

DOI:10.3724/SP.J.1008.2011.01009

酵母基因组尺度遗传相互作用网络的拓扑性质

崔幼一^{1*}, 张 卓²

1. 吉林铁道职业技术学院, 吉林 132001
2. 第二军医大学长海医院心血管内科, 上海 200433

[摘要] **目的** 分析酵母遗传相互作用的基因组尺度的总体性质。**方法** 用网络的图论方法, 计算酵母遗传相互作用网络的拓扑性质。**结果** 酵母全基因组的遗传相互作用中, 基因的相互作用个数服从幂律分布, 幂指数接近3。平均相互作用的基因为87, 约2/3的基因只需通过1个基因就有遗传相互作用, 约1/3只需通过2个基因就有遗传相互作用。遗传相互作用网络的平均聚集系数是0.047。**结论** 酵母的基因通常具有多种而不是单一的功能, 而且基因之间遗传相互作用高度密集。

[关键词] 酵母菌; 遗传相互作用; 基因组尺度; 拓扑学

[中图分类号] R 394.21 **[文献标志码]** A **[文章编号]** 0258-879X(2011)09-1009-04

Topological properties of genetic interaction network in yeast

CUI You-yi^{1*}, ZHANG Zhuo²

1. Jilin Railway Vocational and Technology College, Jilin 132001, China
2. Department of Cardiovasology, Changhai Hospital, Second Military Medical University, Shanghai 200433, China

[Abstract] **Objective** To analyze the topological properties of genomic-wide genetic interaction network in yeast. **Methods** The topological properties of genomic-wide genetic interaction network in yeast were calculated by the graph theory. **Results** A power law fit the degree distribution well in the genomic-wide genetic interaction network in yeast, with the exponent approaching 3. The Node's degree followed a wide-tail distribution, with the average degree being 87. Two thirds of the genes had genetic interaction via only one gene, and about one third of the genes had genetic interaction via only 2 genes. The average clustering coefficient was 0.047. **Conclusion** The gene function in yeast is usually multiple and the functional interaction between genes is highly condensed.

[Key words] yeasts; genetic interaction; genome-scale; topology

[Acad J Sec Mil Med Univ, 2011, 32(9):1009-1012]

功能基因组学是后基因组学时代的生物学研究的核心问题^[1]。解析基因组中各个成员的功能需要付出巨大的努力, 但更困难的是研究基因组中各成员之间复杂的相互作用。对于各种进化距离不同的生物, 其基因组的差别远远落后于它们表现出来的形态、生理、功能和适应性的差别。如果不只关注基因组的规模, 而考虑到基因组中各个成员的相互作用, 那么这种巨大的复杂度是可能理解的。目前, 关于遗传相互作用的研究大多还只关注个别基因之间的关系。在一些低等生物如线虫也开展了小规模遗传相互作用的测试^[2]。当然, 由于这一问题的高度复杂性, 目前的研究主要模式生物还是集

中在酵母上。酵母是单细胞真核生物, 包含有大约6 000个编码基因, 潜在的二元相互作用就有即3 000万~4 000万个。以前对遗传相互作用做过分析, 形成的网络包含了大约300个目的基因的结点, 虽然这已经是在其他物种不大可能完成的实验, 但是数据规模仍然相对小, 不足酵母整个基因组的1/20^[3]。最近, 利用合成遗传相互作用芯片技术, 首次大规模测量了酵母的遗传相互作用, 包含了1 670余个目的基因与4 400余个基因的相互作用, 大致代表了一个基因组尺度的二元相互作用网络^[4]。

网络的拓扑学分析对于遗传相互作用是重要的一步, 能从中得知一些普遍性质。由于网络的规模

[收稿日期] 2011-06-27

[接受日期] 2011-08-22

[作者简介] 崔幼一, 讲师。

* 通信作者(Corresponding author). Tel: 0432-4868955, E-mail: cuiyouyi@yahoo.cn

对于网络的拓扑学性质有重要的影响,显然,详细地分析一个大尺度的真实的遗传相互作用网络能给予我们一个关于生物复杂性的清楚的概念,而且对于高等生物也有重要的提示作用。本研究分析了酵母遗传相互作用网络的拓扑学性质,包括连通度分布、聚集性、最短路径,并估计了人类相互作用的情况。

