

等输入模型下 DNA 序列点突变的熵性质

王晓芳, 袁志发, 张瑞明

(西北农林科技大学 理学院 应用数学系, 陕西 杨凌 712100)

[摘要] 目的: 研究在等输入模型下 DNA 序列熵随世代变化的规律及其性质. 方法: 基于 Shannon 信息熵理论, 在等输入模型下建立有限长 DNA 序列点突变微分方程组模型, 利用 Lagrange 乘数法进行理论证明, 并用 Mathematica 软件模拟等输入模型下 DNA 序列熵的变化过程. 结果: DNA 序列在等输入模型下发生突变, 其熵随世代增加随机波动, 且趋于最大值. 结论: DNA 序列点突变具有“保熵性”, 并验证了生物是朝着多样性增大方向进化的理论.

[关键词] 生物多样性; 核苷酸替代; Shannon 信息熵; 等输入模型

[中图分类号] O29; Q347 **[文献标志码]** A **[文章编号]** 1000-9965(2016)05-0436-06

doi:10.11778/j.jdxb.2016.05.016

Entropy properties of the point mutations in DNA sequence using equal-input model

WANG Xiaofang, YUAN Zhifa, ZHANG Ruiming

(Department of Applied Mathematics, College of Science, Northwest A&F University, Yangling 712100, China)

[Abstract] **Aim:** To further understanding the foundation of the mechanism of biological evolution, the changing of the entropy of DNA sequences among the generations was characterized using the equal-input model, which is a more complicated substitution model. **Methods:** Based on Shannon entropy theory, a set of differential equations was founded about point mutations of DNA sequence with limited length in the equal-input model, the theoretical demonstration was proved by the Lagrange multiplier method, and Mathematica was used to simulate the changing of the entropy of DNA sequences. **Results:** The mutations in the DNA sequences occurred in the equal-input model, the entropy fluctuated as the generations, and tended to be maximum. **Conclusion:** The point mutations in the DNA sequences can maintain their entropy, and it was verified that biodiversity of species increased during evolvement.

[Key words] biodiversity; Shannon entropy; nucleotides substitution; equal-input model

生物具有多样性, 以适应复杂多变的生存环境. 因此, 基因(遗传)多样性的存在是生物进化的基础. 这正是达尔文关于生物由简单到复杂、由低级到

高级逐渐进化理论的核心所在, 即生物进化是生物为适应环境而保持或改变成新的多样性的过程^[1-2]. 生物多样性主要包括基因(遗传)多样性、

[收稿日期] 2016-07-01

[基金项目] 国家自然科学基金项目(11371294)

[作者简介] 王晓芳(1991-), 女, 研究方向: 生物数学

通信作者: 张瑞明, 教授, 博士, 研究方向: 理论数学、应用数学与计算机软件, E-mail: RuimingZhang2010@163.com

物种多样性、生态系统多样性. 而基因(遗传)多样性一般体现在 DNA 序列一级结构的多样性上^[3-4], 这种多样性的改变主要依赖于 DNA 复制过程中的碱基突变.

