

遗传绝对距离差异显著性检验方法

惠刚盈¹, 王韩民², 胡艳波¹

(1. 中国林业科学研究院林业研究所, 北京 100091; 2. 西北工业大学资源与环境信息化工程研究所, 西安 710072)

摘要:Gregorius 提出的用于比较等位基因差异的遗传绝对距离方法可用来比较两个种群或群落的差异,也可比较两个样地是否来自相同的总体。该方法由于能够给出两个分布的具体差异量而在林分直径分布比较中得到广泛应用。在分析 Gregorius 遗传绝对距离应用于群落结构比较的可能性的基础上,提出了该遗传绝对距离的差异显著判别标准,当两个分布的差异 $d_{xy} \geq d_a = d_{\max} \cdot (1 - 1/k) \cdot \sqrt{-0.2\ln(\alpha/2)}$ 时,差异显著。为检验所提出的判别差异显著性方法的有效性,特给出了 40 个模拟林分的直径分布并用 $K-S$ 检验进行了验证,结果表明所提出的方法与 $K-S$ 检验的符合率达 100%。研究还指出,遗传绝对距离方法可有效进行群落结构比较,它既可进行直径分布检验也可对物种的相似性做出判定,自然也可以用于等位基因的分析。结合研究提出的差异显著性检验方法,从而使度量结构差异量和判断此差异量显著与否的林分结构分析与比较一体化。

关键词:遗传绝对距离;临界值;群落结构比较

文章编号:1000-0933(2005)10-2534-06 中图分类号:Q141, S718 文献标识码:A

Significance test of difference for genetic absolute distance

HUI Gang-Ying¹, WANG Han-Min², HU Yan-Bo¹ (1. Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China; 2. Institute of Resource and Environment Information Engineering, Northwestern Polytechnical University, Xian 710072, China). *Acta Ecologica Sinica*, 2005, 25(10): 2534~2539.

Abstract: The “genetic absolute distance”, developed by Gregorius, has been used not only in quantifying the difference between allelic distributions, but also in measuring the difference between the diameter distributions or other attributes of two forests. Now it can be used in analyses of the difference between two populations or communities. A critical value of the genetic absolute distance that can determine the significance of the difference between two distributions is developed in this paper. To demonstrate the approach, we have used 40 examples. In this study, we propose that the difference between two distributions is significant if $d_{xy} \geq d_a = d_{\max} \cdot (1 - 1/k) \cdot \sqrt{-0.2\ln(\alpha/2)}$, and that the test is comparable to a $K-S$ test. This method may be used to advantage for comparing forest community structures. The Genetic Absolute Distance parameter offers the advantage can be used both for quantifying the difference among populations and for significance testing. It can be used to express the difference in terms that are easy to understand and it can also be used as a test criterion.

Key words: genetic absolute distance; critical value; comparison between community structures

森林的结构决定了它的功能。实现森林可持续经营的关键在于明了森林的结构。从结构入手能够深入了解森林生态系统的组织形式和运作机制,恰当地描述森林的结构是破译森林之谜的基础^[1],对森林结构的比较分析方法是制定合理的森林经营方案的有效手段。因此,多年来人们一直致力于群落结构比较方法的研究。出现了群落相似性系数(similarity of community)^[2,3]、 $K-S$ (Kolmogoroff—Smirnoff)检验^[4]和遗传绝对距离(Genetic distance)^[5~8]等方法。其中的群落相似性系数在植被生态学中用来比较两个群落的种类异同,无法进行种群中个体如分布的比较; $K-S$ 检验在生物统计中得到广泛应用,它能对不同形式的分布函数的差异做出十分有效的比较,但难以给出具体的差异量,也无法对物种的相似性做出判定。遗传绝对距离方法在数量遗传

