

随机化的Eigen模型研究

乔丽华¹⁾²⁾³⁾ 赵同军^{1)†} 顾建中²⁾ 卓益忠²⁾

1)(河北工业大学理学院, 生物物理研究所, 天津 300401)

2)(中国原子能科学研究院核物理研究所, 北京 102413)

3)(河北医科大学基础医学院, 石家庄 050017)

(2013年6月23日收到; 2014年2月12日收到修改稿)

为了使Eigen模型能够更真实地描述物种的演化过程, 将确定性Eigen模型改造成随机模型. 以Eigen模型为理论框架, 把基因序列中每一个位点的突变率看作一个高斯随机变量, 从而导出随机性Eigen模型. 对于此随机性Eigen模型, 当突变率的涨落强度较小时, 准物种的误差阈位置几乎没有改变, 仍是个相变点; 而当突变率的涨落强度变大时, 误差阈由一个相变点变为一个转变区域. 在真实的物种演化过程中, 误差阈应是一个转变区域, 而且在解决实际问题时应考虑该转变区域的上限.

关键词: 高斯随机变量, 突变率, 准物种, 误差阈

PACS: 87.10.Mn, 87.10.Rt, 87.15.A-

DOI: 10.7498/aps.63.108701

1 引言

在物种演化模型中, 最有代表性的是Eigen模型^[1,2]和Crow-Kimura模型^[3,4], 它们是建立在单峰适应面上的确定性模型. 大量研究表明, 准物种分布会随着突变率的增加出现一个误差阈(相变点)^[5,6]. 这一相变使准物种从适应面的高峰突然趋向消亡. 然而, 这一结论是在确定性模型下得到的. 事实上, 用随机性模型研究准物种分布和误差阈更符合实际情况. 目前, 上述领域的研究已取得了很大的进展^[7-11], 而随机化问题又是该领域中的一个研究热点^[12-14]. 但是, 将物种演化模型作随机化的研究还比较少, 而且大多数研究只是在确定性方程中加上一项随机力, 使之成为形如Langevin方程或Fokker-Planck方程.

本文将确定性方程中的有关物理量随机化, 即将具有单一值的物理量处理成高斯随机变量. 此随机化的理论方法物理图像清楚、简单易行, 特别适用于处理那些具有内禀随机性的问题. 因此, 这一理论方法具有普遍意义. 文献^[15, 16]将Eigen模

型中的适应度这一物理量做了高斯随机化处理, 并得到了有价值的研究结果.

在分子水平上物种演化受多方面复杂因素的影响^[17], 无论是基因因素还是基因之外的环境因素都能导致突变率随空间和时间的不规则改变, 因此, 突变率应该是随机的, 这已被实验所证实^[18]. 所以, 本文仍以Eigen模型为基础, 进一步把突变率这一重要物理量处理成高斯随机变量, 从而使Eigen模型更符合实际, 更能够正确地描述物种演化过程.

2 Eigen模型的随机化

影响有自复制能力的生物群体演化过程的最主要因素是适应度和突变率^[1,2]. 适应度使演化群体沿着适应面达到局部最大值, 而突变率使其偏离原来的最大值并趋向新的最大值. 我们将以Eigen模型为例研究其随机化问题.

Eigen模型下准物种^[19-21]方程为

$$\dot{x}_i = \sum_{j=0}^n x_j f_j q_{ji} - \phi(\mathbf{x}) x_i$$

† 通讯作者. E-mail: zhaotj66@126.com

$$(i = 0, 1, 2, \dots, n), \quad (1)$$

其中, x_i 为包含 n 个碱基的第 i 类分子序列的浓度, 序列分类依据为 Hamming 距离; f_j 为适应度; q_{ji}

为从第 j 类序列到第 i 类序列的突变率; ϕ 为平均适应度, 且满足 $\phi(\mathbf{x}) = \sum_{i=1}^n f_i x_i$.