生物进化是朝着基因(遗传)多样性增大的方向进行的,这是一个自发优化且不可逆的过程^[2]. Lewontin^[5]曾建议采用 Shannon 信息指数,即熵来度量基因变异. 但 Nei^[6]认为,Shannon 信息指数与任何遗传物质无关,是用来测定信息工程中的信息量的. 袁志发等^[7]和郭满才等^[8-10]自 2002 年开始运用最大信息熵原理对多种孟德尔群体证明了“平衡群体信息熵最大定理”,以 Shannon 信息熵为中心建立了多样性指标体系,给 Shannon 信息熵作为基因(遗传)多样性指标赋予了遗传学意义,是一个最好的多样性度量指标. 汪小龙等^[11]及张宏礼等^[12]从群体水平上阐明了生物多样性与熵之间的关系,提出最大熵分布就是 Hardy-Weinberg 平衡分布. 但目前分子水平上是否存在与最大熵原理一致的平衡原理,以及在这一水平上的 Shannon 信息熵性质的研究还比较少. Jukes 等^[13]以及 Kimura^[14]分别在单参数和双参数模型下研究了 DNA 序列点突变的性质,但均未涉及与熵相关的性质. 马国际等^[3]在单参数和双参数模型下分别建立了有限长 DNA 序列点突变的微分方程模型,并得出了 A,T,C,G 4 种碱基随世代变化的函数关系式,对点突变的熵性质在分子水平上进行了探索,得到生物是朝着多样性增大,即熵增加的方向进化的结论. 郭子湖等^[15-16]在 Tamura 模型下研究了 DNA 序列点突变的熵性质,说明生物进化具有“保熵性”. 本研究在马国际等^[3]以及郭子湖等^[15-16]基础上,基于等输入模型,研究有限长 DNA 序列点突变的熵性质,并对 DNA 序列熵随世代变化过程进行计算机模拟,以期从更细微的层次上揭示 DNA 分子碱基突变、Shannon 信息熵及基因(遗传)多样性之间的关系,并为进一步了解生物进化机理奠定基础.

1 资料与方法

1.1 DNA 序列的熵

熵的概念起源于热力学,由法国物理学家克劳修斯于 1854 年首次提出,用来描述物质热力状态分布的混乱程度. 1948 年,Shannon^[17]在《通信的数学理论》中利用概率分布理论给出了 Shannon 信息熵的定义. 设有一离散分布(表 1).

表 1 Shannon 信息熵中各符号的概率
Table 1 The probability of the symbols in the Shannon entropy

X	x_1	x_2	\cdots	x_n
$P(X)$	p_1	p_2	\cdots	p_n

表中 X 为状态空间, P 为 X 的概率空间, x_i 为信息符号, p_i 为 x_i 的概率.

若分布中 x_i 不能量化,那么将 X 和 P 均称为信息源,将 x_i 的信息量定义为 $-\ln p_i$ (显然, p_i 越小, x_i 的信息量越大),而将 x_i 的平均信息量称为信息源 X 的 Shannon 信息熵,记为 $S(X)$, $S(X)$ 刻画了总体 X 的多样性,计算公式为:

$$S(X) = - \sum_{i=1}^n P_i \ln P_i \tag{1}$$

如今,生物学领域也普遍地应用熵的概念来描述生物系统的复杂性. 对某条 DNA 序列,把 A,T,C,G 4 种核苷酸看成信号,加上在该序列中出现的频率 P_i ($i=1,2,3,4$,分别对应 A,T,C,G),构成 1 个信息源,其熵可以表示如下:

$$S(X) = - \sum_{i=1}^4 P_i \ln P_i \tag{2}$$

1.2 有限长 DNA 序列在等输入模型下点突变的熵性质

在研究核苷酸替代的过程中,学者提出了不同的核苷酸替代模型,等输入模型是由 Felsenstein^[18]以及 Tajima 等^[19]独立提出的. 该模型需假定核苷酸的静态分布(g_A, g_T, g_C, g_G)来估计核苷酸替代数,其碱基转化过程(图 1).

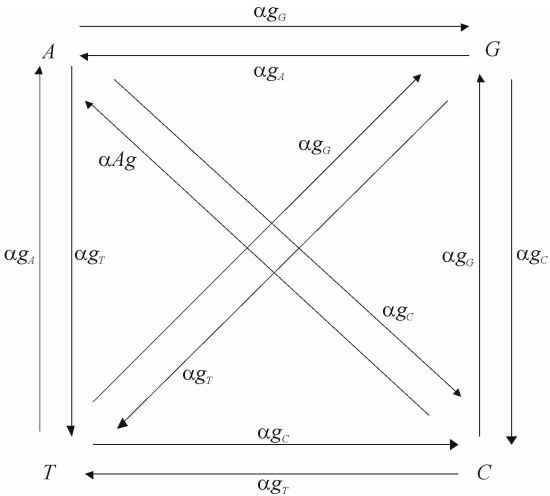


图 1 等输入模型中的碱基转化示意图
Fig. 1 The base transformation process in the equal-input model