基金项目:德国洪堡基金资助项目

收稿日期:2004-07-20;修订日期:2005-02-10

作者简介:惠刚盈(1961~),男,博士,研究员,主要从事天然林经营模拟研究。E-mail:hui@forestry.ac.cn

Foundation item:Supported by Alexander von Humboldt Foundation

Received date:2004-07-20;Accepted date:2005-02-10

Biography:HUI Gang-Ying, Ph. D., Professor, mainly engaged in natural forest management and simulation. E-mail:hui@forestry.ac.cn

传学中得以发展应用,它能给出两个种群分布的具体差异量,并能对物种的频次分布差异进行对比。但它目前还没有明确指出差异显著与否的判断标准即差异多少可称为显著。可见,遗传绝对距离方法还有待进一步完善。本研究的目的在于给出遗传绝对距离差异显著性的判别方法。

1 遗传绝对距离

Gregorius 提出了一种比较等位基因差异的遗传绝对距离方法,该方法可用来比较两个种群或群落的差异,也可比较两个样地是否来自相同的总体。遗传绝对距离公式如下:

$$d_{xy} = \frac{1}{2} \sum_i^k |x_i - y_i| \tag{1}$$

式中, x_i 为群落 X 中遗传类型 i 的相对频率; y_i 为群落 Y 中遗传类型 i 的相对频率; k 为遗传类型的数量; $\sum_1^k x_i = 1$ 和 $\sum_1^k y_i = 1$ 。

遗传绝对距离的重要特性: d 是一个非负实数。距离对称,也就是说, $d_{xy} = d_{yx}$; 如果 X 和 Y 的遗传结构完全一致,则 $d_{xy} = 0$; 也就是说有共同的等位基因或具有相同的频率; 如果 X 和 Y 的遗传结构完全不一致,则 $d_{xy} = 1$; d 满足三角不等式 $X - Y$ 的群落遗传距离受 $X - Z$ 、 $Y - Z$ 群落遗传距离之和的限制(Z 为第 3 群落)即:

$$d_{xy} \leq d_{xz} + d_{yz} \tag{2}$$

2 遗传绝对距离临界值的确定

按照遗传距离特性(2 式)有:

$$d_{xy} < 2 \cdot \max(d_{xz}, d_{yz}) \tag{3}$$

$$\text{令} \quad \max(d_{xz}, d_{yz}) = d_{\max} \tag{4}$$

$$(3) \text{式可以写成} \quad d_{xy} < 2 \cdot d_{\max} \tag{5}$$

可见, d_{xy} 永远小于 $X - Z$ 、 $Y - Z$ 群落遗传距离中最大者的 2 倍。显然 $d_{\max}$ 的大小与 Z 群落有关。因此,有必要首先给出 Z 群落的定义。

从种群生态学得知,种群中的某个类型占有的频率非常低时,该种群就面临最大的遗传亏损危险。相反,各类型的频率相等即种群类型为均匀分布时,则该种群具有最低可能的遗传亏损。所以,种群生态学中将种类是否接近均匀分布作为评价种群生态稳定性的标准。据此,可将 Z 群落定义为均匀分布。均匀分布时各类型占的相对频率皆为 $1/k$ 。可见, $d_{\max}$ 与 k 的大小有关。

由(5)式出发, d_{xy} 的值必然在 $0 \sim 2d_{\max}$ 之间。根据中心极限定理,多数随机变量的频率分布将近似正态^[4,10]。正态分布为统计学中绝大多数检验方法提供了理论基础^[4]。所以,在此假设, d_{xy} 的频率分布也近似正态(如图 1)。

由图 1 可见,临界值(d_a)与 $d_{\max}$ 有关。由于 $d_{\max}$ 与 k 的大小有关,所以,临界值(d_a)也必然与 k 有关。 k 直接由 X 或 Y 群落的等级数(遗传类型)所决定。而 $k-1$ 正好表达的是卡方(X^2)检验的自由度,按照卡方检验原理,自由度愈大,临界值就愈大。所以认为 k 与遗传距离临界值成正比。另外根据统计学原理,临界值必然与显著性标准 α 有关,参照 $K-S$ 检验中 $K_{(\alpha)}$ 的基本形式^[4],在此不妨用 $K_{(\alpha)} = \sqrt{-0.2 \ln(\alpha)}$ 来表达。

