第 22 卷第 6 期
2002 年 6 月

生 态 学 报
ACTA ECOLOGICA SINICA

Vol. 22, No. 6
Jun., 2002

分子生态学重要概念——遗传距离及其测度的理论研究概况

张爱兵，王正军，谭声江，李典谟*

（中国科学院动物研究所，农业虫鼠害综合治理研究国家重点实验室，北京 100080）

摘要：综述了遗传距离的概念、背景，有关遗传距离的几种基本的突变模型以及和遗传距离有关的参量和几种常用统计量，指出在处理蛋白质数据、分子数据以及序列数据时，如何选择相应的统计量和可用的软件包，同时还着重指明了各种模型的假设前提，为处理实际的蛋白质或分子数据时选择合适的模型，和对数据的最终解释提供一些帮助。

关键词：分子生态；遗传距离；测度；突变模型

The Important Concept of Molecular Ecology——Genetic Distance and Its Measuring

ZHANG Ai-Bing, WANG Zheng-Jun, TAN Sheng-Jiang, LI Dian-Mo (*The State Key Laboratory of Integrated Management of Insect Pests and Rodents, Institute of Zoology, Chinese Academy of Sciences*).
***Acta Ecologica Sinica*, 2002, 22(6): 943～949.**

Abstract: Genetic distance is the extent of gene differences between populations or species that is measured by some numerical quantity. The statistical and genetic properties of several typical distance measures will be classified into two groups, i. e., measures for population classification and measures for evolutionary study. Several basic mutation models, such as the infinite alleles mutation model, the stepwise mutation model, the mutation in nucleotide sequences model and so on, together with some parameters and their statistics relating to genetic distance, are reviewed too. Authors suggest that suitable models and software packages should be chosen before analyses of all types of molecular data.

Key words: molecular ecology; genetic distance; measuring; mutation model

文章编号：1000-0933(2002)06-0943-07 **中图分类号**：Q75，Q141，Q346 **文献标识码**：A

1992 年《分子生态学》(Molecular Ecology)杂志的创刊，标志着分子生态学这门学科的正式诞生[1]。分子生态学作为宏观生物学与微观生物学的交叉学科，是当前国际生物学的研究热点之一[2～6]。国内这方面的研究也方兴未艾，如颜廷芬等[7]用 RAPD (Random amplified polymorphic DNA)方法研究了多年生药用草本植物高山红景天 (*Rhodiola sachaliensis*)多态位点在各居群中的分化；高亦珂等[8]用 RAPD 方法分析了东北地区天然白桦(*Betula platyphylla*)不同地理种群的遗传结构；张民照等[9]对蝗虫的研究，等等。然而对于分子生态学中一个重要概念——遗传距离及其相关的突变模型却鲜见系统报道，尽管这些概念在国内一些经典的遗传学教科书[10,11]中有些介绍；另一方面由于许多新的分子标记，如微卫星分子标记等的出现，对于遗传距离的测度也会产生相应的改变[12～15]。因而对遗传距离及其测度理论的系统性了解，将有助于对分子数据进行正确地统计分析，从而最大限度地从分子实验数据获取有用的信息，加深生态学

基金项目："瑞典国际基金"IFS 资助项目(D/3204-1)；国家重点基础发展规划资助项目(G：2000016210)；中国科学院创新方向资助项目(KSCX2-1-02)。

收稿日期：2001-07-06；修订日期：2001-10-22

作者简介：张爱兵(1974～)，男，江苏宝应人，博士。主要从事系统生态，分子生态学研究。

* 通讯联系人 Author for correspondence.

