

基因网络研究进展*

彭华正 潘建伟 朱睦元**

(浙江大学生命科学学院, 杭州 310012)

摘要 分子生物学的深入发展揭示了复杂的生命现象是大量基因相互作用的结果, 传统的以描述为主的生物学和分解分析的研究方法受到挑战. 随着 DNA 芯片和分子阵列技术的应用, 快速检测生物基因组的表达已成为可能. 在生命科学领域, 基因网络作为一种系统的、定量的研究方法正在受到重视, 该方法建立在分子生物学、非线性数学和信息学等多学科交叉的基础上. 基因网络是动力系统模型, 具有稳定性、层次性等一系列非线性系统的特性. 通过基因表达的大量数据, 结合一定的分析和计算方法可以构建合适的基因网络拓扑结构模拟系统的行为. 反过来, 利用已建立的基因网络可以指导进一步的实验. 计算机工具和 Internet 资源是基因网络研究的重要手段. 基因网络研究将在后基因组研究中发挥重要的作用.

关键词 基因网络, 基因表达, 发育模拟

学科分类号 Q61

1953 年, Watson 和 Crick 发现了 DNA 的双螺旋结构, 以此开创了分子生物学的新时代. 从分子水平上研究生物的生长、发育等现象, 使科学家认识到基因调控的重要意义. 受技术条件的限制, 早期的科学家只能描述性地研究少数几个基因在特定生命现象中所起的作用. 随着 DNA 芯片和分子阵列等技术的应用, 科学家已能够在短时间内获得生物体基因表达的大量数据. 例如, 给出一个诸如酵母一样含有大约 6 000 个基因的生物体, 一个科学家在几个月内就能测出其整个细胞周期内全部基因的表达水平. 现在, 不少模式生物的全部基因组序列已被或即将被全部测出. 这样, 随之而来的问题是: 零碎的、描述性的传统研究方法已难以深入, 科学家需要一种定量地从整体上解释实验结果的方法.

近年来, 对系统科学的研究正在推动生物学家从系统角度看待具有高度复杂性的生命现象. 生命系统是一个有内外相互作用的自组织系统, 即储存生命信息的各基因并不是孤立地发挥作用, 而是通过形成“基因网络”这样一个复杂系统来推动生命演化的. 因此, 一些科学家已经开始利用系统动力学的手段来研究基因网络, 并且也取得了较好的效果^[1].

1 基因网络的研究内容

1.1 基因网络的原理

以前, 由于技术水平的限制, 分子生物学家通

常采用分解分析的方法, 即根据基因及其产物是否对特定的表型起支配作用而进行分类, 这样一来, 在解释由几个基因及其产物共同作用的现象时便遇到了困难. 随着分子生物学研究的深入和分子阵列等大规模测定技术的应用, 我们已经能够从整个基因组规模, 在基因表达水平和蛋白质水平上认识生物系统的结构和动态信息, 这就使得上述困难更加突出. 从系统的观点看, 基因表达的大量数据正是反映了基因网络这样的复杂信息处理系统在分子层次上的运转. 因此, 我们可以用带有反馈回路的基因网络有效地处理它.

对某一物种或组织中全部基因的表达关系进行整体性研究是基因网络理论的重要目的. 在研究过程中, 首先是按照同步或反同步表达, 以及表达强度的变化, 系统地识别各基因的特点, 再用聚类的方法将各基因归类, 在此基础上构建基因调控网络, 分析相关的控制参数^[2]. 要构建一个合适的基因网络框架. 并非需要非常完整的基因表达数据, 事实上基因网络的构建与实验数据的采集是相互作用的过程, 有时在实验数据不完善时, 可以适当引入假设^[1].

* 国家自然科学基金资助项目 (39770420, 30100115) 和浙江省自然科学基金资助项目 (300255).