综上所述,假设 $d_{xy} \geq d_a = d_{\max} \cdot (1 - 1/k) \cdot K_{(\alpha)}$,那么就认为差异是显著的。

为检验上述假设,特以经典的 $K-S$ 检验为标准。 $K-S$ 检验的原理如下:

要检验两次独立的抽样是否来自同一个总体,可应用 Kolmogoroff 和 Smirnoff 提出的方法^[4]。该方法能揭示所有分布类型的差异。计算两个待检验抽样的相对累计曲线(从小到大排序再累加)以及两者之间的最大绝对差异,以该最大绝对差异为检验变量与临界值进行比较。公式为:

$$\hat{D} = \max \left| \left(\frac{\hat{F}_1}{n_1} - \frac{\hat{F}_2}{n_2} \right) \right| \tag{6}$$

式中, $\hat{F}_1$ 、 $\hat{F}_2$ 为两次抽样各自的累计频数; n_1 、 n_2 为两次抽样各自的样本数。

临界值通过下式计算:

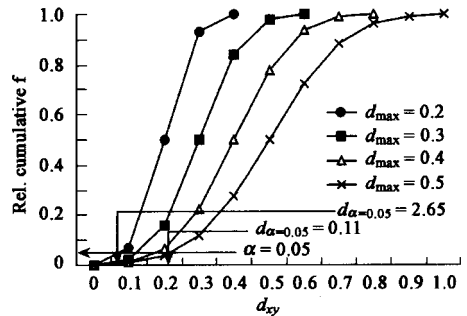


图 1 临界值(d_a)与 $d_{\max}$ 关系

Fig. 1 The relationship between d_a and $d_{\max}$ for 4 values of $d_{\max}$

$$D_{(\alpha=0.05)}=1.36\sqrt{\frac{n_1+n_2}{n_1n_2}}$$

(7)

如果 $\hat{D} \geqslant D_{\alpha}$ 则,差异显著。否则,差异不显著。

现以表 1 的 40 个理论样地的直径分布数据为基础,首先用 K - S 检验进行直径分布差异显著程度的检验,然后再进行遗传绝对距离的计算和检验。检验结果见表 2。

表 1 理论样地(x,y)的直径分布数据

Table 1 Diameter distributions of theoretical sample plots(x,y)

序号 No.	径阶(cm) Diameter	x	y											
			y1	y2	y3	y4	y5	y6	y7	y8	y9	y10	y11	y12
1	2	10	20	10	10	10	40	20	10	10				
	4	20	20	20	10	10	10	40	20					
	6	40	40	10	20	40	20	40	20	20				
	8	20	10	20	40	20	20	20	20	10				
	10	10	10	40	10	20	10	10	10	40				
2	2	10	15	10	50	50	15	10	15	10	15	5		
	4	15	10	50	10	15	50	10	10	10	50	20		
	6	50	50	15	15	10	10	50	50	15	15	50		
	8	15	15	15	15	15	15	15	10	15	10	20		
	10	10	10	10	10	10	10	15	15	50	10	5		
3	2	5	5	5	5	10	50	20	30	20	20	5	0	5
	4	10	10	5	10	5	20	50	20	30	20	10	15	0
	6	15	20	20	15	15	10	10	20	20	10	50	20	20
	8	50	40	30	15	15	10	10	10	10	30	15	40	30
	10	15	20	20	50	40	5	5	10	10	10	15	15	30
	12	5	5	20	5	15	5	5	10	10	10	5	10	15
4	2	1	5	4	5	4	1	1	2	1	1	4		
	4	5	10	5	10	10	4	3	3	4	5	5		
	6	10	10	10	20	25	5	5	5	5	5	5		
	8	15	15	20	25	20	10	7	5	5	10	10		
	10	20	20	25	15	15	15	10	10	5	25	10		
	12	25	25	15	10	10	25	20	15	10	20	20		
	14	10	5	10	5	5	20	25	20	15	15	25		
	16	5	5	5	5	5	10	15	25	20	10	15		
	18	5	4	5	4	5	5	10	10	25	5	5		
	20	4	1	1	1	1	5	4	5	10	4	1		