图 1 中, A, T, C, G 分别代表腺嘌呤、胸腺嘧啶、胞嘧啶和鸟嘌呤 4 种核苷酸, g_A, g_T, g_C, g_G 分别代表 4 种核苷酸的平衡频率, $\alpha g_A, \alpha g_T, \alpha g_C$ 和 αg_G 为 4 种核苷酸之间的瞬时转化速率, α 是 1 个参数 ($\alpha > 0$). 平衡频率是指 4 种核苷酸之间的相互转化虽仍然存在, 但各自频率不再发生变化^[1]. 显然, $g_A + g_T + g_C$

$$\begin{bmatrix} \frac{dP_A(t)}{dt} \\ \frac{dP_T(t)}{dt} \\ \frac{dP_C(t)}{dt} \\ \frac{dP_G(t)}{dt} \end{bmatrix} = \begin{bmatrix} -\alpha(g_T + g_C + g_G) & \alpha g_T & \alpha g_C & \alpha g_G \\ \alpha g_A & -\alpha(g_A + g_C + g_G) & \alpha g_C & \alpha g_G \\ \alpha g_A & \alpha g_T & -\alpha(g_A + g_T + g_G) & \alpha g_C \\ \alpha g_A & \alpha g_T & \alpha g_C & -\alpha(g_A + g_T + g_G) \end{bmatrix} \begin{bmatrix} P_A(t) \\ P_T(t) \\ P_C(t) \\ P_G(t) \end{bmatrix} \quad (3)$$

式(3)中, $P_i(t), i = A, T, C, G$ 分别代表 t 时刻 A, T, C, G 4 种核苷酸的频率, t 代表 DNA 序列发生突变的世代. 在初值条件 $(P_A(0), P_T(0), P_C(0), P_G(0))^T$ 下, 该模型的解为

$$\begin{cases} P_A(t) = (P_A(0) - g_A)e^{-\alpha t} + g_A \\ P_T(t) = (P_T(0) - g_T)e^{-\alpha t} + g_T \\ P_C(t) = (P_C(0) - g_C)e^{-\alpha t} + g_C \\ P_G(t) = (P_G(0) - g_G)e^{-\alpha t} + g_G \end{cases} \quad (4)$$

显然, 当 $t \rightarrow +\infty$ 时, A, T, C, G 4 种核苷酸的频率分别为 g_A, g_T, g_C, g_G , 即 $t \rightarrow +\infty$ 时, 4 种核苷酸的频率达到平衡状态.

下面考查 4 种核苷酸频率满足什么条件时, 等输入模型下的 DNA 序列熵取得最大值. 先假设 A, T, C, G 4 种核苷酸出现初始频率分别为 P_1, P_2, P_3 和 P_4 , 则该问题可转化为在约束条件

$$\begin{cases} \sum_{i=1}^4 P_i = 1 \\ \lim_{t \rightarrow \infty} P_1 = g_A \\ \lim_{t \rightarrow \infty} P_2 = g_T \\ \lim_{t \rightarrow \infty} P_3 = g_C \\ \lim_{t \rightarrow \infty} P_4 = g_G \end{cases} \quad (5)$$

下, 求目标函数 $S(X) = -\sum_{i=1}^4 P_i \ln P_i$ 的最大值. 对此, 采用 Lagrange 乘数法, 首先构造 Lagrange 函数

$$\begin{aligned} f(P_1, P_2, P_3, P_4, \lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \\ -\sum_{i=1}^4 P_i \ln P_i + (\ln \lambda_0 + 1)(P_1 + P_2 + P_3 + \end{aligned}$$