在考虑点突变的情况下 [22,23],

$$q_{ji} = \sum_{l=l_{\min}}^{l_{\max}} \left[\binom{i}{l} \binom{n-i}{j-l} (1-u)^{n-i-j+2l} u^{i+j-2l} \right], \quad (2)$$

其中,

$$l_{\min} = \max\{0, j+i-n\},$$

$$l_{\max} = \min\{j, i\},$$

u 为每一位点的突变率. 在先前的一些工作中, 一直沿用经典 Eigen 模型的突变率, 该模型中假定突变率与序列的特定位置无关, 也与其他位点是否突变无关, 因而突变率是单一值. 这是一个非常简化的模型, 也称为均匀误差率模型. 本文仍假定所有位点有相同的平均突变率, 但是把突变率 u 看成高斯随机变量, 因此可以称为随机突变率模型, 突变率 u 的概率密度分布 $p(u)$ 满足如下关系:

$$p(u) = \frac{1}{\sqrt{2\pi\omega_0^2}} \exp\left[-\frac{(u-u_0)^2}{2\omega_0^2}\right], \quad (3)$$

其中, u_0 为突变率的中心值, ω_0^2 为突变率的方差, ω_0/u_0 为突变率的涨落强度.

将随机化的 Eigen 方程线性化 [24] 可以得到

$$\frac{dX_i}{dt} = \sum_{j=0}^n X_j f_j q_{ji}, \quad (4)$$

其中 $X_i(t) = x_i(t) \exp(\varphi(t))$. 解一阶线性微分方程 (4), 得到

$$x_i(t) = \frac{\sum_{k=0}^n l_{ik} \sum_{j=1}^{n+1} h_{kj} x_j(0) \exp(\lambda_k t)}{\sum_{i=1}^{n+1} \sum_{k=0}^n l_{ik} \sum_{j=1}^{n+1} h_{kj} x_j(0) \exp(\lambda_k t)} \quad (i = 0, 1, 2, \dots, n+1). \quad (5)$$

(5) 式中各变量的意义详见文献 [1].

当 $t \rightarrow \infty$ 时, 得到

$$x_{i0} = \frac{l_{i0}}{\sum_{j=1}^{n+1} l_{j0}}, \quad (6)$$

其中 x_{i0} 为达到稳态时准物种群体中各类分子序列的相对浓度.

3 计算结果及讨论

3.1 对突变率 u 高斯随机分布的检验

我们将上述模型中的突变率 u 做了高斯随机化处理, 并编写了计算稳态时相对浓度的程序. 为了验证突变率 u 是否符合高斯随机分布, 将 u 的概率密度的数值计算结果 (虚线) 与由 (3) 式得到的概率密度 (实线) 做比较, 结果如图 1 所示. 从图 1 可以看出, 当突变率的中心值 u_0 为 0.1, 涨落强度为 0.02 时, 突变率 u 的确是高斯随机变量.

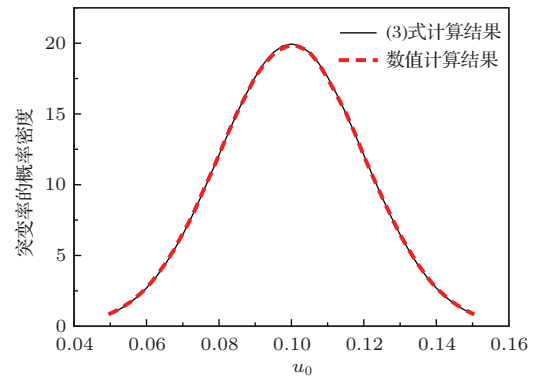


图 1 (网刊彩色) 突变率 u 的概率密度分布

3.2 随机突变率模型下的准物种演化

图 2 给出了随机突变率模型下计算得到的准物种分布, 其中, 基因序列长度 n 取为 20, 突变率中心值 u_0 的取值范围为 0.0—1.0, 而突变率的涨落强度 ω_0/u_0 取为零 (零涨落). 当 $u_0 = 0.0$ 时, 初始相对浓度为 1.0 的曲线代表的是主序列 (野生型) 的演化, 而初始相对浓度为 0.0 的曲线代表的是突变序列的演化. 图 2 完全再现了确定性 Eigen 模型的准物种分布结果 [1,25,26], 这是对本文程序正确性的验证. 从图 2 可以看出, 误差阈的位置在 0.11. 当 $u_0 > 0.11$ 时, 各个分子序列以等概率出现. 图 2 显示的是各类分子序列的浓度, 即各个分子序列浓度乘以简并度以后得到的各类分子序列的浓度. 当突