万方数据

者对诸如种群的基因流、迁移动态等生态学过程的理解。

1 遗传距离概念

遗传距离最初是用来估计不同种群之间遗传分化程度的一个指标。遗传距离的概念最早见于 Sanghvi[16]一篇关于进化的文献，但是类似的用于测度种群差异的概念则可追溯到 Czekanowski[17]和 Pearson[18]的工作。关于遗传距离的一般假定是：遗传距离是起源于共同祖先的相同基因进化趋异的一种测度。在这个假定下，遗传距离的理想测度应能够表达这样的含义：基因之间的差异与它们起源于共同祖先的时间成比例。Nei[19]在其经典著作《分子进化遗传学》中，指出："遗传距离是指不同的种群或种之间的基因差异的程度，并且以某种数值进行度量"。因此，本质上，遗传差异的任何数值测度，只要是在序列水平或是基因频率水平上，由不同个体、种群或种的数据计算而来，皆可定义为遗传距离。但是 Nei[19]接着又指出由于历史的缘故，遗传距离通常是指由基因频率的某个函数所确定的基因差异。

2 基本的突变模型

2.1 无限基因突变模型

该模型[20]假定每次突变产生一种新的当前种群不存在的等位基因。这就意味着如果两个等位基因相同，则没有突变发生；如果两个等位基因不同，则至少发生一次突变，但不清楚这种情况下突变发生的具体次数。如果把两个等位基因相同记为事件 0 发生，两个等位基因不同记为事件 1 发生，突变率为 μ，则突变的分布服从泊松（Poisson）分布，t 代前具有共同祖先的两个基因发生突变的平均数目为 $2\mu t$，因而对于无限基因突变模型，t 代之前具有共同祖先的两个等位基因相同的概率服从平均数为 $2\mu t$ 的泊松分布：

$$2ut : p(0) = e^{-2ut}$$

2.2 逐步突变模型

逐步突变模型最初是针对不同蛋白同功酶电荷差异的分布而设计的[21]。蛋白质的电泳迁移基本上取决于蛋白质所带的总的净电荷。如果带正电荷的氨基酸被一个电中性的氨基酸替代，则蛋白质的泳动能力就会发生变化。该模型假定蛋白质的电泳迁移仅由其所带总电荷确定，并且根据所带电荷表达一个基因的等位基因状态（Allelic state），则突变的发生是逐步的，即每次基因突变只改变蛋白质一个单位的电荷，而且向电荷增加和减少方向突变的概率相同，然而，事实上，基因突变对电荷的改变并不总是以逐步的方式发生的，有时突变会使蛋白质所带电荷直接从正电荷变化到负电荷，或者相反。但是这种由突变导致的蛋白质所带电荷的两步变化的概率很小，以致可以忽略不计[19]。简言之，该模型认为突变将或多或少逐步改变同功酶所带电荷，因而在电泳凝胶上迁移相同或类似距离的同功酶应比迁移不同距离的同功酶对应着更少的突变。

这种思想最近又被重新提出运用于微卫星数据的遗传距离的测度[13,22~23]。对于微卫星位点，该模型假定突变以逐步的方式或多或少地改变等位基因（Allele）的长度。

2.3 核苷酸序列的突变模型

最简单的突变模型是碱基的点突变模型，然而每个给定的核苷酸位点只有 4 种可能的碱基，若在一个位点发生多次突变，则无法加以区分，这就导致了进化上的非同源相似（Homoplasy）。因而，与序列变化数目成线性函数关系的遗传距离测度将过低估计序列从共同祖先分化而来的时间。为了解决这个问题，Jukes-Cantor[24]距离模型被提出来，该模型考虑了同一位点的多次突变，但是没有考虑突变概率的差异。在该模型的条件下，每个位点核苷酸替换数目的期望值为 $d=2\lambda t$，且 d 由下式估计：

$$\hat{d} = -\frac{3}{4}\log_e\left(1-\frac{4}{3}P\right)$$

其中：λ 为核苷酸每年的突变率；P 为两个种群不同核苷酸的比例。

d 的方差由下式（Kimura 和 Ohta[25]）给出：

$$V(\hat{d}) = \left[\frac{\mathrm{d}d}{\mathrm{d}p}\right]^2 V(P) = \frac{9P(1-P)}{(3-4P)^2 n}$$

