论著

文章编号:1000-5404(2007)08-0702-03

基因调控网络建立的生物动力学方程研究

刘 岭,易 东 (第三军医大学军事预防医学院卫生统计学教研室,重庆 400038)

提 要:目的 基于基因表达的时空特性,寻求一种新的描述基因调控网络的生物动力学方程模型。**方法** 从生物物种竞争动力学系统的角度出发,提出一种新的 Lotka-Volterra 微分方程模型,并应用于基因的时间序列表达数据。**结果** 将建立的微分方程应用于酵母基因的表达调控网络研究,得到了调控关系的预测结果,并与试验结果进行了比较研究。**结论** 本研究的预测结果与试验结果基本吻合,该模型弥补了其他模型的不足,是一种新的有价值的基因表达时空微分方程模型。

关键词: 基因调控网络;Lotka-Volterra 方程;酵母基因表达

中图法分类号: R311;R394 **文献标识码:** A

Study on bio-dynamic equation for gene regulation networks

LIU Ling, YI Dong (Department of Medical Statistics, College of Preventive Medicine, Third Military Medical University, Chongqing 400038, China)

Abstract: Objective To explore a new bio-dynamic differential equation to establish gene regulatory network based on the time-space of gene expression. Methods According to bio-dynamic competitive model and differential equation, a new time-space Lotka-Volterra differential equation was established and was applied to time-order expression data. Results The model was applied to set up the regulation network of yeast genes. The regulatory relation was found and compared with the experiment results. Conclusion The results from the new model were almost consistent with the real situation. So this model can remedy the shortage of other models. It's a new valuable differential equation model.

Key words: gene regulation network; Lotka-Volterra equation; yeast gene expression

随着 cDNA 芯片技术的不断发展,在全基因组水平上描述基因表达调控网络关系的研究已成为热点。当前,已有一些数学模型基于基因表达时间序列数据来描述基因之间的调控关系。例如聚类分析[1],Somogyi 等[2]的布尔模型,Liang 等[3]的反向工程布尔调控函数,Thieffry 等[4]的广义逻辑模型,McAdams 等[5]的混合模型,Chen 等[6,7]提出的线性微分方程模型。

本研究借助"多物种生态系统动态发展理论"的思想[8],从种群竞争过程的动力学观念出发,将整个基因的调控环境视为一个动态竞争过程,用 Lotka-Volterra 方程来描述发育过程中基因之间的相互关系。

1 材料与方法

1.1 基因的选择

基因的选择来源于 Spellman 等创建的公共数据库。该数据库包括芽殖酵母菌的 6 178 个开放阅读框(open read frame, ORF)在一个细胞周期中的基因表达数据(网址 http://genome-www.standford.edu/cellcycle)。

我们在 CDC-28 细胞株 17 个时间点(每 10 分钟取 1 次)的基因中,筛选出已被传统的分子生物学试验证实的在 4 个细胞周期中被调控的目标基因和有调控作用的调控基因。

目标基因为:Aga1、Clb3、Clb4、Far1、Fus1、Hsm1、Ho、Hta1、Htb1、Hht1、Ndd1、Pcl2、Svs1。

调控基因为:Clb2、Cln2、Cln3、Hsl7、Mcm1、Spo12、Sic1、Sim1、Ste12、Swe1、Swi4、Swi5。

1.2 数据预处理

对每个基因表达在各时间点的导数,本研究首先采用 3 次样条函数对表达数据进行拟合,然后求出基因表达每个时间点的变化如调控方程中的 $X(t)$。同时可以消除原始数据的一部分误差。

1.3 Lotka-Volterra 方程的建立

我们将以上 25 个基因的调控网络视为一个竞争过程,于是构造了 Lotka-Volterra 方程,具体如下:

$$\frac{dX_i(T)}{dt} = X_i(t)\left[\left(a_i + \sum_{j=1}^{n} W_{ij}X_j(t)\right)\right] \quad (1)$$

其中,$X_i(t)$ 表示 t 时点,第 i 个基因的表达量($i=1,2,\cdots,25$);a_i 表示第 i 个基因的内禀增长率;W_{ij} 表示第 j 个基因对第 i 个基因的调控作用。

基金项目:国家自然科学基金(60371034)
Supported by the National Natural Science Foundation of China (60371034)
作者简介:刘 岭(1969-),女,江苏省江阴市人,硕士研究生,讲师,主要从事基因调控网络数据分析方面的研究。电话:(023)68752343,E-mail: liuling_505@sina.com
通讯作者:易 东,电话:(023)68752344, E-mail:yd_house@hotmail.com
收稿日期:2006-11-17;修回日期:2006-12-01

