第 30 卷　第 5 期　　西北农林科技大学学报(自然科学版)　　Vol. 30　No. 5
2002 年 10 月　　Jour. of Northwest Sci-Tech Univ. of Agri. and For. (Nat. Sci. Ed.)　　Oct. 2002

性连锁群体平衡的信息模型研究*

刘建军, 郭满才, 解小莉, 张宏礼, 周静芋, 袁志发

(西北农林科技大学 生命科学学院, 陕西 杨陵 712100)

[**摘　要**]　以 XY 型性染色体为模型, 运用信息论方法研究了性连锁群体平衡的熵变规律。结果表明, 在随机交配下, 性连锁平衡群体的熵最大; 性连锁不平衡群体, 其平衡过程是信息熵增大的过程, 平衡时信息熵达到最大。

[**关键词**]　性连锁基因; 平衡群体; 非平衡群体; 配子; 熵

[**中图分类号**]　S11⁺4　　[**文献标识码**]　A　　[**文章编号**] 1671-9387(2002)05-0114-03

位于性染色体上的基因被称为性连锁基因。在性连锁群体中, 雌性为同型配子 XX, 雄性为异型配子 XY(或 XO)。用信息论方法研究群体平衡问题, 国内外学者[1~5]做了一些工作, 然而对性连锁群体的信息分析尚未见报道。本研究将建立性连锁群体的 Shannon 信息量模型, 并进行讨论分析。

1　性连锁群体的平衡与信息熵

为方便讨论, 视性连锁群体为雄性和雌性 2 个群体(信源)

$$G(♀):\begin{bmatrix} AA & Aa & aA & aa \\ p_{11} & p_{12} & p_{13} & p_{14} \end{bmatrix} \text{和} G(♂):\begin{bmatrix} A & a \\ p_{15} & p_{16} \end{bmatrix} \quad (1)$$

其中, Aa 与 aA 为正、反交。$p_{11}+p_{12}+p_{13}+p_{14}=1$, $p_{15}+p_{16}=1$。雌、雄群体合在一起考虑, 其基因信源为

$$A(P):\begin{bmatrix} A & a \\ p & q \end{bmatrix} \quad (2)$$

$$p=\frac{1}{3}(2p_{11}+p_{12}+p_{15}),$$

$$q=\frac{1}{3}(2p_{14}+p_{12}+p_{13}+p_{16}), p+q=1 \quad (3)$$

据 Shannon 信息熵的定义, G(♀), G(♂) 和 A(P) 的信息熵分别为

$$S[G(♀)]=-(p_{11}\ln p_{11}+p_{12}\ln p_{12}+p_{13}\ln p_{13}+p_{14}\ln p_{14}) \quad (4)$$

$$S[G(♂)]=-(p_{15}\ln p_{15}+p_{16}\ln p_{16}) \quad (5)$$

$$S[A(P)]=-(p\ln p+q\ln q) \quad (6)$$

性质 1　当 $[S[G(♀)]+S[G(♂)]]_{max}=3S[A(P)]$ 时, 性连锁群体(1)达到平衡。即 $p_{11}=p^2$, $p_{12}=p_{13}=pq$, $p_{14}=q^2$, $p_{15}=p$, $p_{16}=q$。

证明: 据最大熵原理, 性质 1 的证明归结为如下的条件极值问题

$$p_{11}+p_{12}+p_{13}+p_{14}=1 \quad (7)$$

$$p_{15}+p_{16}=1 \quad (8)$$

$$\frac{1}{3}(2p_{11}+p_{12}+p_{13}+p_{15})=p \quad (9)$$

$$\frac{1}{3}(2p_{14}+p_{12}+p_{13}+p_{16})=q \quad (10)$$

$$S[G(♀)]+S[G(♂)]=\max \quad (11)$$

运用拉格朗日(Lagrange)乘数法, 引入乘数 λ_0, λ_1, λ_2 和 λ_3, 构造目标函数

$$f(G)=S[G♀]+S[G(♂)]+(\ln\lambda_0+1)(\sum_{j=1}^{4}p_{1j}-1)+(\ln\lambda_1+1)(\sum_{j=5}^{6}p_{1j}-1)+\ln\lambda_2[\frac{1}{3}(2p_{11}+p_{12}+p_{13}+p_{15})-p]+\ln\lambda_3[\frac{1}{3}(2p_{14}+p_{12}+p_{13}+p_{16})-q] \quad (12)$$