万方数据

其中：

$$V(P)=P(1-P)/n$$

如果 d 值很小或不同位点以等概率进行突变,上式是 d 的一个很好估计。如果上述两条件不成立,则估计是偏低的。

上述两模型隐含了一个假定:不同核苷酸座位上具有相同的突变速率。而这一点,目前看来在许多情况下是与实际的情形不吻合的[26],比如许多脊椎动物、无脊椎动物线粒体的控制区,在进化速率上,有相当的异质性。Kimura[25]提出了 Kimura 两参数模型(Kimura's two-parameter model),该模型基于核苷酸转换速率高于颠换速率的情况,假定核苷酸每位点每年转换、颠换速率是不同的,分别为 α 和 2β,则总的替换速率为 $\alpha+2\beta$,则序列 X 和 Y 之间,每个位点替换数目的期望值为:

$$d = 2\lambda t = 2\alpha t + 4\beta t = -1/2\log_e[1 - 2P - Q_{\text{sequr}}(1 - 2Q)]$$

其中,λ 为突变率;t 为两个序列起源于共同祖先的时间(a);P 为转换频率,Q 为颠换频率。

D 的估计值的方差为:

$$V(\hat{d}) = \frac{1}{n}[a^2P + b^2Q - (aP + bQ)^2]$$

其中:

$$a = \frac{1}{1 - 2P - Q} \qquad b = \frac{1}{2}[\frac{1}{1 - 2P - Q} + \frac{1}{1 - 2Q}]$$

Tamura-Nei Gamma 距离模型则更一般地考虑了不同位点之间突变速率的差异,该模型假定位点之间突变速率服从珈玛随机变量(Gamma random variables)[27]。

3 常用的遗传距离

3.1 几何距离

几何距离作为遗传距离的测度,不依赖于任何种群遗传模型。在这种情况下,群体间的遗传距离可用等位基因频率表示。对于 v 个等位基因来说,两群体间的几何距离为:

$$d_{pQ} = \sqrt{\sum_{u}^{v}(p_u - q_u)^2}$$

其中,v:等位基因个数;P_u:群体 P 的等位基因频率;q_u:群体 q 的等位基因频率。

另一种常用的几何距离是 Cavalli-Sforza 和 Edward 的弦距离(Chord distance),Dc[28]。该距离是把基因频率的平方根看作与等位基因个数同维的超球面上的点:

$$Dc = (2/\pi)[2(1 - \cos\theta)]^{1/2}$$

即为代表不同种群基因频率的两点之间的弦距离。

另一个相近的距离测度是 Nei,等[29]的 D_A。其具体公式为:

$$D_A = \sum_{k=1}^{r}(1 - \sum_{i=1}^{m_k}(\sqrt{x_{ik}y_{ik}}))/r$$

其中,m_k 为第 k 个位点等位基因的数目,r 为所检测位点的个数。D_A 在 0~1 之间取值,当 D_A 较小时,与进化时间近似地成线性函数关系。

Takezaki 和 Nei[15]在用不同距离测度重构种群分化的拓扑(Topology)结构的模拟研究中,认为几何距离在重构种群分枝的拓扑结构时更加可靠,而基于遗传模型的距离测度更易于重构分枝长度。

3.2 赖氏距离(Nei's D)

应用最广泛的基于无限基因突变模型的距离测度是赖氏距离 D[29,30]。其表达式为:

$$D = -\ln(J_{xy}/sqrt(J_x \times J_y)).$$

其中:$J_{xy}=1-Hb$,对种群 x 和种群 y

$J_x=1-Hw$,对种群 x

$J_y=1-Hw$,对种群 y

当祖先种群和衍生种群具有相同的有效种群大小时 Nei's 距离 D 的期望值为[30,31]:$\bar{e}^{2ut}$。

可以看出,在理想的情况下,Nei'D 与 T 成正比,而与种群大小无关。

万方数据

如果基因的替换速率对于所有位点都相同，则该定义非常合适，在这种情况下，D 测度了每个位点基因替换的累积数目；如果基因替换的速率对于所有位点并不相同，则 D 会过低估计每个位点基因替换的累积数目[32]。当基因的替换速率随位点而异，且所有的 Ij 都比较大，则一个更合适的遗传距离的测度由下式给出[32]：

$$D' = -\log_e I'$$

其中，$I_j = j_{XY}/(j_X\, j_Y)^{1/2}$，$j_X$ 为随机取自 X 种群的两个基因相同的概率；j_Y 为随机取自 Y 种群的两个基因相同的概率；j_{XY} 为分别取自 X 种群和 Y 种群的两个基因相同的概率；$I' = J'_{XY}/(J'_X J'_Y)^{1/2}$，$J'_X$，$J'_Y$ 和 J'_{XY} 分别为 j_X，j_Y 和 j_{XY} 的几何平均。