** 通讯联系人.

Tel: 0571-88273325, E-mail: lsczhumy@mail.hz.zj.cn

收稿日期: 2001-01-10, 接受日期: 2001-02-23

在真核生物中, 整个基因组很可能是通过基因簇这一重要形式进行基因调节的, 即基因组通常由基因簇组成一个复杂的、紧密联系的控制系统. 这些基因簇往往有多个调节位点, 在这些调节位点上, 转录因子特异地与之结合. 例如, 在对海胆 (sea urchin) 的研究中, 便提供了这样一个分析材料. Davidson^[3]从基因组角度分析了海胆发育过程中转录的空间调控, 为进行基因网络的整体分析打下基础. 这种功能上相关基因组成的基因簇在生理上是否具有优势, 可能取决于信号分子的区域化和穿透能力, 这样就能使得各转录因子协同组成高度特异性的调节模块. 但这些方面仍需更多的了解.

1.2 基因网络的基本特性分析

基因网络在本质上是一个连续的动力系统, 即动力系统网络, 考虑到简化问题求解的需要, 有些基因网络可表示为离散型系统, 有些则可表示为布尔型的. 布尔型基因网络系统是指基因及其产物的表达状态只有“开-关”两种状态, 而所谓离散型网络是介于动力系统模型和布尔网络模型之间.

基因网络具有稳定性, 即基因网络系统能够通过自动调节达到稳定. 基因和其他生物分子构成的网络体系由许多生物化学参数确定, 在细胞的代谢、发育过程中这些参数势必发生变化. 而基因网络的拓扑结构要能够承受这种变化, 使得系统维持稳定. 不少学者认为, 基因网络系统中的自组织和负反馈回路限制了系统中各组分的波动范围, 从而为系统的稳定性提供保障. Becskei 和 Serrano^[4]通过在 *E. coli* 中构建由一个调节基因和转录模块组成的简单基因作用回路来说明这个问题. 应用这一理论, 人们可以解释某些基因网络在总体上具有的“开关”特性^[5]. 另外, Dassow 等^[1]用基因网络理论对果蝇体节发育进行模拟后, 认为该网络的一些主要拓扑结构的存在保证了这一发育模块的稳定性.

基因网络的稳定性也体现在生物体缓解突变的影响方面. 生物在发育和生理过程中抵御突变的影响主要体现在两个方面: 一是基因功能的重叠性, 即一个基因在突变中丧失的功能, 可以由一个或更多具有相似功能的基因所代偿, 以减弱这一突变对表型造成的影响; 另一种机理源于那些在功能上不相关基因之间的复杂的相互作用, 这种作用已经在生物代谢等系统中得到证明. 而在整个基因组范围内, 这个稳定性机理更加受到关注. Wagner^[6]以模式生物酿酒酵母 (*Saccharomyces cerevisiae*) 的

基因组数据为例, 证明在功能上无关基因之间的相互作用是系统抵抗突变影响的主要原因. 这很可能也是生物在进化中保持稳定性的原因.

基因网络有着高度的复杂性. 生物具有的大量基因可能组成各个模块, 而不同的基因网络模块可以在不同层次上发生相互作用, 即便在同一基因网络中, 同一个基因可能参与各种不同的分子机理, 即有不同的蛋白质因子和顺式作用元件加入^[7]. 这样一来, 基因活化就涉及了许多不同作用元件之间的相互作用, 或者组成复杂的复合物, 或者相互之间有竞争作用. 因此, 进一步搞清分子网络中各成分的相互关系对于建立合理的基因网络模型是十分重要的.

2 基因网络的应用

基因网络系统研究的一个重要方面是定量地描述生物个体的发育过程. 目前, 科学家已经对一些模式生物的生长发育现象进行了基因网络的分析和模拟, 涉及微生物, 动物和植物, 并且得到了一些重要的结论. 值得注意的是, 由于基因网络研究尚处于探索阶段, 在对少量基因进行分析时, 似乎传统的定性描述方法也可适用, 但当涉及到大规模基因表达时 (这正是基因网络研究的重要目的), 必须借助于合理的算法和计算机强大的计算能力.