令 $\frac{\partial f}{\partial p_{1j}}=0$, $j=1,2,\cdots,6$

得　$p_{11}=\lambda_0\lambda_2^{\frac{2}{3}}$, $p_{12}=\lambda_0\lambda_2^{\frac{1}{3}}\lambda_3^{\frac{1}{3}}$, $p_{13}=\lambda_0\lambda_2^{\frac{1}{3}}\lambda_3^{\frac{1}{3}}$,

$p_{14}=\lambda_0\lambda_3^{\frac{2}{3}}$, $p_{15}=\lambda_1\lambda_2^{\frac{1}{3}}$, $p_{16}=\lambda_1\lambda_3^{\frac{1}{3}}$

代入约束条件, 有 $\lambda_1^2=\lambda_0$, 解之得

$$\lambda_2^{\frac{1}{3}}=\frac{p}{\sqrt{\lambda_0}} \qquad \lambda_3^{\frac{1}{3}}=\frac{q}{\sqrt{\lambda_0}} \quad (13)$$

故 $p_{11}=p^2$, $p_{12}=p_{13}=pq$, $p_{14}=q^2$, $p_{15}=p$, $p_{16}=q$。显然有

* [收稿日期]　2002-04-27
[基金项目]　西北农林科技大学重点科研基金项目(0808)
[作者简介]　刘建军(1964-), 男, 陕西蓝田人, 讲师, 在读硕士, 主要从事农业应用数学研究。

$$S[G(♀)] = -(p^2\ln p^2 + 2bq\ln pq + q^2\ln q^2) = 2S[A(P)]$$

$$S[G(♂)] = -(p\ln p + q\ln q) = S[A(P)]$$

即 $S[G(♀)] + S[G(♂)]_{max} = 3[A(P)]$

设群体(1)平衡, 则由性质 1, G(♀)和 G(♂)的基因信源分别为

$$A(♀):\begin{bmatrix} A & a \\ p & q \end{bmatrix} \text{和} A(♂):\begin{bmatrix} A & a \\ p & q \end{bmatrix} \quad (14)$$

即 $$S[A(♀)] = S[A(♂)] = S[A(P)] \quad (15)$$

性质 2　性连锁群体(1)达到平衡的必要条件是: $S[A(♀)] = S[A(♂)] = S[A(P)]$。这样的群体如未平衡, 那么再进行一代随机交配, 群体就平衡了。

2　平衡的建立及其熵变规律

设起始的连锁群体为(0 代)不平衡群体, 即

$$G_0(♂):\begin{bmatrix} A & a \\ p_m & q_m \end{bmatrix} \text{和} G_0(♀):\begin{bmatrix} AA & Aa & aa \\ D & H & R \end{bmatrix} \quad (16)$$

其中 $p_m + q_m = 1, D + H + R = 1$。它们的基因信源分别为

$$A_0(♂):\begin{bmatrix} A & a \\ p_m & q_m \end{bmatrix} \text{和} A_0(♀):\begin{bmatrix} A & a \\ p_f & q_f \end{bmatrix} \quad (17)$$