$+ g_G = 1$, 当 $g_A = g_T = g_C = g_G = \frac{1}{4}$ 时, 等输入模型转化为 Jukes^[13] 单参数模型和 Kimura^[14] 双参数模型. 根据图 1 可以建立等输入模型下 DNA 序列点突变的微分方程组模型为:

$$\begin{bmatrix} \alpha g_C & \alpha g_G \\ \alpha g_C & \alpha g_G \\ -\alpha(g_A + g_T + g_G) & \alpha g_C \\ \alpha g_C & -\alpha(g_A + g_T + g_G) \end{bmatrix} \begin{bmatrix} P_A(t) \\ P_T(t) \\ P_C(t) \\ P_G(t) \end{bmatrix} \quad (3)$$

$$P_4 - 1) + \ln \lambda_1 (P_1 - g_A) + \ln \lambda_2 (P_2 - g_T) + \ln \lambda_3 (P_3 - g_C) + \ln \lambda_4 (P_4 - g_G) \quad (6)$$

求 $f(P_1, P_2, P_3, P_4, \lambda_0, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ 对 P_1, P_2, P_3 和 P_4 的偏导, 并使其分别为 0, 带入约束条件

$$\begin{cases} \frac{\partial f}{\partial P_1} = -\ln P_1 - 1 + \ln \lambda_0 + \ln \lambda_1 + 1 = 0 \\ \frac{\partial f}{\partial P_2} = -\ln P_2 - 1 + \ln \lambda_0 + \ln \lambda_2 + 1 = 0 \\ \frac{\partial f}{\partial P_3} = -\ln P_3 - 1 + \ln \lambda_0 + \ln \lambda_3 + 1 = 0 \\ \frac{\partial f}{\partial P_4} = -\ln P_4 - 1 + \ln \lambda_0 + \ln \lambda_4 + 1 = 0 \\ P_1 + P_2 + P_3 + P_4 = 1 \\ \lim_{t \rightarrow \infty} P_1 = g_A, \lim_{t \rightarrow \infty} P_2 = g_T, \lim_{t \rightarrow \infty} P_3 = g_C, \lim_{t \rightarrow \infty} P_4 = g_G \end{cases} \quad (7)$$

由上式可得: $P_1 = g_A, P_2 = g_T, P_3 = g_C, P_4 = g_G$,

即在此条件下, $S(X) = -\sum_{i=1}^4 P_i \ln P_i$ 取得最大值. 此时, DNA 序列熵取得最大值, 这与由式(4)取得最大熵的条件一致.

以上分析表明, 在等输入模型下, 由于需要对核苷酸频率的静态分布进行假定, 所以和单参数模型与双参数模型的不同点是, 当 $t \rightarrow +\infty$ 时, DNA 在等输入模型下各核苷酸的频率并非都是 $\frac{1}{4}$, 但对有限

长度的 DNA 序列在等输入模型下发生点突变, 其熵值随世代的变化随机波动, 且随世代增加逐渐趋于最大值是显而易见的, 这是因为与单参数和双参数模型相比, 该模型需要对核苷酸频率的静态分布进行假设, 说明生物在复杂的环境中具有保持多样性

的趋势,生物进化具有“保熵性”。

2 结果

2.1 有限长 DNA 序列点突变熵性质的计算机模拟

在平均意义上,生物的遗传物质 DNA 在复制过程中是按一定的突变率发生突变的. Denver 等^[20]经过 2 组实验多次验证,发现个体基因组发生突变的次数服从泊松分布,即:

