果,不仅与研究材料特性,而且与田间试验设计类型以及分析方法有关。除此之外,参试子代样本数大小也对遗传分析结果有着重要的影响^[1]。本文以杉木 9×9 半双列交配试验的数据为例,来说明参试子代样本数对遗传分析结果的影响。

1 材料与方法

1.1 研究材料

研究材料为七年生的杉木 9×9 半双列全同胞子代试验林。关于亲本来源等方面更详细的信息可参阅文献[2]。田间试验设计为平衡随机区组,单株小区,重复 40 次。1988 年对这片七年生试验林进行了全林调查,测量并推算了材积、树高、胸径及枝下高等性状。

1.2 研究方法

以单株观察值参与统计分析,方差分析的线性模式为: $y_{ijk} = \mu + F_{ij} + b_k + \epsilon_{jk}$ ($i = 1, 2, \dots, 9$; $j = 1, 2, \dots, 9$; $k = 1, 2, \dots, 40$)。式中: μ 为群体平均效应, F_{ij} 为第 i 个母本与第 j 个父本杂交产生组合的平均效应, b_k 为区组效应, ϵ_{jk} 为观察值 y_{ijk} 的试验误差。

配合力分析按 Griffing 方法进行,其线性模型为: $F_{ij} = g_i + g_j + S_{ij}$, 式中, g_i (或 g_j) 为第 i 个亲本的一般配合力效应, S_{ij} 为第 ij 个组合的特殊配合力效应。

全同胞家系的广义遗传力为: $h^2 = 1 - 1/F$, 这里 F 值为方差分析检验值。

1996—12—03 收稿。

齐明助理研究员(中国林业科学研究院亚热带林业研究所 浙江富阳 311400)。

本文由亚林所杉木课题组提供研究材料。

$$h_f^2 \text{ 的标准误}^{[3]} \text{ 为: } \delta_{h_f^2} = \frac{(1 - h_f^2)[1 + (b - 1)h_f^2]}{b \times (b - 1) \times (f - 1)/2}$$

式中, b 为区组数; f 为参试家系数目。

$$\text{试验分析的精确性: } \frac{MSe}{b},$$

$$\text{单株广义遗传力 } h_i^2 = \frac{4\delta_g^2 + 4\delta_s^2}{2\delta_g^2 + \delta_s^2 + \delta_b^2 + \delta_e^2}$$

式中, δ_g^2 为一般配合力方差分量, δ_s^2 为特殊配合力方差分量, δ_b^2 为区组方差分量, δ_e^2 为随机误差的方差分量。

$$h_i^2 \text{ 的标准误}^{[4]} \text{ 为: } \delta_{h_i^2} = \frac{(1 - 1/2h_i^2)[1 + 1/2(NBS)h_i^2]}{(NBS)^2 \times (F - 1)/2}$$

式中, NBS 为每个全同胞家系的参试子代样本数; F 值为方差分析检验值。

为了研究参试子代样本数对遗传分析结果的影响, 从全部试验资料 40 个区组重复中, 分别随机地抽取部分资料: 8, 16, 24 及 32 个区组重复数(即参试子代样本数分别为 8, 16, 24, 32 及 40), 分别进行方差分析和配合力分析, 既从随机模型, 又从固定模型的角度, 系统地考察参试子代样本数对诸多遗传参数的大小及精度方面的影响。