其中, $p_f = D + \frac{H}{2}, q_f = R + \frac{H}{2}, p_f + q_f = 1, p_m \neq p_f$。雌、雄群体的平均基因信源为

$$A_0(P):\begin{bmatrix} A & a \\ p & q \end{bmatrix}, p = \frac{2}{3}p_f + \frac{1}{3}p_m,$$

$$q = \frac{2}{3}q_f + \frac{1}{3}q_m \quad (18)$$

各基因信源的信息熵分别为

$$S[A_0(♂)] = -(p_m\ln p_m + q_m\ln q_m) \quad (19)$$

$$S[A_0(♀)] = -(p_f\ln p_f + q_f\ln q_f) \quad (20)$$

$$S[A_0(P)] = -(p\ln p + q\ln q) \quad (21)$$

群体(16)经过一代随机交配, A(♀)配子(X 染色体携带)与 Y 染色体随机结合形成一代雄性群体 $G_1(♂)$; A(♀)配子与雄性 X 染色体所携带的 A(♂)配子随机结合, 形成了一代雌性群体 $G_1(♀)$, 它们分别为

$$G_1(♂):\begin{bmatrix} A & a \\ p_f & q_f \end{bmatrix} \quad (22)$$

$$G_1(♀):\begin{bmatrix} AA & Aa & aA & aa \\ p_m p_f & q_m p_f & p_m q_f & q_m q_f \end{bmatrix} \quad (23)$$

一代的基因信源分别为

$$A_1(♂):\begin{bmatrix} A & a \\ p_f & q_f \end{bmatrix} \quad (24)$$

$$A_1(♀):\begin{bmatrix} A & a \\ \frac{p_m + p_f}{2} & \frac{q_m + q_f}{2} \end{bmatrix} \quad (25)$$

$$A_1(P):\begin{bmatrix} A & a \\ p & q \end{bmatrix} \quad (26)$$

它们的信息熵分别为

$$S[G_1(♂)] = -(p_f\ln p_f + q_f\ln q_f) \quad (27)$$

$$S[G_1(♀)] = -(p_m\ln p_m + q_m\ln q_m + p_f\ln p_f + q_f\ln q_f) \quad (28)$$

$$S[A_1(♂)] = -(p_f\ln p_f + q_f\ln q_f) \quad (29)$$

$$S[A_1(♀)] = -\left(\frac{p_m + p_f}{2}\ln\frac{p_m + p_f}{2} + \frac{q_m + q_f}{2}\ln\frac{q_m + q_f}{2}\right) \quad (30)$$

$$S[A_1(P)] = -(p\ln p + q\ln q) \quad (31)$$

式(27)～(31)说明, 性连锁群体(16)经过一代随机交配, 有如下性质:

性质 3　经过一代随机交配, 性连锁群体的雄性基因信源为上代雌性基因信源, 因而有 $S[A_1(♂)] = S[A_0(♀)]$。这个性质说明, 平衡过程是下代雄性基因信源及其信息量变为上代雌性基因信源及其信息量的交递振荡过程, 即 $S[A_i(♂)] = S[A_{i-1}(♀)], i = 1, 2, \cdots\cdots$。

性质 4　$S[G_1(♀)] = S[A_0(♂)] + S[A_0(♀)]$。这个性质说明平衡过程是用上代雌、雄基因信源信息量表达下代雌性群体信息量的过程, 即 $S[G_i(♀)] = S[A_{i-1}(♂)] + S[A_{i-1}(♀)], i = 1, 2, \cdots\cdots$。当 $S[A(♂)] = S[A(♀)] = S[A_0(P)]$时, 再经过一代随机交配就平衡了。

性质 5　群体中基因平均信源及其信息量不变, 即 $S[A_1(P)] = S[A_0(P)]$。这个性质说明, 在随机交配下, 群体基因平均信源及其信息量逐代不变, 即 $S[A_i(P)] = S[A_0(P)], i = 1, 2, \cdots\cdots$。

性质 6　经过一代随机交配, 雌性基因信源 A, a 的频率为上代雌、雄基因信源同种基因频率的平均值, 即 $p_1(A♀) = \frac{p_m + p_f}{2}, q_1(a♀) = \frac{q_m + q_f}{2}$。