变率超过误差阈以后, 各个分子序列出现的概率非常小, 所有分子序列都会消亡.

下面我们考察当突变率 u 的涨落强度由小变大时准物种分布是如何改变的, 将着重关注误差阈的变化.

图3给出了突变率 u 的涨落强度由零增大到 0.001 时计算得到的准物种分布, 是系综平均后的结果. 从图3可以看出, 在 ω_0/u_0 趋近于零时, 误差阈的位置也在 0.11, 仍是一个相变点.

当突变率的涨落强度逐渐增大至 0.03 时, 误差阈开始有明显变化, 结果如图4所示. 误差阈由一个位点变为有一定宽度的转变区域, 大致在 0.11 到 0.12 之间. 此时的涨落强度可以作为引起转变区域产生的临界涨落强度.

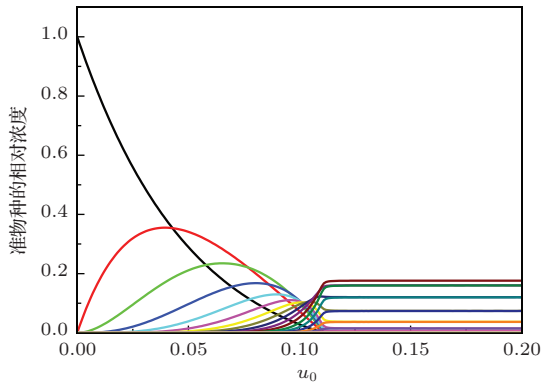


图2 (网刊彩色) 当 $n = 20, f_0 = 10, f_i = 1 (i \neq 0)$, $\omega_0/u_0 = 0.0$ 时, 稳态情形下准物种的相对浓度分布

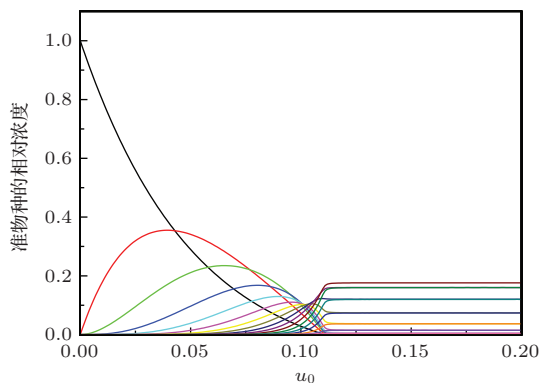


图3 (网刊彩色) 当 $n = 20, f_0 = 10, f_i = 1 (i \neq 0)$, $\omega_0/u_0 = 0.001$ 时, 稳态情形下准物种的相对浓度分布

图5给出了突变率的涨落强度增大为 0.1 时计算得到的准物种分布结果. 从图5可以看出, 误差阈变宽, 大致在 0.11 到 0.13 之间. 图6是突变率的涨落强度为 0.19 时计算得到的准物种分布结果. 从图6可以看出, 误差阈的宽度更宽, 大致在 0.12 到 0.15 之间, 而且误差阈中心位置向后漂移.