$$P(\gamma,\lambda)=\frac{e^{-\lambda}\lambda^{\gamma}}{\gamma!}\tag{8}$$

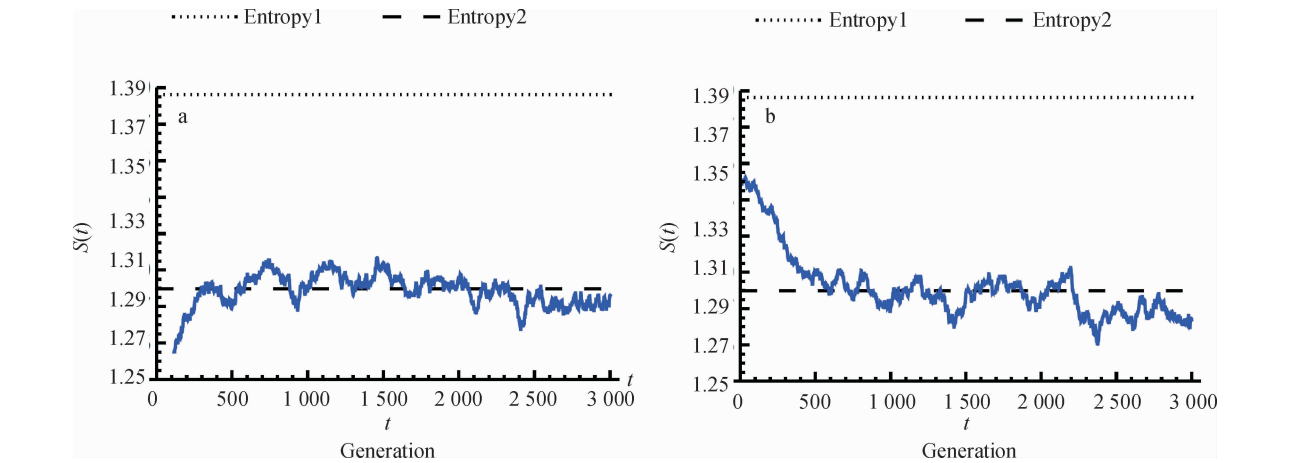
式(8)中, γ 为突变次数, $\lambda=\mu\times G$ 为常数(μ 为基因组突变率, G 为基因组大小). 并且,在模拟过程中多为突变次数服从泊松分布^[21-23].

利用 Mathematica 软件模拟等输入模型下 DNA

序列熵的变化过程. 研究中适当放大了突变率,以便更清楚地观察到模拟效果. 由于每种核苷酸发生突变的概率与其达到平衡时频率有关,所以,等输入模型中假定 4 种核苷酸平衡时频率之比为 0. 1: 0. 2: 0. 3: 0. 4.

2.2 对 4 种核苷酸初始频率不同的 DNA 序列进行模拟

利用计算机产生两条长度均为 3 000 bp 的 DNA 序列,第 1 条序列中 A, T, C, G 4 种核苷酸的初始频率之比为 0. 04: 0. 22: 0. 32: 0. 42, 第 2 条序列中 A, T, C, G 4 种核苷酸的初始频率之比为 0. 25: 0. 15: 0. 25: 0. 35. 设定每个世代突变的次数为 15, 分别模拟 3 000 个世代, 计算每条序列每个世代的 DNA 序列熵, 模拟结果(图 2).



a, b 分别表示核苷酸初始频率比例为 0. 04: 0. 22: 0. 32: 0. 42, 核苷酸初始频率比例为 0. 25: 0. 15: 0. 25: 0. 35, entropy1 和 entropy2 分别为均匀分布时的熵和等输入模型下的最大熵

图 2 等输入模型下不同初始频率的 DNA 序列熵的变化过程

a, b respectively represents the different initial frequencies 0. 04: 0. 22: 0. 32: 0. 42 and 0. 25: 0. 15: 0. 25: 0. 35; The entropy1 is the entropy of normal distribution, and the entropy2 is the maximum entropy of the equal-input model

Fig. 2 The entropy variation process of 4 nucleotides in the DNA sequences of different initial frequencies in the equal-input model

由模拟结果得出:图 2(a) 中 DNA 序列熵随世代随机波动逐渐趋于最大值,并在最大熵值附近波动;图 2(b) 中 DNA 序列熵随世代随机波动逐渐趋于最大值,并在最大熵值附近波动. 这说明 DNA 序列在等输入模型下发生突变,无论 4 种核苷酸的初始频率之比如何,即初始序列熵值如何,其序列熵总是趋于最大熵值的,与最大熵原理一致.