4 *Fst*、类 *Fst*、*Gst*、*Rst*、*Nst* 和 *Φst*

4.1 *Fst*

Wright 的 Fst 也是种群之间遗传差异的一种测度[33]，尽管 Wright[35] 反对 Fst 用作遗传距离，因为 Fst 不满足三角不等关系，如果研究的目的是为了推断基因流的方式，Fst 则是遗传相似性的一个有用测度[36]。对于取自两个种群的样本，Wright[33] 定义了 Fst，Nei[37] 对其进行了推广，可以写为下式：

$$Fst = \frac{f_0 - \bar{f}}{1 - \bar{f}}$$

其中：

$$\bar{f} = \frac{f_1 + f_0}{2}$$

f_1 为取自不同种群的基因血统相同的概率；f_0 为取自同一种群的基因血统相同的概率。

如果种群是新近起源，并且突变率对于所研究的标记来说不是太高，那么近似而言，种群之间的差异可能是随机漂变效应。在这种情况下，一个较好的估计量是 Fst 估计量。Fst 是在没有迁移、侵入发生时，同一亚种群内两个基因起源于该亚种群内同一祖先的概率[35]。

$$Fst = e^{-T/2Ne}$$

Fst 可由 $(fw - fb)/(1 - fb)$ 估计，其中：fw 为取自同一种群两个基因血统相同的概率的估计量；fb 为取自不同种群两个基因血统相同的概率的估计量。

在种群之间进行两两比较时，Fst 的另一个合适的估计量是 Weir 和 Cockerham[38] 的 θ。Weir 和 Cockerham[38] 集中讨论了估计 Fst 的几个统计量，其中大多数统计量依赖于代表基因频率取样的种群数目。统计量 θ 由 Weir 和 Cockerham[38] 定义，其具体公式随所用假设条件、等位基因数目和所考察的位点数而异。Cockerham 和 Weir[39] 讨论了 θ 的其它特性，认为其用作岛屿模型 N_m 的估计量时，其偏差相对较低。

4.2 类 *Fst*

类 Fst 估计量由 $1 - \mathrm{d}w/\mathrm{d}x$ 估计，其中：$\mathrm{d}w$ 为一个种群内所有基因两两之间距离的平均值；

$\mathrm{d}x$ 可以是 $d_{b'}$ 或 $d_{a'}$，其中 $d_{b'}$ 由不同种群两两基因的距离计算而得；$d_{a'}$ 由所有基因两两之间的距离计算而得，而不考虑这些基因来自哪个种群。

基于 $1\text{-}\mathrm{d}w/\mathrm{d}a$ 的类 Fst 统计量具有一个特性，其期望值依赖于样方的数目。

4.3 *Gst*

基于 da 的应用最广泛的类 Fst 估计量是 Nei 的 Gst[30,37]。亚种群间基因分化程度相对于整个种群的基因分化，由 Gst 来测度，Gst 由下式给出：

$$Gst = Dst/H_T$$

其中，Dst 为亚种群之间的平均基因多样度，包含亚种群与其自身的比较；H_T 为整个种群的基因多样度。

该测度依赖于所研究的种群，在一个种群获得的估计值不能与另一个种群的估计值进行比较，除非两个种群的繁育系统类似。Gst 等价于 Wright 的 Fst，可以称之为基因变异系数。当所研究的位点只有两个等位基因时，$Gst = Fst = \delta_x^2/x(1-x)$，其中，$x$ 和 δ_x^2 分别为亚种群中等位基因频率的平均数和方差。当有多个

等位基因时，Gst 相当于 Fst 对所有等位基因的加权平均。

4.4 Rst、Nst 和 Φst

种群间基因多样度相对于种群内基因多样度的比例由 Rst 测度，Rst 由下式给出：

$$Rst = sDst/(s-1)Hs$$

其中，s 为亚种群的个数；Hs 为亚种群内的基因多样度；Dst 为亚种群间基因多样度。

在序列数据的情况下，Fst 的合适的估计量是 Lynch 和 Crease 的 Nst，$Nst=1-\delta_w/\delta_b$[13]。该统计量类似于 Wright 的 Fst，事实上，当取自不同种群的样本数相同且足够小时，$Nst=\theta$。在无限位点、无限基因突变模型时，与 Slatkin 在计算 Fst 期望值的同样的假定下，Nst 的期望值为：$(T/2Ne)/(1+T/2Ne)$。序列数据的情况下，Fst 另一个合适的估计量是 Excoffier 等[40]的 Φst。