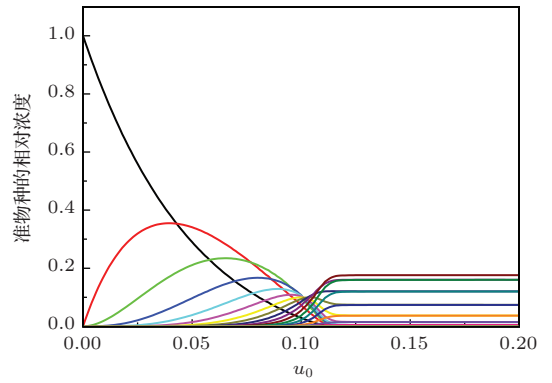


图4 (网刊彩色) 当 $n = 20, f_0 = 10, f_i = 1 (i \neq 0)$, $\omega_0/u_0 = 0.03$ 时, 稳态情形下准物种的相对浓度分布

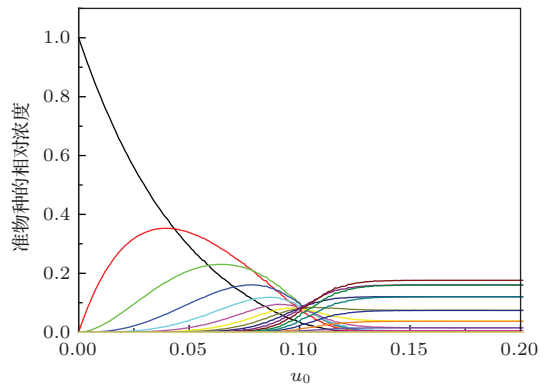


图5 (网刊彩色) 当 $n = 20, f_0 = 10, f_i = 1 (i \neq 0)$, $\omega_0/u_0 = 0.1$ 时, 稳态情形下准物种的相对浓度分布

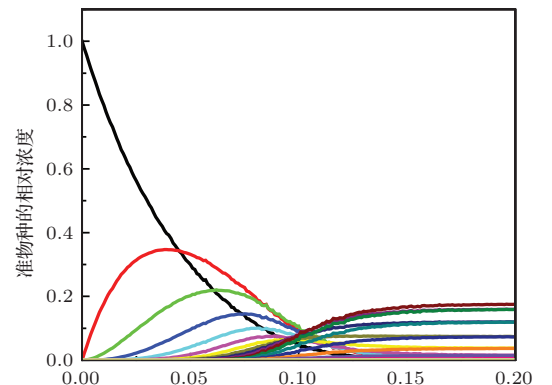


图6 (网刊彩色) 当 $n = 20, f_0 = 10, f_i = 1 (i \neq 0)$, $\omega_0/u_0 = 0.19$ 时, 稳态情形下准物种的相对浓度分布

如果继续增大突变率的涨落强度, 我们会发现准物种的浓度开始出现负值或紊乱, 该结果不再有意义. 这说明突变率的涨落强度应该是有上限的, 不能无限制地增大.

3.3 突变率随机化与适应度随机化的比较

图7给出了将适应度作高斯随机化处理^[15,16]得到的结果, 其中, 序列长度仍取 $n = 20$, 主序列

和突变序列适应度的涨落强度均为0.1. 图8给出了同样序列长度下突变率的涨落强度为0.1时计算得到的结果.

比较图7和图8可以看出, 在同样参数下, 适应度随机化时对误差阈的影响比较小, 而突变率随机化时对误差阈的影响比较大. 这说明在物种演化中突变率的随机化起着更重要的作用.