2.3 对突变率不同而长度相同的 DNA 序列进行模拟

利用计算机生成 3 条长度均为 3 000 bp 的 DNA 序列, A, T, C, G 4 种核苷酸的初始频率之比为

0. 04: 0. 22: 0. 32: 0. 42. 设定每个世代突变的次数分别为 15, 35 和 55, 并模拟 3 000 个世代, 计算每条序列每个世代的 DNA 序列熵, 模拟结果(图 3).

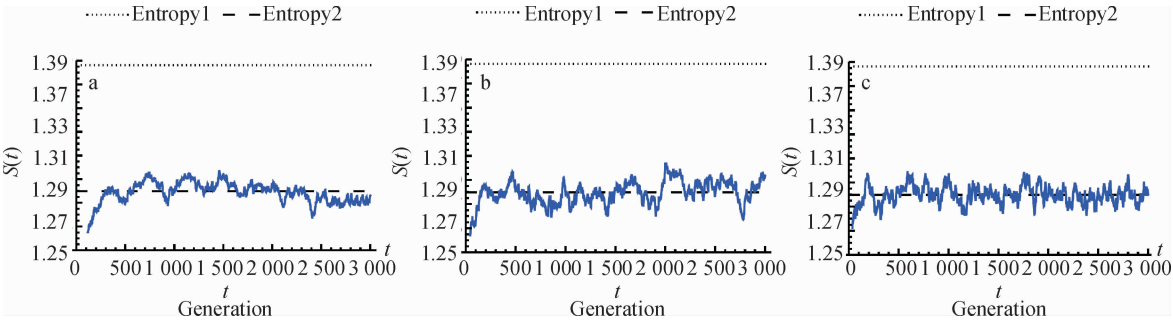
由图 3 可以看出,在突变率不同而长度相同的 DNA 序列中, DNA 序列熵随世代增加随机波动,并逐渐趋于最大值. 每个世代突变次数为 15 的 DNA 序列达到最大熵值所需世代最多,其熵随世代波动较为平缓;而每个世代突变次数为 55 的 DNA 序列达到最大熵值所需世代最少,其熵随世代波动较剧烈. 这说明基因组突变率越高,序列熵达到最大值的速度越快,随世代波动也越明显. 与马国际等^[3]在

单参数模型和双参数模型下以及郭子湖等^[15]在 Tamura 模型下模拟所得结论是一致的。

2.4 对长度不同而突变率相同的 DNA 序列进行模拟

利用计算机生成 3 条长度分别为 3 000,5 000,

7 000 bp 的 DNA 序列,A,T,C,G 4 种核苷酸的初始频率之比为 0.04:0.22:0.32:0.42. 设定每个世代突变的次数均为 15,并模拟 3 000 个世代,计算每条序列每个世代的 DNA 序列熵,模拟结果(图 4)。