由表 2 可见,上述判断遗传绝对距离差异显著与否的方法和传统的 K - S 检验结论完全相同。40 个样例的检验结果中,两种方法没有出现 1 例矛盾结论。符合率为 100%。可见,按上述方法进行检验是可行的,即若:

$$d_{xy} \geqslant d_{\alpha} = d_{\max} \cdot (1 - 1/k) \cdot \sqrt{-0.2 \cdot \ln(\alpha)}$$

则认为差异是显著的。

这样就可以直接用遗传绝对距离法同时进行差异多少计算和是否显著的检验。

3 遗传绝对距离法的应用实例

将上述遗传绝对距离法应用于我国东北阔叶红松林的两块试验地的林木直径分布(图 2)比较和树种(图 3)相似性分析。计算遗传绝对距离并按本文提到的检验方法进行差异显著性分析(表 3、表 4)。

由表 3 可知,两个样地的直径分布差异为 22.6%,但由于 $d_{xy} < d_{\alpha}$,故认为两个样地的直径分布未达到差异显著水平。 K - S 检验($\hat{D}=0.160 < D_{\alpha}=0.208$)结果进一步证实了这种判断。

由表 4 可知,两个样地的树种频次分布相差 44%,由于 $d_{xy} > d_{\alpha}=0.374$,故认为这个差异是显著的。

表 2 K -S 检验与遗传绝对距离计算结果

Table 2 The result of K -S test and genetic absolute distance test

序号 No.	K -S				遗传绝对距离 Genetic absolute distance			
	最大差异 D_n	临界值 D_a	注释 Note		x - y 的遗传绝对距离 $d_{x,y}$	最大遗传距离 $d_{\max}$	临界值 d_a	注释 Note
1	x - y 1	0.1	0.19	x 、 y 1 差异不显著 *	0.10	0.20	0.14	x 、 y 1 差异不显著
	x - y 2	0.3	0.19	x 、 y 2 差异显著 * *	0.30	0.20	0.14	x 、 y 2 差异显著
	x - y 3	0.2	0.19	x 、 y 3 差异显著	0.20	0.20	0.14	x 、 y 3 差异显著
	x - y 4	0.1	0.19	x 、 y 4 差异不显著	0.10	0.20	0.14	x 、 y 4 差异不显著
	x - y 5	0.3	0.19	x 、 y 5 差异显著	0.30	0.20	0.14	x 、 y 5 差异显著
	x - y 6	0.1	0.19	x 、 y 6 差异不显著	0.10	0.20	0.14	x 、 y 6 差异不显著
	x - y 7	0.2	0.19	x 、 y 7 差异显著	0.20	0.20	0.14	x 、 y 7 差异显著
	x - y 8	0.3	0.19	x 、 y 8 差异显著	0.30	0.20	0.14	x 、 y 8 差异显著
2	x - y 1	0.05	0.19	x 、 y 1 差异不显著	0.05	0.30	0.21	x 、 y 1 差异不显著
	x - y 2	0.35	0.19	x 、 y 2 差异显著	0.35	0.30	0.21	x 、 y 2 差异显著
	x - y 3	0.40	0.19	x 、 y 3 差异显著	0.40	0.30	0.21	x 、 y 3 差异显著
	x - y 4	0.40	0.19	x 、 y 4 差异显著	0.40	0.30	0.21	x 、 y 4 差异显著
	x - y 5	0.40	0.19	x 、 y 5 差异显著	0.40	0.30	0.21	x 、 y 5 差异显著
	x - y 6	0.05	0.19	x 、 y 6 差异不显著	0.05	0.30	0.21	x 、 y 6 差异不显著