5 Sb、Δ 和 ASD

对于逐步突变模型，等位基因长度（对于微卫星长度以重复数计算）或蛋白质电荷的均方差类似于无限基因突变模型中的基因杂合度或基因多样度。对于种群之间的两两比较，Sb[14]、Δ[13]、ASD[13]都可以作为一种距离测度。这些距离与平均共祖时间 成线性函数关系。$\Delta=(\Delta\mu)^2=(\mu_A-\mu_B)^2$，其中 μ_A、μ_B 分别为种群 A 和种群 B 中等位基因大小的平均数。两个种群 A 和 B 之间的平均平方距离 ASD 为 $\sum\sum(i-j)^2FiFj$，其中，Fi、Fj 分别为种群 A 中第 i 个等位基因频率和种群 B 中第 j 个等位基因频率。对于 T 代前从共同祖先分化来的两个种群，在严格的逐步突变模型下[13~14]，ASD 的期望值为：

$$2v\mu(T+2Ne)$$

其中，v 为每一突变步骤长度改变的方差；Ne 为祖先种群的有效大小。

对于较短的分离时间，$(\Delta\mu)^2$ 距离比 ASD 具有更低的变异系数。$(\Delta\mu)^2$ 比较适合于测度亲缘关系较远的类群之间的遗传距离，而标准遗传距离则更适合于测度亲缘关系较近的类群之间的遗传距离[13]。$(\Delta\mu)^2$ 的另一个优点是其依赖于进化时间而与种群大小无关。

6 有关序列数据、RFLP 数据、微卫星数据的软件包

序列分析的第一步是进行序列对比。有关序列对比的算法很多，目前己有许多软件包能够执行其中的诸多算法，这些软件包有些可以共享，有些则是商品化的。Clustal 是一个可共享的执行多重序列对比的软件包，该软件包的最新版本是 Clustal W[42]，可以在 VMS、UNIX、MAC 和 PC 平台上运行。PHYLIP[43]（http://evolution. genetics. washington. edu/phylip. html）和 MEGA[44]是两个利用序列对比计算各种遗传距离的软件包。

对于 RFLP 数据，最基本的分析方法是把各种不同的单倍型（Haplotyple）看作探针杂交位点的等位基因，在此基础上则可进一步计算出各种经典的遗传距离。Biosys[45]、Fstat[46]、Genepop[46]等软件包可以处理这种类型的数据。

微卫星序列很大程度上表现了孟德尔遗传。对于微卫星位点的突变过程，有两种突变模型可以运用：无限基因突变模型和逐步突变模型。在前一种模型的情况下，微卫星变异可以用类似于同功酶的方式进行分析，或者运用赖氏基因多样度（Nei's gene diversity）作为基本的测度。在后一种模型的情况下，以重复单位数或核苷酸数作为测度的等位基因长度的均方差可以进行类似的分析。对于微卫星数据可用的软件包有：Biosys[45]、Fstat[46]、Genepop[12]、GDA[31]（http://biologyool. unm. edu. ～lewis/gda. htm）、Microsat[47]（http://lotka. stanford. edu/distance. html）、WinAmova[40]（http://anthroplogie. unige. ch/ftp/comp/win/amova）、Popgene[48]（http://www. rr. ualberta. ca. /profs/yeh. htm）和 Kinship[49]（http://www. bioc. rice. edu/～kfg/Gsoft. html）。

总之，遗传距离是分子生态学中一个十分重要的概念，准确理解遗传距离的概念和在不同的情况下选择合适的度量方式，对于生态学者理解许多分子生态学过程是非常重要的。

万方数据

参考文献

[1] Zu Y G (祖元刚), Yan T F (颜廷芬), Yu J H (于景华). Formation and development of Molecular Ecology. In: Zu Y G (祖元刚), Sun M (孙梅),Kang L (康乐)ed. *Theory, method and application of Molecular Ecology*(in Chinese). Beijing: China Higher Education Press and Springer-Verlag Heidelberg,1999. 1～16.