据性质 3 和性质 6, 各随机交配代数的雌性群体中基因 A 的频率为:

$$p_0(A♀) = p_f$$

$$p_1(A♀) = \frac{p_m}{2} + \frac{p_f}{2}$$

$$p_2(A♀) = \frac{1}{4}p_m + \frac{3}{4}p_f$$

$$p_3(A♀) = \frac{3}{8}p_m + \frac{5}{8}p_f$$

$$p_4(A♀)=\frac{5}{16}p_m+\frac{11}{16}p_f$$

……

$$p_n(A♀)=\frac{(\frac{2n}{3}-(-1)^n\frac{1}{3})p_m}{2^n}+\frac{(\frac{2^{n+1}}{3}+(-1)^n\frac{1}{3})p_f}{2^n}$$

由此易得:

性质7 性连锁不平衡群体,逐代随机交配下去,终将平衡,使$S[G(♂)]+S[G(♀)]$达到最大值$3S[A_0(P)]$,即

$$\lim_{n\to\infty}p_n(A♀)=\frac{2}{3}p_f+\frac{1}{3}p_m=p \tag{32}$$

$$\lim_{n\to\infty}q_n(a♀)=\frac{2}{3}q_f+\frac{1}{3}q_m=q \tag{33}$$

$$\lim_{n\to\infty}S(G_n(♀)]=\lim_{n\to\infty}(S[A_{n-1}(♂)]+S(A_{n-1}(♀)])=2S[A_0(P)] \tag{34}$$

$$\lim_{n\to\infty}(S[G_n(♀)]+S[G_n(♂)])=\lim_{n\to\infty}(S[G_n(♀)+S[A_{n-1}(♀)])=3S[A_0(P)] \tag{35}$$

3 讨 论

由性连锁群体的平衡及其信息熵的分析可知,平衡时信息熵最大,这在伴性遗传基因定位上很有意义。另外,性连锁不平衡群体,在随机交配下,其平衡过程就是信息熵增大的过程,即种内进化是熵增大过程,也是生物保持其遗传多样性的过程。普里高津[6]认为,理论物理的热力学第二定律说明自发过程是向熵增大方向发展,而生物进化是向熵减小的方向发展,因而他说:"生物学与理论物理学之间仍然存在着巨大的鸿沟"。本研究认为物理学的自发过程和生物进化过程都是向熵增大的方向发展。

[参考文献]

[1] 王身立.生物物理遗传学[M].长沙:湖南科学技术出版社,1992.
[2] 袁志发,郭满才,宋世德,等.相对Shannon信息量与基因变异的测量[J].西北农业大学学报,1998,26(4):30-34.
[3] 郭满才,宋世德,周静芋,等.非平衡群体基因变异测量的Shannon信息量方法[J].生物数学学报,2001,16(3):341-347.
[4] 周士谔,衡红刚,张国庆.数量遗传学中一种新的求综合性状的方法[J].遗传学报,1989,16(4):269-275.
[5] 柯卫东,刘采芹.数量遗传学中的信息问题初探[J].自然杂志,1998,13(4):210-212.
[6] 湛垦华,沈小峰.普里高津与耗散结构理论[M].西安:陕西科学技术出版社,1982.21.

Study on information model of sex-linked equilibrium population

LIU Jian-jun, GUO Man-cai, XIE Xiao-li, ZhANG Hong-li, ZhOU Jing-yu, YUAN Zhi-fa
(College of Life Sciences, Northwest Sci-Tech University of Agriculture and Forestry, Yangling, Shaanxi 712100, China)

Abstract: Based on the model of XY type sex chromosome, the entropy variation regularity of sex-linked population equilibrium was studied by using the information theory method. The results showed that the entropy of sex-linked equilibrium population was the maximum under the random mating system. To sex-linked disequilibrium population from disequilibrium to equilibrium, the entropy capacity would increase gradually. When it reached the equilibrium state, the entropy would go up to the maximum.

Key words: sex-linked gene; equilibrium population; disequilibrium population; gametal; entropy