1 方法和结果

1.1 资料来源 酵母遗传相互作用资料来源于文献[5]。

1.2 计算方法 关于鉴别是否具有遗传相互作用,我们采用了中等程度的阈值(intermediate cutoff, $|\epsilon| > 0.08, P < 0.05$)作为计算依据。后续的计算把相互作用网络用图表示,运用链表结构存储数据。拟合度分布时,采用最小二乘法计算幂指数。计算节点之间的距离时,采用广度优先搜索算法。计算程序用C语言编程,可以与作者联系后获取。

1.3 酵母遗传相互作用网络及其拓扑性质分析 合成遗传相互作用芯片技术(synthetic gene array, SGA)是用一张含有酵母全基因组所有成员突变体的芯片,每一个目标基因的突变体与SGA杂交,就产生了大规模的双突变体,它们蕴含了目标基因的相互作用信息[5]。目前SGA数据包含了1 674个待测的目标基因,4 419个相互作用基因,191 890个相互作用[4]。如果把每一个基因看成一个结点,遗传相互作用看成结点之间的一条边,我们研究的网络便具有4 419个结点和191 890连通度,在此基础上研究它们的拓扑学性质。

由于相互作用不表现出方向性,我们把网络转换成无向网络。因此,对于网络的结点 a 和 b ,如果 a 是 b 的上件,那么我们规定 b 也是 a 的上件。另外,由于网络包含了4 419个结点,而目的基因只有1 674个,目的基因数据必然比非目的基因密集,因此,下面所有的计算都分成只包含计算目的结点(称为目的集)和包含全部结点(称为全集)这两个集合分析。

1.3.1 结点的度 首先,我们分析了平均度,全集的度是87,而目的集是229。接着,观察了连接度分布,见图1。该分布是一个明显的宽尾分布。我们用指数分布来逼近,采用最小二乘法拟合数据,当计算全结点时得出 $p(k) \sim Ak^{-2.954}$;当计算目的结点时得出 $p(k) \sim Ak^{-1.998}$ 。这些结果说明,该网络可

能是一个小世界网络。

1.3.2 结点的聚集程度 一个结点的聚集系数定义为 $C = 2k'/k(k-1)$,其中, k 为结点的连接度, k' 为结点的连接点之间的连通度。 C 的平均值,即平均聚集系数,对于全集和目的集,分别是0.047和0.043,相差很小。 C 对连通度 k 作图的函数见图2。由图2可见,大多数结点聚集系数都较小。

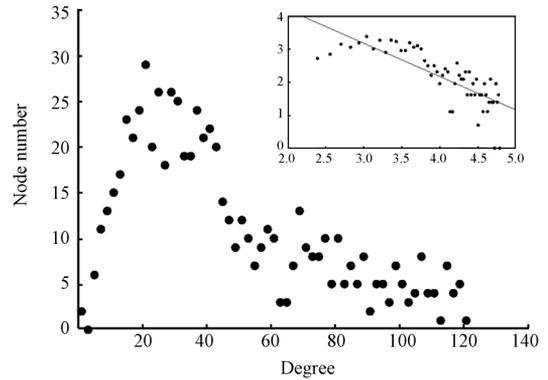


图1 结点的度分布曲线

Fig 1 Distribution profile of node's degree

The insert in the top right is the same distribution profile with logarithmic scale

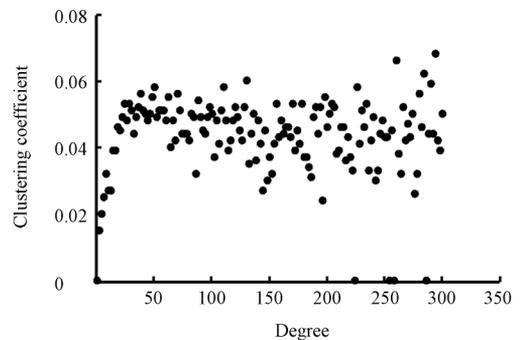


图2 结点的聚集系数的分布曲线

Fig 2 Distribution profile of clustering coefficient

The clustering coefficient is quite small, and the higher the nodes' degree, the lower the coefficient