	x - y 7	0.05	0.19	x 、 y 7 差异不显著	0.10	0.30	0.21	x 、 y 7 差异不显著
	x - y 8	0.40	0.19	x 、 y 6 差异显著	0.40	0.30	0.21	x 、 y 8 差异显著
	x - y 9	0.40	0.19	x 、 y 7 差异显著	0.40	0.30	0.21	x 、 y 9 差异显著
	x - y 10	0.05	0.19	x 、 y 8 差异不显著	0.10	0.30	0.21	x 、 y 10 差异不显著
3	x - y 1	0.05	0.19	x 、 y 1 差异不显著	0.10	0.33	0.24	x 、 y 1 差异不显著
	x - y 2	0.20	0.19	x 、 y 2 差异显著	0.25	0.33	0.24	x 、 y 2 差异显著
	x - y 3	0.35	0.19	x 、 y 3 差异显著	0.35	0.33	0.24	x 、 y 3 差异显著
	x - y 4	0.35	0.19	x 、 y 4 差异显著	0.40	0.33	0.24	x 、 y 4 差异显著
	x - y 5	0.55	0.19	x 、 y 5 差异显著	0.55	0.37	0.26	x 、 y 5 差异显著
	x - y 6	0.55	0.19	x 、 y 6 差异显著	0.55	0.37	0.26	x 、 y 6 差异显著
	x - y 7	0.40	0.19	x 、 y 7 差异显著	0.45	0.33	0.24	x 、 y 7 差异显著
	x - y 8	0.40	0.19	x 、 y 8 差异显著	0.45	0.33	0.24	x 、 y 8 差异显著
	x - y 9	0.25	0.19	x 、 y 9 差异显著	0.30	0.33	0.24	x 、 y 9 差异显著
	x - y 10	0.35	0.19	x 、 y 10 差异显著	0.35	0.33	0.24	x 、 y 10 差异显著
	x - y 11	0.05	0.19	x 、 y 11 差异不显著	0.15	0.33	0.24	x 、 y 11 差异不显著
	x - y 12	0.25	0.19	x 、 y 12 差异显著	0.30	0.33	0.24	x 、 y 12 差异显著
4	x - y 1	0.09	0.19	x 、 y 1 差异不显著	0.09	0.30	0.23	x 、 y 1 差异不显著
	x - y 2	0.13	0.19	x 、 y 2 差异不显著	0.13	0.30	0.23	x 、 y 2 差异不显著
	x - y 3	0.29	0.19	x 、 y 3 差异显著	0.29	0.30	0.23	x 、 y 3 差异显著
	x - y 4	0.28	0.19	x 、 y 4 差异显著	0.28	0.30	0.23	x 、 y 4 差异显著
	x - y 5	0.16	0.19	x 、 y 5 差异不显著	0.16	0.30	0.23	x 、 y 5 差异不显著
	x - y 6	0.30	0.19	x 、 y 6 差异显著	0.30	0.30	0.23	x 、 y 6 差异显著
	x - y 7	0.36	0.19	x 、 y 7 差异显著	0.37	0.30	0.23	x 、 y 7 差异显著
	x - y 8	0.46	0.19	x 、 y 8 差异显著	0.46	0.30	0.23	x 、 y 8 差异显著
	x - y 9	0.10	0.19	x 、 y 9 差异不显著	0.15	0.30	0.23	x 、 y 9 差异不显著
	x - y 10	0.22	0.19	x 、 y 10 差异显著	0.28	0.30	0.23	x 、 y 10 差异显著