本研究我们关心的是对如何确定方程(1)中的系数矩阵 $W=(W_{ij})$，方程(1)在一般情况下没有解析表达式，并且由于有25个自变量，因此从理论上讲需要26个时间点的表达数据才能确定系数矩阵，但是，由于受实验条件的限制，不可能做这样大规模的实验，本研究4个周期，每个周期仅有 $K_i(i=1,2,3,4)$ 个时间点数据。为此，我们结合基因表达的特点，根据假定提出了一种插值方法，并求得方程的数值解法。这种数值解法基于的假定为：所选基因的表达是一个连续平稳的过程，即不存在突变值对调控关系的影响。由此，我们可以对每个基因的表达曲线进行3次样条拟和，得到1条连续的表达曲线。然后，在$[1,K_i]$时间区间上，以 $K_i/26$ 为间隔，等距离选取26个时间点，其相应的曲线上的点作为表达强度的模拟值(含1和 K_i 两个时点)。所以方程的解可表达为：

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{25} \\ W_{11} & W_{21} & \cdots & W_{25,1} \\ \cdots & \cdots & \cdots & \cdots \\ W_{1,25} & W_{2,25} & \cdots & W_{25,25} \end{pmatrix} = \begin{pmatrix} 1 & X_1(1) & \cdots & X_{25}(1) \\ & \cdots & \cdots & \\ 1 & X_1(26) & & X_{25}(26) \end{pmatrix}^{-1} \begin{pmatrix} \frac{dX_1(1)}{dt}\cdot X_1(1) & \cdots & \frac{dX_{25}(1)}{dt}\cdot X_{25}(1) \\ \frac{dX_1(2)}{dt}\cdot X_1(2) & \cdots & \frac{dX_{25}(2)}{dt}\cdot X_{25}(2) \\ \cdots & \cdots & \cdots \\ \frac{dX_1(26)}{dt}\cdot X_1(26) & \cdots & \frac{dX_{25}(26)}{dt}\cdot X_{25}(26) \end{pmatrix} \quad (2)$$

2　结果

2.1　基于 Lotka-Volterra 方程的调控网络模型

根据 Spellman 等试验取样的情况，时间点大致对应的细胞周期见表1。

表1　时间点与细胞周期对应情况

细胞周期	对应时间点(min)
G_1	0,10,20,90,100
S	30,40,110,120
G_2	50,60,130,140
M	70,80,150,160

根据各时间点的表达数据得到4个数据矩阵(由于数据量过大，未列出)。在实际生物情况中，1个基因只受有限个调控基因的控制，所以我们将同一周期的矩阵作点乘得到相应的周期调控矩阵，并根据调控矩阵，采用 SAM 软件作图1～4，各图中实线表示正调控，虚线表示负调控，无连续表示不存在调控关系。