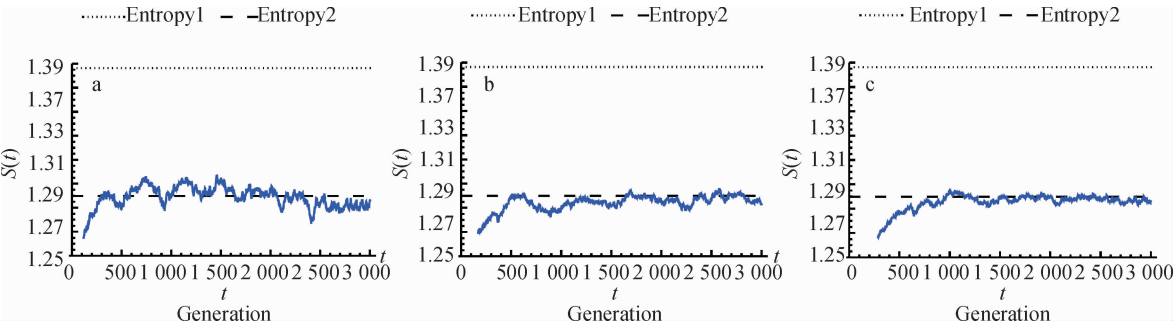


a,b,c 分别表示每个世代突变次数为 15,35,55; entropy1 和 entropy2 分别为均匀分布时的熵和等输入模型下的最大熵

图 3 等输入模型下突变率不同而长度相同的 DNA 序列熵的变化过程

a,b,c respectively represents the number of each generation mutant 15,35 and 55;The entropy1 is the entropy of normal distribution,and the entropy2 is the maximum entropy of the equal-input model

Fig. 3 The entropy variation process of 4 nucleotides in the DNA sequences of different mutation rates and equal length in the equal-input model



a,b,c 分别表示 DNA 序列长度为 3 000,5 000,7 000 bp; entropy1 和 entropy2 分别为均匀分布时的熵和等输入模型下的最大熵

图 4 等输入模型下长度不同而突变率相同的 DNA 序列熵的变化过程

a,b,c represents the length of the DNA sequence respectively are 3 000, 5 000 and 7 000 bp;The entropy1 is the entropy of normal distribution,and the entropy2 is the maximum entropy of the equal-input model

Fig. 4 The entropy variation process of 4 nucleotides in the DNA sequences of different lengths and equal mutation rate in the equal-input model

图 4 中在长度不同而突变率相同的 DNA 序列,DNA 序列熵随世代变化随机波动,并随世代增加逐渐趋于最大值. 长度为 3 000 bp 的 DNA 序列达到最大熵值所需世代最少,其波动较剧烈;长度为 7 000 bp 的 DNA 序列达到最大熵值所需世代最多,其波动较平缓. 在突变率相同的情况下,基因组越小,序列熵达到最大值的速度越快,随世代波动越大,说明基因组越小的生物体经受突变的能力越弱,其多样性的改变越容易,进化也越快. 反之,基因组越大的生物体经受突变的能力越强,其多样性的改变越难,

进化也越慢,与单参数模型和双参数模型^[3]以及 Tamura 模型^[15]下点突变的规律表现一致。

3 讨论

本研究在等输入模型下,研究了 DNA 序列点突变的熵性质,并对等输入模型下 DNA 序列熵的变化规律进行了计算机模拟,得出了生物遗传物质 DNA 的序列熵随世代而变化,并逐渐趋于最大值的结论. 在计算机模拟过程中对 4 种核苷酸平衡时频率之比进行了假定,因而在等输入模型下取得的最大熵与

马国际等^[3]在单参数模型和双参数模型以及郭子湖等^[15]在 Tamura 模型下取得的最大熵有所不同,这是因为与单参数以及 Tamura 模型相比,等输入模型增加了新的约束条件,从而进一步说明,在不同的核苷酸替代模型下,DNA 序列熵依然随世代变化逐渐趋于最大值,并验证了生物是朝着多样性增大,即熵增加的方向进化的理论。

生物进化的过程是一个多样性增大的过程,其总趋势是由简单到复杂,由低级到高级。基因(遗传)多样性是生物多样性的的重要组成部分,其本质体现在生物遗传物质 DNA 序列一级结构的多样性,生物进化过程中碱基的随机突变是影响这种多样性的主要因素,当然,还有许多其他因素的影响^[24]。因此,在原有的研究基础上加上新的影响因素,建立新的核苷酸替代模型来对 DNA 一级结构组成及多样性改变的性质进行研究,有待进一步的研究,从而更加精确地展示生物的进化过程及基因组的进化和遗传的本质。