2 结果与分析

2.1 在随机模型条件下的分析结果

七年生杉木 9×9 半双列交配试验, 当参试子代样本数变化时, 各性状遗传分析结果列于表 1。

表 1 随机模型条件下, 参试子代样本数对遗传分析的影响

样本数	相对分量(%)				h_f^2 (%)	$\delta_{h_f^2}$ (%)	h_i^2 (%)	$\delta_{h_i^2}$ (%)	精确性	相对分量(%)				h_f^2 (%)	$\delta_{h_f^2}$ (%)	h_i^2 (%)	$\delta_{h_i^2}$ (%)	精确性
	GCA	SCA	ERROR							GCA	SCA	ERROR						
	材 积									树 高								
8	12.78	22.93	64.29	42.60	7.41	14.72	31.22	5.358	6	25.59	15.21	59.20	49.67	7.30	18.32	28.43	0.328	3
16	47.55	1.96	50.49	61.46	6.17	16.28	15.03	3.746	3	50.16	7.82	42.02	68.30	5.58	20.52	14.37	0.231	7
24	62.56	1.59	35.85	74.72	4.75	19.58	10.51	2.854	1	58.62	9.90	31.48	77.20	4.41	21.65	10.33	0.185	9
32	60.04	1.97	37.99	72.22	5.00	13.97	8.32	2.486	8	57.32	12.69	29.99	79.69	4.02	19.47	8.76	0.157	9
40	63.02	0.86	36.12	73.35	5.18	12.20	6.94	2.212	1	58.95	12.35	28.70	80.89	4.05	17.62	8.35	0.142	3
	胸 径									枝 下 高								
8	方差分析结果家系间差异不显著									32.04	27.02	40.94	65.85	6.21	30.54	24.28	0.254	9
16	19.76	12.05	68.19	38.16	6.51	4.87	15.82	0.415	5	57.02	22.93	20.05	84.59	3.30	39.85	12.73	0.159	1
24	29.57	2.99	67.44	44.87	6.44	6.34	11.39	0.332	6	72.02	3.00	24.98	82.61	3.59	28.10	10.22	0.137	6
32	47.68	7.69	44.63	65.95	5.62	10.82	8.30	0.287	5	68.86	9.06	22.15	84.84	3.19	25.79	8.38	0.117	4
40	55.74	3.42	40.84	70.01	5.60	10.56	6.89	0.263	7	63.71	18.55	17.75	87.11	2.91	25.38	7.23	0.106	6

表 1 可见, 随着参试子代样本数的增加, 遗传分析结果的可靠性在不断增加, 这可从试验分析的精确性以及遗传参数的误差大小看出来。当参试子代样本数为 40 时, 即全部资料参与分析时, 具有最高的精确性和最小的遗传参数误差, 其结果的可靠性最高, 而当参试子代样本数为 8 时, 分析结果最不可靠, 不仅因为整个试验分析的精确性最差, 而且因为此时遗传参数的误差很大。

在同一研究性状中, 参试子代样本数大小对遗传分析结果有着十分显著的影响。例如在材

积这一研究性状中,当参试子代样本数为 8 时,其配合力分析结果是随机误差(ERROR)居第一位,特殊配合力(SCA)次之,一般配合力(GCA)处第三位,这与全部资料分析结果完全相反。另外此时全同胞家系遗传力偏小,误差偏大,单株遗传力的误差远比遗传力值大,可靠性差。当参试子代样本数增加到 16 时,其配合力分析结果与全部资料分析结果间仍存在着差异。只有当参试子代样本数增加到 24 时,各种遗传参数才表现出接近全部资料分析结果的趋势。当参试子代样本数为 32 时,材积性状部分资料的分析结果与全部资料分析结果完全一致。全同胞家系遗传力及单株遗传力趋于稳定,其误差也基本稳定,这种趋势与整个试验分析的精确性趋于稳定是一致的。

参试子代样本数的大小对遗传分析结果的影响,随研究性状的变化而有所不同。例如,在参试子代样本数为 8 时,胸径方差分析其全同胞家系间差异性不显著;材积的配合力分析结果完全是错误的;而在树高及枝下高这两个性状上,其配合力分析结果与全部分析资料分析结果仍有较大的差异。此时在所有的研究性状上,全同胞家系遗传力偏低,误差偏大;单株遗传力的误差太大,遗传参数这种低可靠性与整个试验分析精确性差是一致的。

2.2 在固定模型条件下的研究结果

在固定模型条件下,当参试子代样本数分别为 8、16、24、32 及 40 时,对上述材料的材积、树高等 4 个性状作方差分析和配合力分析,以全部资料分析结果作为参比标准,每组部分资料分析结果分别与之比较:(1)以方差分析获得的全同胞家系平均效应为依据,考察该效应值间的相似性、误差大小,以及以此为准进行全同胞家系选择时失误比率大小;(2)以配合力分析获得的亲本配合力效应值大小为依据,考察部分资料与全部资料每组试验中,亲本一般配合力效应值秩的相同或相异比率,以便了解亲本评定失误的可能性有多大。结果列于表 2。