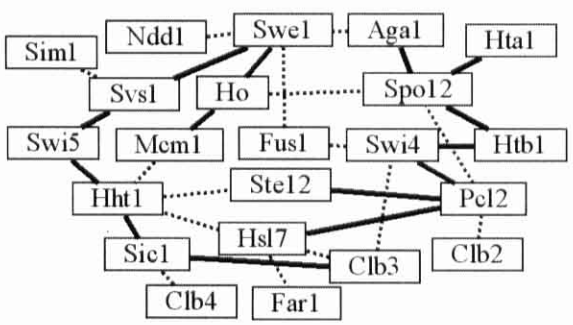


图1　G_1 期调控基因对目标基因的调控情况

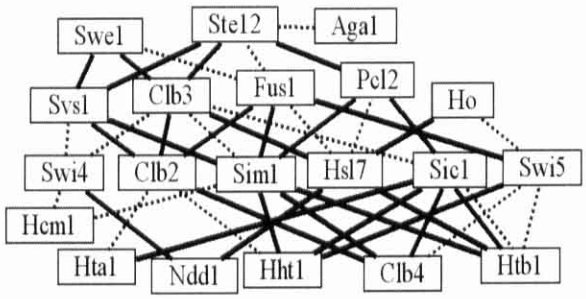


图2　S 期调控基因对目标基因的调控情况

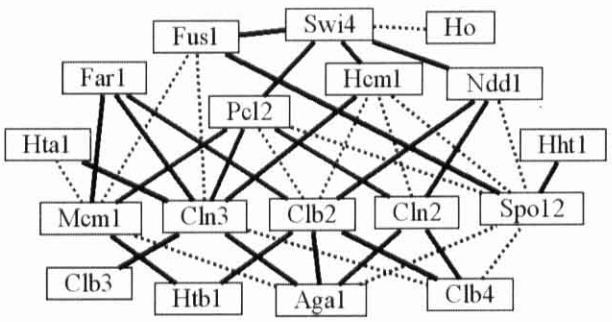


图3　G_2 期调控基因对目标基因的调控情况

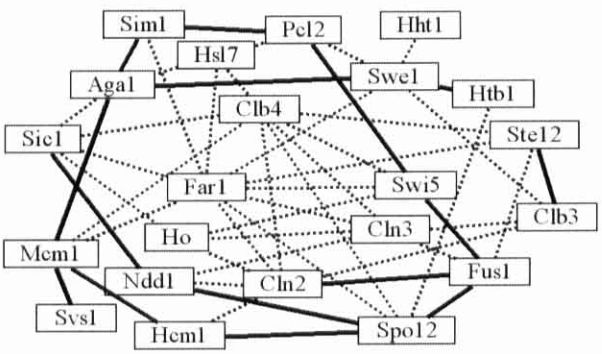


图4　M 期调控基因对目标基因的调控情况

2.2　网络模型的生物学验证

Cln2 是酵母染色体中一个重要的细胞周期调控基因，它编码细胞周期蛋白，同 Cdc28 结合形成 Cln2/Cdc28 复合体。这个复合体在细胞的出芽和纺锤体的极化等 G_2、M 期重要细胞学事件中发挥重要的作用。本模型中 Cln2 对所选的13个目标基因在 G_1 和 S 期没有作用，在 G_2 期对 Aga1、Clb4、Ndd1、Pcl2 有促进作用，对 Hcm1 有抑制作用，在 M 期对 Fus1 有促进作用，对 Clb3、Clb4、Far1、Hcm1、Ho、Ndd1 有抑制作用。

SBF(Swi4、Swi6 细胞周期盒黏附因子)是转录调控因子，在细胞周期 G_1/S 期结合许多基因的启动子，以达到调控作用。SBF 包括 Swi6 亚单元和 DNA 结合蛋白 Swi4。在 G_1 期和 S 期，本模型显示，Swi4 各对4个基因有作用；作用的性质有抑制也有促进。在 M 期对所选 13 个基因没有调控作用，这说明了 Swi4 的调节作用主要在 G_1 和 S 期。

在基因 Swi4、Swe1、Ndd1、Svs1 和 Hcm1 中都包含有 SBF 结合位点。在 S 期，Swi4 对 Ndd1 的转录起正调节作用，这与生物学试验中 Swi4 是转录调控因子 SBF 的 DNA 结合亚单元相吻合。

Spo12 基因对细胞从 M 进入减数分裂期起着正调控的作用，但其分子机制不明确。Spo12 的调节作用如表 2 所示，在 S 期无作用。

表 2 Spo12 基因调控情况

细胞周期	正调节	负调节
G_1	Aga1、Ho、Hta1、Htb1	Pcl2
G_2	Fus1、Hht1	Aga1、Clb4、Hcm1、Ndd1、Pcl2
M	Fus1、Hcm1、Ndd1	Clb4、Far1、Htb1

Swe1 的分子功能是调节细胞周期蛋白依赖性激酶的活性，它在 G_1 期突然表达。我们模型表明 Swe1 在 G_1 期对 Aga1、Ndd1 有负调控作用，对 Fus1、Ho 和 Svs1 基因有正调节作用。

Clb 类的基因编码的细胞周期蛋白对于 G_2/M 期的基因表达有很大的影响，其中 Clb2 最为典型。Mcm1、Fkh1/2 和 Ndd1 黏附 Clb2 启动子的，是 Clb2 表达的调控因子。有文献表明 S 期缺乏 Clb2 的细胞，Clb3 和 Clb4 也不表达，Clb2 可能启动 Clb3、Clb4 两基因的转录，这一点也在本模型中体现出来（Clb2 对 Clb3 和 Clb4 的作用为促进作用。