[2] Stow A J, Sunnucks P, Briscoe D A, *et al*. The impact of habitat fragmentation on dispersal of Cuningham's skink (*Egernia cunninghami*): evidence from allelic and genotypic analysis of microsatellites. *Molecular Ecology*, 2001, **10**: 867～878.

[3] Jordano P, Godoy J A. RAPD variation and populaion genetic structure in *Prunus mahaleb* (Rosaceae), an animal-disperseal tree. *Mol. Ecol.*, 2000, **9** (9): 1293～1305.

[4] Fisher M, Husi R, Prati D, *et al*. RAPD variation among and within small and large populaions of the rare clonal plant *Ranunculus reptans* (Ranunculaceae). *Am. J. Bot.*, 2000, **87** (8): 1128～1137.

[5] Goossens B, Chikhi L, Taberlet P, *et al*. Microsatellite analysis of genetic variation among and within Alpine marmot populations in the French Alps. *Molecular Ecology*, 2001, **10**: 41～52.

[6] King T L, Kalinows S T, Schill W B, *et al*. Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, 2001, **10**: 807～821.

[7] Yan T F (颜廷芬), Yan X F (阎秀峰), Zhou F J (周福军), *et al*. Research on the distribution and differentiation of RAPD polymorphic fragments for *Rhodiola sachalinensis*. In:Zu Y G (祖元刚), Sun M (孙梅),Kang L (康乐)ed. *Theory, method and application of Molecular Ecology*(in Chinese). Beijing: China Higher Education Press and Springer-Verlag Heidelberg,1999. 167～176.

[8] Gao Y K (高亦珂), Nie S Q(聂绍荃), Zu Y G(祖元刚). Genetic structure analysis by RAPD in *Betula platyphylla* nature population in Northeast of China. In: Zu Y G (祖元刚), Sun M (孙梅),Kang L (康乐)ed. *Theory, method and application of molecular ecology* (in Chinese). Beijing: China Higher Education Press and Springer-Verlag Heidelberg,1999. 196～205.

[9] Zhang M Z(张民照), Kang L(康乐). Extraction of total DNA from locusts and optimization of reaction conditions for RAPD analysis. *Zoological Research*(in Chinese) (动物学研究), 2001, **22** (1): 20～26.

[10] Guo P Z (郭平仲). *Introduction to population genetics*(in Chinese). Beijing: Agriculture Press,1993. 366～371.

[11] Bruce S Weir ed. Xu Y B(徐云碧), Wang Z N(王志宁), Yu Z H(俞志华) translated. *Analysis of genetics data——analysis methods of discrete data of population genetics*(in Chinese). Beijing: Agriculture Press, 1996. 162～170.

[12] Goldstein D B, Linares A R, Cavalli-Sforza L L, *et al*. An evaluation of genetic distances for use with microsatellite loci. *Genetics*, 1995, **139**: 463～471.

[13] Goldstein D B, Linares A R, Cavalli-Sforza L L, *et al*. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA*, 1995, **92**: 6723～6727.

[14] Slatkin M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, 1995, **139**: 457～462.

[15] Takezaki N and Nei M. Genetic distances and reconstruction of plylogenetic trees from microsatellite DNA. *Genetics*, 1996, **144**: 389～399.

[16] Sanghvi L D. Comparison of genetical and morphological methods for a study of biological differences. *Amer. J. Phys. Anthrop.*, 1953, **11**: 385～404.

[17] Czekanowski J. Zur differentialaldiagnose der neandertalgruppe, *Korrespondenzblatt Deutsch. Ges. Anthropol. Ethnol. Urgesch.*, 1909, **40**: 44～47.

[18] Pearson K. On the cofficient of racial likeness. *Biometrika*, 1926, **18**: 337～343.

[19] Nei M. *Molecular evolutionary genetics*. New York : Columbia University Press,1987.