[参考文献]

- [1] 袁志发. 群体遗传学、进化与熵[M]. 北京: 科学出版社, 2011.
- [2] 袁志发, 周静芊, 郭满才, 等. 基因多样性与 Shannon 信息熵[J]. 动物生物技术通讯, 2002, 8(1): 353 - 358.
- [3] 马国际, 梁丽静, 范彦辉, 等. 点突变的熵性质[J]. 科学通报, 2008, 53(19): 2318 - 2323.
- [4] KIMURA M, OHTA T. Protein polymorphism as a phase of molecular evolution[J]. Nature, 1971, 229(5285): 467 - 469.
- [5] LEWONTIN R C. The apportionment of human diversity[J]. Evol Biol, 1972, 6: 381 - 398.
- [6] NEI M. Molecular population genetics and evolution[M]. Amsterdam: American Elsevier Pub Co, 1975: 128 - 132.
- [7] 袁志发, 郭满才, 宋世德, 等. 相对 Shannon 信息量与基因异变的测量[J]. 西北农业大学学报, 1998, 26(4): 30 - 34.
- [8] 郭满才, 宋世德, 周静芊, 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.
- [9] 郭满才, 袁志发, 陈 宏, 等. 遗传多样性指标体系研究[J]. 动物生物技术通讯, 2002, 8(1): 359 - 364.

- [10] 郭满才. 群体遗传变异的信息学模型研究[D]. 陕西杨凌: 西北农林科技大学, 2002.
- [11] 汪小龙, 袁志发, 郭满才, 等. 最大信息熵原理与群体遗传平衡[J]. 遗传学报, 2002, 29(6): 562 - 564.
- [12] 张宏礼, 张鸿雁. 关于最大信息熵原理与群体遗传平衡一致性的探讨[J]. 遗传, 2006, 28(3): 324 - 328.
- [13] JUKES T H, CANTOR C R. Evolution of protein molecules[C]//Munro H N. Mammalian Protein Metabolism. New York: Academic Press, 1969: 21 - 132.
- [14] KIMURA M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences[J]. J Mol Evol, 1980, 16(2): 111 - 120.
- [15] 郭子湖, 张瑞明, 袁志发. Tamura 模型下 DNA 序列点突变熵性质的研究[J]. 西北农林科技大学学报(自然科学版), 2013, 41(2): 189 - 194.
- [16] 郭子湖. 核苷酸替换模型中若干问题的研究[D]. 陕西杨凌: 西北农林科技大学, 2013.
- [17] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 27: 379 - 423, 623 - 656.
- [18] FELSENSTEIN J. Evolutionary trees from DNA sequences: A maximum likelihood approach[J]. J Mol Evol, 1981, 17(6): 368 - 376.
- [19] TAJIMA F, NEI M. Biases of the estimates of DNA divergence obtained by the restriction enzyme technique[J]. J Mol Evol, 1982, 18(2): 115 - 120.
- [20] DENVER D R, MORRIS K, MICHAEL L. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome[J]. Nature, 2004, 430(7000): 679 - 682.
- [21] KEIGHTLEY P D. Inference of genome-wide mutation rates and distributions of mutation effects for fitness traits: a simulation study[J]. Genetics, 1998, 150(3): 1283 - 1293.
- [22] DENG H W. Estimating (over)dominance coefficient and discriminating dominance vs. overdominance as the genetic cause of heterosis[J]. Genetics, 1998, 148: 2003 - 2014.
- [23] DENG H W. Characterization of deleterious mutations in outcrossing populations[J]. Genetics, 1998, 150(2): 945 - 956.
- [24] 张艳红, 刘帅帅, 王 信, 等. 三斑海马线粒体基因组核苷酸全序列结构与分析[J]. 动物学杂志, 2016, 51(3): 413 - 422.

[责任编辑:王景周, 朱颖娜]