* no-significantly deference; * * Significantly deference

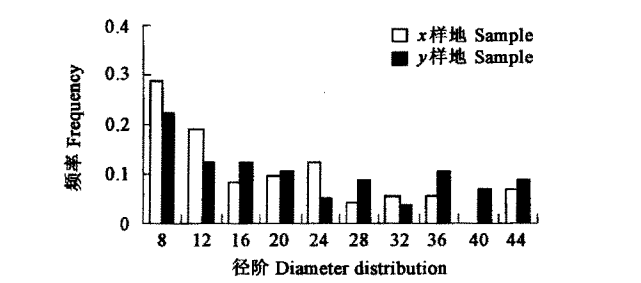


图 2 两块样地直径分布

Fig. 2 Diameter distribution in two sample plots

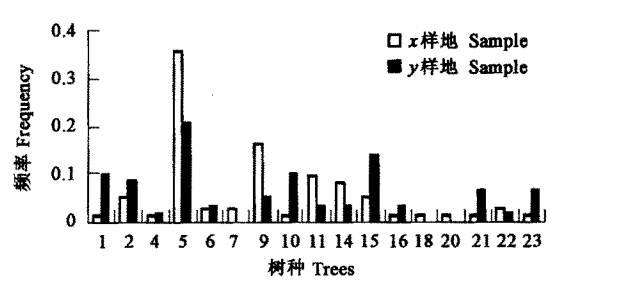


图 3 两块样地树种频次分布

Fig. 3 The species frequency distribution in two samples

表 3 两块阔叶红松林试验地的林木直径分布遗传绝对距离计算												
Table 3 Genetic absolute distance test of diameter distribution in two sample plots in mixed forest of broad-leaved and Korean pine												
径级 (cm) Diameter class	x y		x 频率分布 Frequency distribution	y 频率分布 Frequency distribution	频率 x-y Absolute value of frequency difference between (x,y)	x-y 的遗传绝对距离 d _{xy}	频率 x-z Absolute value of frequency difference between (x,z)	x-z 的遗传绝对距离 d _{xz}	频率 y-z Absolute value of frequency difference between (y,z)	y-z 的遗传绝对距离 d _{xy}	最大遗传距离 d _{max}	临界值 d _a
8	21	13	0.288	0.224	0.064	0.226	0.188	0.303	0.124	0.172	0.303	0.234
12	14	7	0.192	0.121	0.071		0.092		0.021			
16	6	7	0.082	0.121	0.038		0.018		0.021			
20	7	6	0.096	0.103	0.008		0.004		0.003			
24	9	3	0.123	0.052	0.072		0.023		0.048			
28	3	5	0.041	0.086	0.045		0.059		0.014			
32	4	2	0.055	0.034	0.020		0.045		0.066			
36	4	6	0.055	0.103	0.049		0.045		0.003			
40	0	4	0.000	0.069	0.069		0.100		0.031			
44	5	5	0.068	0.086	0.018		0.032		0.014			