表 2 固定模型条件下,参试子代样本数对遗传分析结果的影响

项 目	材 积				树 高			
	B= 8 与 B= 40 比较	B= 16 与 B= 40 比较	B= 24 与 B= 40 比较	B= 32 与 B= 40 比较	B= 8 与 B= 40 比较	B= 16 与 B= 40 比较	B= 24 与 B= 40 比较	B= 32 与 B= 40 比较
家系平均效应值间相关	0.715 6	0.867 8	0.940 5	0.971 2	0.777 4	0.860 2	0.934 2	0.973 5
相对误差 $E = \frac{\delta_x}{x}$ (%)	14.05	9.01	6.18	3.53	4.18	2.89	1.98	1.11
分析精度 $1-E$ (%)	85.95	90.99	93.82	96.47	95.82	97.11	98.02	98.89
前七名家系彼此重叠情况 ^①	7(3/4)	7(4/3)	7(5/2)	7(5/2)	7(5/2)	7(6/1)	7(5/2)	7(7/0)
亲本 GCA 效应值秩变化 ^②	9(6/3)	9(7/2)	9(7/2)	9(9/0)	9(3/6)	9(3/6)	9(7/2)	9(9/0)
	胸 径				枝 下 高			
家系平均效应值间相关	0.812 2	0.896 9	0.960 2	0.804 3	0.931 5	0.948 3	0.979 5	
相对误差 $E = \frac{\delta_x}{x}$ (%)	3.79	2.27	1.45	19.53	9.52	5.98	3.40	
分析精度 $1-E$ (%)	96.21	97.73	98.55	80.47	90.48	94.02	96.60	
前七名家系彼此重叠情况	7(4/3)	7(5/2)	7(6/1)	7(6/1)	7(6/1)	7(6/1)	7(6/1)	
亲本 GCA 效应值秩变化	9(2/7)	9(3/6)	9(4/5)	9(4/5)	9(5/4)	9(5/4)	9(7/2)	

①假定入选率为 20%,该栏中,例如 7(5/2)表示 7 个入选家系中,5 个家系相同,2 个不同。另外在不同组间结果相同的,但入选的家系并不重叠。②表示在两组试验中,亲本 GCA 效应值秩的相同或相异比率,该栏中,例如 9(7/2)表示 9 个亲本中 7 个亲本 GCA 效应值秩相同,2 个相异;同样在不同组间结果相同的,但评定的亲本并不重叠。

从表 2 可见,不同的参试子代样本数对同一研究性状遗传分析结果有明显的影响。例如材积,在 B= 8 与 B= 40 这组试验中,虽然两者全同胞家系平均效应值之间具有明显相关($r = 0.715 6$),但相对误差较大($E = 14.05\%$),以部分资料分析结果进行全同胞家系评选,失误率为

4/7; 以部分资料的配合力分析结果进行亲本评定, 也会产生明显失误, 失误比率为 3/9。随着参试子代样本数增加, 部分资料分析结果有接近全部资料分析结果的倾向。在 $B=16$ 与 $B=40$ 这组试验中, 全同胞家系平均效应值间的相关达到 0.8678, 相对误差降到 9.01%, 全同胞家系评选失误率降到 3/7, 亲本配合力评定失误率为 2/9。而在 $B=24$ 与 $B=40$ 这组试验中, 全同胞家系平均效应值间相关为 0.9405, 相对误差为 6.18%, 全同胞家系评选失误率降至 2/7, 亲本配合力评定失误率仍为 2/9。在 $B=32$ 与 $B=40$ 这组试验中, 全同胞家系平均效应值间的相关又有进一步加强, 高达 0.9712, 全同胞家系平均值间相对误差下降为 3.53%, 此时部分资料关于样本配合力的评定结果与全部资料分析结果完全一致, 但是全同胞家系评选仍会失误, 失误比率为 2/7。

与材积分析结果类似的, 还有胸径及枝下高。只有树高这一性状的分析结果稍有例外。在 $B=8$ 与 $B=40$ 这组试验中, 全同胞家系平均效应值间相关系数 $r=0.7774$, 相对误差 4.18%, 全同胞家系评选失误率为 2/7, 亲本配合力评定失误率为 6/9。随着参试子代样本数的增加, 部分资料分析结果有接近全部资料分析结果的趋势。在 $B=32$ 与 $B=40$ 这组试验中, 全同胞家系平均值间相关达 0.9735, 相对误差为 1.11%; 此时全同胞家系评选结果, 亲本配合力评定结果, 在两个试验中完全一致。这表明, 参试子代样本数对遗传分析结果的影响, 随研究性状的变化而有所不同。对树高的遗传分析, 32 个以上的参试子代样本数是合适的。