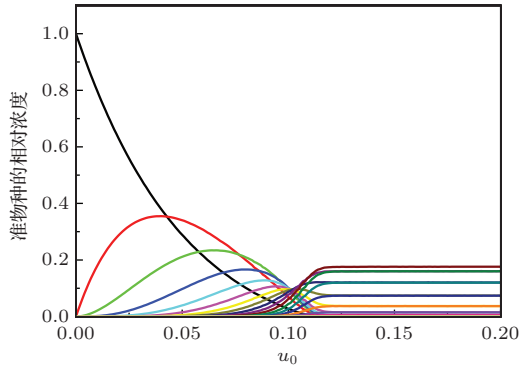


图7 (网刊彩色) 适应度处理成高斯随机变量后, 稳态情形下准物种的相对浓度分布, $n = 20$, 适应度涨落强度为0.1

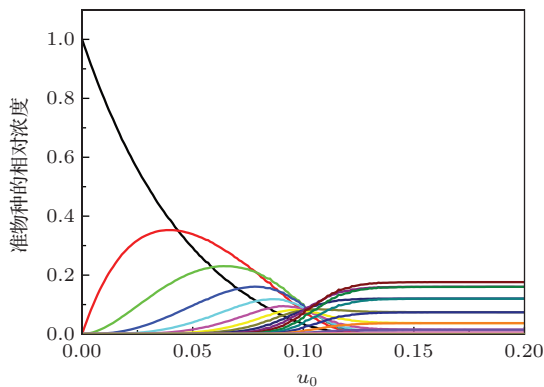


图8 (网刊彩色) 突变率处理成高斯随机变量后, 稳态情形下准物种的相对浓度分布, $n = 20$, $\omega_0/u_0 = 0.1$

3.4 误差阈随序列长度和突变率涨落强度的变化

表1列出了误差阈随序列长度和突变率涨落强度的变化. 这里当序列长度 $n = 10, 20, 30$ 时, 对应的突变率涨落强度的上限分别为0.28, 0.19, 0.18. 由表1可以得出以下两点: 第一, 在其他参数确定的情况下, 序列长度 n 越大, 误差阈越小, 即误差阈与 n 成反比, 这与确定性Eigen模型中得到的误差阈存在条件一致^[5,6]; 第二, 对于同一基因序列长度, 突变率的涨落强度很小时, 误差阈变化不明显, 仍是一个相变点, 这说明对于小的扰动, 误差阈是稳定的 (涨落可以看成一种扰动). 当涨落强度逐

渐变大时, 误差阈开始由一个相变点转变成一个有一定宽度的转变区域, 而且区域的上限向突变率增大的方向漂移, 当序列长度 $n = 10, 20, 30$ 时, 对应的突变率临界涨落强度 $(\omega_0/u_0)_c$ 分别为0.1, 0.03, 0.02.

表1 误差阈随序列长度 n 和突变率涨落强度 ω_0/u_0 的变化

序列长度 n	误差阈/100%		
	$\omega_0/u_0 = 0$	$(\omega_0/u_0)_c$	$\omega_0/u_0 = 0.01$
10	0.22	0.23	0.225
20	0.11	0.12	0.115
30	0.075	0.08	0.075

4 结 论

本文将描述物种演化的确定性Eigen模型改造成一个随机模型, 使得Eigen模型能够更真实地描述物种的演化过程. 对于同一个基因序列, 突变率的涨落强度趋近于零时, 误差阈是一个相变点, 说明对于小的扰动, 误差阈是稳定的. 当突变率的涨落强度逐渐增大时, 误差阈由一个相变点变为一个有一定宽度的转变区域. 这是一个很重要的现象, 它对于解决实际问题具有重要意义. 例如, 在研究抗病毒案例时, 原来的确定性模型所使用的剂量是使其突变率达到误差阈即可, 但考虑突变率的随机性后, 使用的剂量应使病毒突变率达到转变区域的上限才能把它完全消灭. 这种改进的Eigen模型具有普遍意义, 利用它可以进一步研究遗传、变异等问题, 也可以把这种随机方法应用于其他物种演化模型. 因此, 把Eigen模型中的突变率高斯随机化处理具有广泛的应用价值.

感谢展永教授、耿金鹏博士、李多芳博士和冯晓利博士的有益讨论.