表 4 两块阔叶红松林试验地的树种频次分布遗传绝对距离计算													
Table 4 The genetic absolute distance test of species frequency distribution in two sample plots in mixed forest of broad-leaved and Korean pine													
树种 Species	代号 No.	x y		x 频率分布 Frequency distribution	y 频率分布 Frequency distribution	频率 x-y Absolute value of frequency difference between (x,y)	x-y 的遗传绝对距离 d _{xy}	频率 x-z Absolute value of frequency difference between (x,z)	x-z 的遗传绝对距离 d _{xz}	频率 y-z Absolute value of frequency difference between (y,z)	y-z 的遗传绝对距离 d _{xy}	最大遗传距离 d _{max}	临界值 d _a
杉松 <i>Abies</i>	1	1	6	0.014	0.103	0.090	0.436	0.045	0.463	0.045	0.364	0.463	0.374
臭冷杉 <i>Abies</i>	2	4	5	0.055	0.086	0.031		0.004		0.027			
白扭槭 <i>Acer</i>	4	1	1	0.014	0.017	0.004		0.045		0.042			
色木槭 <i>Acer mono</i>	5	26	12	0.356	0.207	0.149		0.297		0.148			
青楷子 <i>Acer tegmentosum</i>	6	2	2	0.027	0.034	0.007		0.031		0.024			
白桦 <i>Bebula platyphylla</i>	7	2	0	0.027	0.000	0.027		0.031		0.059			
千金榆 <i>Carpinus</i>	9	12	3	0.164	0.052	0.113		0.106		0.007			
水曲柳 <i>Frazinus mandshurica</i>	10	1	6	0.014	0.103	0.090		0.045		0.045			
核桃楸 <i>Juglans mandshurica</i>	11	7	2	0.096	0.034	0.061		0.037		0.024			
鱼鳞云杉 <i>Picea</i>	14	6	2	0.082	0.034	0.048		0.023		0.024			
红松 <i>Pinus</i>	15	4	8	0.055	0.138	0.083		0.004		0.079			
香杨 <i>Populus</i>	16	1	2	0.014	0.034	0.021		0.045		0.024			
蒙古栎 <i>Qercus</i>	18	1	0	0.014	0.000	0.014		0.045		0.059			
水榆 <i>Sorbus</i>	20	1	0	0.014	0.000	0.014		0.045		0.059			
暴马丁香 <i>Syringa reticulata</i> var.	21	1	4	0.014	0.069	0.055		0.045		0.010			
紫椴 <i>Tilia</i>	22	2	1	0.027	0.017	0.010		0.031		0.042			
裂叶榆 <i>Ulmus</i>	23	1	4	0.014	0.069	0.055		0.045		0.010			

4 结论与讨论

Gregorius 提出的遗传绝对距离能给出两个分布具体的差异量,同时显示了差异的来源,易于对差异的理解,也便于对差异的调控。但它没有明确指出判断差异显著与否的标准即差异百分之几称为显著。结合本研究提出的判断遗传绝对距离差异显著与否的方法,即若: $d_{xy} \geq d_{\alpha} = d_{\max} \cdot (1 - 1/k) \cdot \sqrt{-0.2 \cdot \ln(\alpha)}$, 就认为差异是显著的。用这种方法实现了各种分布类型差异的计算及其显著性检验。弥补了现有群落结构比较方法的缺陷。用遗传绝对距离法既可以进行直径分布检验也可以对物种的相似性做出判定,自然也可以用于等位基因的分析。它将在群落结构研究中发挥重要作用。

References:

[1] Gadow K V, Hui G Y. Analysis of Forest Structure and Diversity——based on neighbourhood relations. Vortrag, IUFRO International Conference 2002 in Lisbon Portugal, 2002.

[2] Song Y C. *Vegetation ecology*. Shanghai: East China Normal University Press, 2001.

[3] Yang H X. Discussion on similarity coefficient. *Natural Resource. J.*, 1980. 1.

[4] Sachs L. *Angewandte Statistik*. Berlin: Springer-Verlag,1978. 545.

[5] Bergmann F. Genetischer abstand zwischen populationen. *Silvae Genetica*,1974,**23**(1-3).

[6] Gregorius H R. A Unique Genetic Distance. *Biom. J.*, 1984,**26**(1):13~18.

[7] Gregorius H R. Genetischer Abstand zwischen Populationen. Zur Konzeption der Genetischen Abstandsmessung. *Silvae Genetica*,1974,**23**:22~27.

[8] Pommerening A. Eine Analyse neuer Ansätze zur Bestandesinventur in strukturreichen Wäldern. Diss. Fakultät für Forstwissenschaften und Waldökologie. Univ. Göttingen. Cuvillier Verl. Göttingen, 1997. 1870.

[9] Zar J H. *Biostatistical Analysis*. Prentice Hall, Englewood Cliffs,1984. 718.

参考文献:

[2] 宋永昌. 植被生态学. 上海:华东师范大学出版社,2001.

[3] 阳含熙. 相似系数探讨. 自然资源, 1980,(1).