[20] Kimura M and Crow J F. The number of alleles that can be maintained in a finite population. *Genetics*, 1964, **49**: 725～738.

[21] Ohta T and Kimura M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, 1973, **22**: 201～204.

[22] Shriver M D, Jin L, Boerwinkle E, *et al*. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Molecular Biology and Evolution*, 1995, **12**: 914～920.

[23] Michalakis Y and Excoffier L. A genetic estimation of population subdivision using distances between alleles with

万方数据

special reference for microsatellite loci. *Genetics*, 1996, **142**: 1061～1064.
[24] Jukes T H and Cantor C R. Evolution of protein molecules. In: Munro H N, ed. *Mammalian Protein Metabolism*, New York: Academic Press, 1969. 21～132.
[25] Kimura M and Ohta T. On the stochastic model for estimation of mutational distance between homologous proteins. *J. Mol. Evol.*, 1972, **2**: 87～90.
[26] Zhang D X and Hewitt G M. The use of DNA markers in population genetics and ecological studies of the Desert Locust *Schistocera gregaria* (Orthoptera: Acridedae). *The Ecologly of Agricultural Pests*, Edited by W. O. C. Symondson and J. E. Liddell. Published in 1996 by Chapman & Hall, London, 1996. 213～229.
[27] Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 1980, **16**: 111～120.
[28] Cavalli-Sforza L L and Edwards A W F. Phylogenetic analysis: Models and estimation procedures. *Amer. J. Hum. Genet.*, 1967, **19**: 233～257.
[29] Nei M, Tajima F and Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.*, 1983, **19**: 153～170.
[30] Nei M. Genetic distance between populations. *Amer. Natur.*, 1972, **106**: 283～292.
[31] Weir B S. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates Inc., Sanderland, Massachusetts, 1996.
[32] Nei M. Interspecific gene differences and evolutionary time estimated from electrophoretic data on protein identity. *Amer. Natur.*, 1971, **105**: 385～398.
[33] Wright S. The genetical structure of populations. *Ann. Eugen.*, 1951, **15**: 323～354.
[34] Reynolds J, Weir B S and Cockerham C C. Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, 1983, **105**: 767～779.
[35] Wright S. Evolution and the genetics of population Vol 4. Variablility whthin and among natrual populations. University of Chicago Press, Chicago III. 1978.
[36] Slatkin M. Inbreeding coefficients and coalescence times. *Genet. Res.*, 1991, **58**: 167～175.
[37] Nei M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, 1973, **70**: 3321～3323.
[38] Weir B S and Cockerham C C. Estimating F-statistics for the analysis of population structure. *Evolution*, 1984, **38**: 1358～1370.
[39] Cockerham C C and Weir B S. Estimation of gene flow from F-statistics. *Evolution*, 1993, **47**: 855～863.
[40] Lynch M and Crease T J. The analysis of population survey data on DNA sequence variation. *Mol. Biol. Evol.*, 1990, **7**: 377～394.
[41] Excoffier L. AMOVA 1.55 (Analysis of Molecular Variance) University of Geneva. 1995.
[42] Thompson J D, Higgins D G and Gibson T J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 1994, **22**: 4673～4680.
[43] Felsenstein J. PHYLIP (Phylogeny Inference Package) version 3.5. University of Washington, Seattle. 1991.
[44] Sudhir K, Tamura K and Nei M. "MEGA: Molecular Evolutionary Genetic Analysis, Version 1.01. The Pennsylvania State University, University Park, PA16802. 1993.
[45] Swofford D L and Selander R B. BIOSYS-1: a FORTRAN program for the comprehensive analysis of electrophretic data in population genetics and systematics. *Journal of Heredity*, 1981, **72**: 281～283.
[46] Goudet J. FSTAT version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, 1995, **86**: 485～486.
[47] Raymond M and Rousset F. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 1995, **86**: 248～249.
[48] Yeh F Cand Boyle T. POPGENE version 1.2. Microsoft Windows based software for population genetic analysis. University of Alberta, Department of Renewable Resources. Edmonton, Canada. 1997.
[49] Goodnight K E, Queller D C and Poznansky T. KINSHIP 1.1.2. Department of Ecology and Evolutionary Biology, Rice University, Texas, USA. 1997.

万方数据