参考文献

- [1] Eigen M 1971 *Naturwissenschaften* **58** 465
- [2] Eigen M, McCaskill J, Schuster P 1989 *Adv. Chem. Phys.* **75** 149
- [3] Crow J F, Kimura M 1970 *An Introduction to Population Genetics Theory* (New York: Harper & Row) pp3-382
- [4] Kimura M 1984 *The Neutral Theory of Molecular Evolution* pp34-50
- [5] Kamp C, Bornholdt S 2002 *Phys. Rev. Lett.* **88** 068104
- [6] Kamp C 2003 *Microbes Infect.* **5** 1397

- [7] Tang F K, Wang Q, Hua N, Lu H, Tang X Z, Ma P 2011 *Chin. Phys. B* **20** 010702
- [8] Tang F K, Wang Q, Hua N, Tang X Z, Lu H, Ma P 2010 *Chin. Phys. B* **19** 120601
- [9] Deng S X, Cheng C, Ni G H, Meng Y D, Chen H 2010 *Chin. Phys. B* **19** 105203
- [10] Wang J, Ma Q L 2008 *Chin. Phys. B* **17** 4424
- [11] Li Z, Chen G H, Zhang L H, Yang Q S, Feng J 2006 *Chin. Phys.* **15** 310
- [12] Sheng L Y, Jia W Y 2005 *Acta Phys. Sin.* **54** 5574 (in Chinese) [盛利元, 贾伟尧 2005 物理学报 **54** 5574]
- [13] Liu F M, Zhai H C, Yang X P 2003 *Acta Phys. Sin.* **52** 2462 (in Chinese) [刘福民, 翟宏琛, 杨晓莘 2003 物理学报 **52** 2462]
- [14] Huang X G, Xu J X, Huang W, Zhu P C 2001 *Acta Phys. Sin.* **50** 2296 (in Chinese) [黄显高, 徐健学, 黄伟, 朱甫臣 2001 物理学报 **50** 2296]
- [15] Feng X L, Gu J Z, Li Y X, Zhuo Y Z 2007 *Commun. Theor. Phys.* **48** 763
- [16] Feng X L, Li Y X, Gu J Z, Zhuo Y Z 2007 *J. Theor. Biol.* **246** 28
- [17] Muller H J 1922 *Am. Nat.* **56** 32
- [18] Drake J W, Holland J J 1999 *Proc. Natl. Acad. Sci.* **96** 13910
- [19] Eigen M, Schuster P 1977 *Naturwissenschaften* **64** 541
- [20] Eigen M 1986 *Chem. Scrip.* **26B** 13
- [21] Schuster P 1986 *Chem. Scrip.* **26B** 27
- [22] LeuthÄusser I 1987 *J. Chem. Phys.* **84** 1884
- [23] LeuthÄusser I 1987 *J. Stat. Phys.* **48** 343
- [24] Thompson C J, McBride J L 1974 *Math. Biosci.* **21** 127
- [25] Eigen M 2000 *J. Biophys. Chem.* **85** 101
- [26] Schuster P 2011 *Theory Biosci.* **130** 71

Eigen model of randomness in species evolution

Qiao Li-Hua¹⁾²⁾³⁾ Zhao Tong-Jun^{1)†} Gu Jian-Zhong²⁾ Zhuo Yi-Zhong²⁾

1) (Institute of Biophysics, School of Science, Hebei University of Technology, Tianjin 300401, China)

2) (Department of Nuclear Physics, China Institute of Atomic Energy, Beijing 102413, China)

3) (School of Basic Medicine, Heibe Medical University, Shijiazhuang 050017, China)

(Received 23 June 2013; revised manuscript received 12 February 2014)

Abstract

We modify the deterministic Eigen model of species evolution into a randomized model in order to render the Eigen model more realistic for the description of species evolution. In the framework of the Eigen model, we regard the locus mutation rate of a genetic sequence as a Gaussian distributed random variable. Thus the Eigen model turns into a random model. In this randomized model, we can see that when the fluctuation strength of the mutation rate is small, the error threshold of the quasispecies changes slightly and still serves as a phase transition point. However, when the fluctuation strength becomes large, the error threshold shifts from a phase transition point into a crossover region. Since the error threshold in the real species evolution is a crossover region, we should consider the upper limit of the crossover region when dealing with practical problems.

Keywords: Gaussian distributed random variable, mutation rate, quasispecies, error threshold

PACS: 87.10.Mn, 87.10.Rt, 87.15.A–

DOI: 10.7498/aps.63.108701

† Corresponding author. E-mail: zhaotj66@126.com