果蝇 (*Drosophila melanogaster*) 体节发育的模拟. 在果蝇受精卵当中, 母性成形成基因, 即母性 mRNA 往往控制着 Gap 基因的表达模式. 这些母性基因通过编码转录调节因子, 再由调节因子促成了配对基因的模式表达. 这一级联调控过程通常发生于受精卵形成多细胞之前. 分节极性基因的重复作用通过配对基因产物留下了印迹, 这样便形成了原始的体节. Dassow 等^[1]认为体节发育可能是模块式的, 各模块对外界的某种刺激作用有着固有的特征性反应. 因而, 进化只是对各模块在输入上进行重排, 而并不改变各模块固有的行为. 他们用计算机模拟了果蝇体节发育的过程, 并提出果蝇的分节极性基因组成了这样一个模块, 该模块的控制其行为的动力学常数具有稳定性.

拟南芥 (*Arabidopsis thaliana*) 花形态发生的基因调控分析. 拟南芥花的形成过程中涉及到大量基因的调控过程, Mendoza 等^[8]考察了其中 10 个基因的相互关系. 他们研究了野生型和突变体的 mRNA 表达模式, 推导出这一调节体系的网络拓扑结构, 然后建立了一个通用的由一系列逻辑表达

式构成的定性模型, 通过对这一模型中参数的约束, 解释了拟南芥四个花器官 (萼片, 花瓣, 雄蕊, 心皮) 分别所处的基因表达模式, 同时也解释了拟南芥无花时的基因表达状态. 利用这一模型还可以模拟和预测拟南芥的各种变异表型. 在模型分析的基础上, 他们预测了该基因网络存在的第六个稳定模式, 也得到了实验的确认. 而且, 他们通过对该模型进行动态分析后, 认为基因 *LFY* 存在不止一个调节位点, 而这种多位点调节很可能在拟南芥从不开花状态向开花状态过渡时起作用.

3 基因网络的研究工具

基因相互作用形成的网络可以用计算机语言来描述, 并进行计算机模拟. 计算机仿真语言以及一些数学处理工具包如 Matlab 等都可用于基因网络的分析. 随着 Internet 的兴起和生物信息学的发展, Java 语言因为其具有跨平台和适宜网络编程等特点, 逐渐受到生物工作者的青睐. Samsonova 等^[9]介绍了一个 Java 小应用程序 (Java Applet) “NetWork”, 它建立在布尔型网络模型基础上, 具有通用的 3W 网络建模界面. 特别值得一提的是 Dassow 等^[1]用 Java 2 开发的基因网络模拟器 “Ingeneue”, 该程序可以模拟果蝇体节发育过程中基因和蛋白质浓度的变化.

基因网络的研究还涉及存放基因相互作用的数据库. 这种数据库应不仅能够存储相关的基因序列, 还能存储基因的相互作用关系. 现在已有不少有关核酸和蛋白质方面的数据库, 但对于基因及蛋白质相互作用的网络资源还较少, 在一些模式生物的数据库网站上有一些这方面的数据^[10]. 近年来发展的面向对象程序设计思想的生物数据库, 在存放基因相互作用的数据库方面可能会有较好的效果^[11]. 这里, 值得一提的是俄罗斯科学院西伯利亚分院细胞学和遗传学研究所开发的 “GeneExpress” 系统^[12], 该系统提供了有关真核基因表达较为丰富的内容, 包括转录调节模块、位点活性预测模块、位点识别模块以及基因网络模块等.

4 基因网络的研究前景

基因网络的研究处于生物学 (主要是分子生物学)、数学 (主要是非线性数学) 和信息学 (主要是程序算法的设计) 三大学科的交叉点上, 是生物信息学 (bioinformatics) 的热点之一, 也是后基因组研究的重要内容. 基因网络的研究无疑有助于探

索生命现象的一些本质问题, 但也应该看到存在的问题, 基因网络构建的困难首先体现在从基因表达的数据反向推断网络的拓扑结构, 即所谓的 “逆向技术 (reverse engineering)”. 该技术要求设计出合适的算法, 并从数学上进行严密的分析和证明, 而且这样的算法还应该对数据采集的误差是不敏感的. 另外, 通过实验数据确定的网络结构及其参数的稳定性和可靠性到底如何, 还需要再进行实验的检验, 而真正复杂的模型对计算机的计算能力也是一个考验. 目前, 只是对少数模式生物和少量基因进行基因网络研究的尝试, 希望能将其中得出的有关结论用于更大规模的基因网络之中. 但是, 在理解和解释这些巨量的生物学数据, 以便建立预测性模型时 (这是任何一门科学的目的所在) 还存在很多挑战.