3 结论与讨论

早在 1960 年 Wright 和 Frelund^[5] 研究了小区大小与区组重复数对林木子代试验分析结果的影响时, 指出单株小区的试验效率最高, 并且在统计分析上具有优越性。1972 年 Kung^[5] 曾尝试将子代试验中用检验到家系间差异显著性时的灵敏度来确定参试的家系数和每个家系内合适的子代数。Johnstone 和 Samuel^[5] 1974 年报道了在 *Picea sitchensis* 和 *Larix* spp. 试验中不同小区大小和区组重复数对子代试验的影响, 当处理内每个家系维持在 36~40 个参试子代时, 可将小区间的方差降低到最小。1984 年 Cotterill 和 James^[5] 报道了林木子代试验的精度与参试子代数和合适小区大小间的关系方面的研究结果: 单株小区或双株不连续小区时, 每个家系 10~20 个参试子代样本便可可靠地评价每个家系。1990 年 Cotterill^[6] 又报道了他在种源试验中, 有关每个种源合适参试家系数及子代数方面的研究结果: 5~10 个家系可以合适地代表一个种源, 对遗传力高的性状, 每个种源有 100 个子代; 对遗传力低的性状每个种源有 200 个子代, 便可可靠地评价种源的平均效应。比较这两份报告可见, Cotterill 更倾向于从精度的角度来研究林木子代试验中参试样本数的问题, 并且他已经注意到了合适的参试子代样本数与研究性状有关。以上种种研究说明关于林木子代试验中参试子代样本数, 迄今还没有得出一个成熟的结论。

根据本文的分析结果, 可以得出以下结论:

(1) 通常 8 个左右的参试子代样本数进行试验分析是行不通的, 其分析结果很可能是错误的, 即使试验数据在方差分析时其差异性达到了显著性水平。由此, 笔者认为, 在进行材性材质的遗传分析时, 参试子代样本数要尽可能地多, 而不应采用小样本进行试验分析。

(2) 就同一研究性状而言, 在固定模型条件下, 每个处理所需要的参试子代样本数比在随机模型条件下要多。如果试验的目的仅仅限于随机模型范围内, 那么每个处理内有 32 个参试

子代样本数, 便可达到试验目的; 而在固定模型条件下, 通常每个处理内少于 32 个参试子代样本数是不合适的。

(3) 根据文献[7], 在固定模型条件下要可靠地测定与评价每个处理(家系)的平均效应值大小, 所需要的参试子代样本数 $n = t_{\alpha}^2(PCV)^2/E^2 + 1$ (式中 t_{α} 是分析结果的可靠性, E 是允许的相对误差, PCV 是性状的表型变异系数)。可见, 参试子代样本数不是一个常数, 而是一个变数, 研究性状相同时, 要求测定分析的可靠性越高, 允许的相对误差越小, 所需要的参试子代样本数就越多。而在可靠性、相对误差一定时, 研究性状的表型变异越大, 每个处理内参试子代样本数就越多, 即对不同的研究性状进行试验分析时, 每个处理(家系)内所需要的参试子代样本数是不同的。在林木子代试验中, 每个处理内合适的参试子代样本数的确定, 应以经济价值高、表型变异系数最大的研究性状所需要的子代样本数为准。杉木遗传改良中, 笔者以为, 应以能可靠地评价材积所需要的子代样本数为准。

参 考 文 献

- 1 潘沈元, 何斯美. 影响数量性状遗传参数估计的两个因素. 遗传, 1991, 13(4): 19 ~ 21.
- 2 齐明. 杉木育种中 GCA 与 SCA 的相对重要性. 林业科学研究, 1996, 9(5): 498 ~ 503.
- 3 戴君惕. 遗传分析方法. 长沙: 湖南科学技术出版社, 1989. 252 ~ 257.
- 4 (美) 乔纳森 W 赖特(郭锡昌, 胡承海译). 森林遗传学. 北京: 中国林业出版社, 1981. 207 ~ 212.
- 5 Cotterill P P, James J W. Number of offspring and plot sizes required for progeny testing. *Silvae Genetica*, 1984, 33(6): 203 ~ 209.
- 6 Cotterill P P. Number of families and progeny required for provenance testing. *Silvae Genetica*, 1990, 39(2): 82 ~ 84.
- 7 北京林学院主编. 数理统计. 北京: 中国林业出版社, 1980. 131 ~ 134.

Influence of Sample Numbers of Offspring Tested on Genetic Analysis of Quantitative Traits

Qi Ming

Abstract Chosing the 9×9 half diallel cross of Chinese fir plantation at the age 7 as research material, partial data with 8, 16, 24, 32 block repeats randomly were sampled from total data (40 block repeats), respectively. Variance analysis and combining ability analysis were carried out. From the viewpoint of both random model and fixed model, we investigate, in an all-round way and in a systematic way, that offspring sample numbers tested influence on genetic analysis of quantitative traits. The results show that there is a significant effect of sample numbers of offspring tested on genetic analysis, and it varied with research traits. This conclusion tells us in progeny testing and in statistic analysis, each treatment must have a suitable sample numbers of offspring tested, and so, the results of experimental design and statistic analysis are correct and reliable.

Key words sample numbers of offspring tested random model fixed model genetic parameters