另一个重要的细胞周期调控因子是 Sic1。根据文献报道，当 Cln2 相关联的激酶活性在开始期升高时，Sic1 被磷酸化，生成的 Sic1p 被 Cdc4 呈递到 SCF（Skp1、Cdc34、Cdc53 和 1 个 F 盒容纳蛋白形成的复合体），形成遍在蛋白质参与蛋白的水解。Sic1 的表达在 M/G_1 期到达最高，Sic1 的调控作用如表 3。

表 3 Scil 基因调控情况

细胞周期	正调控	负调控
M	Ndd1	Aga1、Clb4、Ho
G_1	Clb3、Hht1	Clb4

Swi5 黏附细胞周期调控蛋白 Cln3，从而对细胞周期的调控起重要作用。同时，Swi5 在 G_1 晚期和 S 早期参与 Ho 基因的转录活化。我们模型结果显示 20、30、40 时间点 Swi5 对 Ho 是抑制作用（－1），而 100、110 时点且是促进作用（＋1），这可能与不同时点取样误差有关。此外，Sim1 是一个同 Swi5 有协同作用的调控子，它们一同使 Cln3 的表达上调。Sim1 作用的可能的分子机制是参与 DNA 复制的调节。

3 个目的基因 Hta1、Htb1 和 Hht1 被归为组蛋白类，它们的表达有很强的保守性，仅仅在 S 期到达高峰。在 S 期，3 个组蛋白受调控的情况如表 4。

表 4 S 期组蛋白受调控情况

目的基因	CLB2	HSL7	SIC1	SIM1	SWI5
HTA1	－1	0	1	0	0
HTB1	0	1	1	1	－1
HHT1	－1	0	1	1	1

在组蛋白基因都存在 SBF 的结合位点，Swi4 应对组蛋白的转录有影响，但我们的模型并不支持这个结论，可能是由于协同作用消除了 Swi4 的调控能力。

Ste12 编码转录调控因子，在 Ste12 缺乏的细胞中，Fus1 和 Aga1 在 G_1 期的表达受到影响以致没有表达，可知 Ste12 对这两种基因的表达存在正调控作用。模型分期结果可见 Ste12 对这 2 个基因都没有作用，但分别考察不同的时间点可见，Ste12 在 5 个 G_1 时间点中有 3 个为正调控。G_2/M 期，Far1 的表达受到 Mcm1 的影响，Mcm1 的表达产物结合 Far1 转录活化域，可使 Far1 的表达增加，同时 Ste12 对 Far1 的 G_1 期表达有影响，可见 Ste12 和 Mcm1 分别对 Far1 的表达进行调控，分期结果表明 G_2 期 Mcm1 对 Far1 为正调控作用，M 期为负调控，而 Ste12 在 G_1 对 Far1 基因没用作用。Hsl7 是 Swe1 在 G_1 晚期的负调控因子，我们的模型明显表明了这一点，与试验结果吻合。

3 讨论

近些年的酵母菌基因表达调控的研究已有大量的文献报道，因此我们对本研究建立的调控网络进行了生物学验证，进而对生物动力学模型的有效性进行了评价。

从我们的结果来看，本动力学方程模型分析结果基本与生物学实际相符合，能够准确、高效地描述酵母菌基因表达调控关系。因此，具有重要的应用价值。

参考文献：

[1] Tamayo P, Slonim D, Mesirov J, *et al*. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation [J]. Proc Natl Acad Sci U S A, 1999, 96(6): 2907－2912.

[2] Somogyi R, Sniegoski C A. Modeling the complexity of Genetic Networks: Understanding Multigenic and Pleiotropic Regulation[J]. Complexity, 1996, 1: 45－63.

[3] Liang S, Fuhrman S, Somogyi R. Reveal, A general reverse engineering algorithm for inference of genetic network architectures [J]. Pac Symp Biocomput, 1998, 3: 18－29.

[4] Thieffy D, Thmos R. Qualitative analysis of gene networks [J]. Pac Symp Biocomput, 1998, 3: 77－88.

[5] McAdms H H, Shapiro L. Circuit simulation of genetic networks [J]. Science, 1995, 269(5224): 650－656.

[6] Chen T, He H L, Church G M. Modeling gene expression with differential equations [J]. Pac Symp Biocomput, 1999, 4: 29－40.

[7] de-Hoon M J, Imoto S, Kobayashi K, *et al*. Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations [J]. Pac Symp Biocomput, 2003, 8: 17－28.

[8] 陈兰荪，宋新宇，陆征一. 数学生态学模型和研究方法[M]. 成都：四川科学技术出版社，2003：5－25.

（编辑 张大春）