总之, 随着微生物到人类基因表达数据的飞速增长, 生物学不再仅仅是基于观察和实验的科学, 基因网络理论将在解释和预测生命现象中发挥积极的作用. 在我国, 描述生物学的根底深厚, 但生物信息学方面与国际前沿差距甚大, 特别是有关基因相互作用的基因网络理论更是很少研究, 这一点应当引起我国生物学工作者的关注.

参 考 文 献

- 1 Dassow G, Meir E, Munro E M, *et al.* The segment polarity network is a robust development module. *Nature*, 2000, **406** (13): 188~192
- 2 Thieffry D. From global expression data to gene networks. *Bioessays*, 1999, **21** (11): 895~899
- 3 Davidson E H. A view from the genome: Spatial control of transcription in sea urchin development. *Curr Opin Genet Dev*, 1999, **9** (5): 530~541
- 4 Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature*, 2000, **405** (6786): 590~593
- 5 Wolf D M, Eeckman F H. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J Theor Biol*, 1998, **195** (2): 167~186
- 6 Wagner A. Robustness against mutations in genetic networks of yeast. *Nature Genetics*, 2000, **24** (4): 355~361
- 7 Blaiseau P L, Thomas D. Multiple transcriptional activation complexes tether the yeast activator Met4 to DNA. *EMBO Journal*, 1998, **17** (21): 6327~6336
- 8 Mendoza L, Thieffry D, Alvarez-Buylla E R. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis. *Bioinformatics Oxford*, 1999, **15** (7~8): 593~606
- 9 Samsonova, M G, Serov V N. NetWork: an interactive interface to the tools for analysis of genetic network structure and dynamics. *Proceedings of the Pacific Symposium on Biocomputing '99*, http://www.csa.ru:81/Inst/gorb_dep/inbios/Dyn_bool/Dyn.htm, 1999
- 10 Brody T. The interactive fly: gene networks, development and the Internet. *Trends in Genetics*, 1999, **15** (8): 333~334

- 11 郝柏林, 张淑誉. 信息生物学手册. 上海: 上海科学技术出版社, 2000. 242~ 244
Hao B L, Zhang S Y. Handbook of Bioinformatics. Shanghai: Shanghai Science and Technology Press, 2000. 242~ 244
- 12 Kolchanov N A, Ponomarenko M P, Frolov A S, *et al.* Integrated databases and computer systems for studying eukaryotic gene expression. Bioinformatics, 1999, 15 (7~ 8): 669~ 686

The Progress of Gene Networks^{*}

PENG HuaZheng, PAN JiarWei, ZHU MuYuan^{**}

(Life Science College of Zhejiang University, Hangzhou 310012, China)

Abstract The progress of molecular biology attributes the complexity of life to the interaction of mass genes, and thus the traditional descriptive method in biology and analysis by disassembling faces the cruel challenge. The application of DNA chip and molecule array enables scientists to monitor the massively parallel expression of genes, which lead to emphasis of gene networks as a systematic, quantitative method in the research of life science. Based on the crossing of subjects on molecule biology, nonlinear maths and informatics, the object of gene networks is to construct the dynamics model which has nonlinear traits such as robustness, hierarchy and so on. The massive data of gene expression combined with fit algorithm can help scientists establish the topology of the interactions, by which the system behavior can be simulated. Contrariwise, the model established can direct the further experiments. To study gene networks the computer and Internet resource are very important. The study of gene networks will play an important role in the post-genome research.

Key words gene networks, gene expression, develop simulation

^{*} This work was supported by a grant from the National Science Foundation of China (39770420, 30100115) and the National Science Foundation of Zhejiang Province (300255).

^{**} Corresponding author. Tel: 86-571-88273325, E-mail: lsczhumy@mail.hz.zj.cn

Received: January 10, 2001 Accepted: February 23, 2001