1.3.3 结点间的最短距离 小世界网络的一个重要性质是结点之间是高度连通的,即最短路径超乎寻常地小[6]。我们采用广度优先搜索算法,计算了任意2个结点之间的最短距离。对于任意节点,它到自身的路径长度定义为0;对于有直接相互作用的节点,它们的距离则定义为1;如果2个节点需要通过一个节点才能连通,则距离定义为2,依此类推。对于全集,平均最短距离为2.323,目的集的是2.030。

1.4 酵母遗传相互作用网络图的高连通度和低聚

集性 现实世界中有多多种类型的网络具有“小世界”的性质, 它们的特征是连通度分布具有宽尾分布特征, 例如幂律分布; 而且结点之间的路径长度非常小, 是 $\log(n)$ 的数量级, 这里 n 是网络的规模, 即结点数目^[7]。很多生物分子网络是小世界网络, 比如前面提到的小规模的酵母遗传相互作用网络, 具有约 300 个目的结点, 整个网络约 1 000 个结点, 它们的平均最短路径是 3.3。按照已有的知识, 我们当前分析的遗传相互作用网络的规模在增大, 平均最短路径居然变得更小, 只有 2.3。这种超乎寻常的高度连通性可以这样来理解: 以前由于规模比较小, 绝大部分相互作用没有被测量到; 现在真实网络的连接度的增加远远超过结点数目的增加, 说明虽然以前网络的高度连通性受到注意, 但是仍然被低估。而本研究中, 相对于全集的平均最短路径 2.323, 目的集的最短路径竟然只有 2.030, 这也提示我们上述的推断是有依据的。为了进一步展示网络的高度连通性, 我们观察了最短路径的分布, 发现大部分的路径长度是 2 或者 3, 只有极少部分是 4 (图 3)。这一异常短的路径长度提示基因的相互作用非常密集和重要。

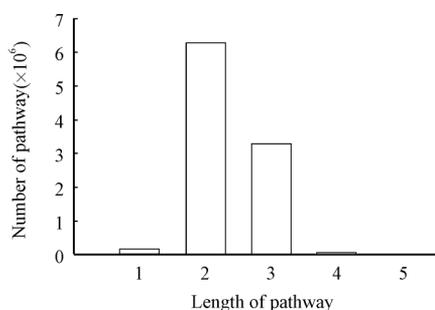


图 3 结点之间最短路径的分布曲线

Fig 3 Length distribution profile of shortest pathway between every pair of nodes

Among all paths, length is 2 for two thirds pathways and 3 for one third pathways, with 4 being the longest paths

另一方面, 结点的聚集性比较弱, 聚集系数仅仅为 0.04~0.05。以前的理论分析表明, 连通度越高, 聚集性越低^[8]。比如, 中心结点 hub 的聚集性一般低于非中心结点。的确, 我们也观察到, 目的集由于连通度比较高, 聚集性也稍微下降(从全集的 0.047 降到 0.043)。这个结果说明基因可能参与多种生物学过程, 因此聚集于同一类的情形很少。这提示我们即使对于研究相对较多的基因, 它们的相互作用和功能仍然需要深入挖掘。例如, p53 作

为抑癌基因, 是研究的“明星分子”, 可是最近仍然发现了它不依赖于转录的抗凋亡作用^[9]。

1.5 酵母遗传相互作用网络图拓扑性质分析示例 我们以一个基因为例仔细观察相互作用网络的拓扑性质。为了不失普遍性, 我们观察了具有典型度的基因。ELM1 与 283 个基因有直接相互作用(与目的集的平均度是 229 相比)。它与网络中的所有基因的平均距离只有 1.982。这说明, 一般而言, ELM1 只通过一个基因就能与基因组中的所有基因有联系, 这是高连通度的表现。另一方面, 它的低聚集性意味着它参与多种生物学过程。由于它的直接相互作用很多, 不能一一列出, 我们只观察与它有强相互作用的基因(stringent cutoff, $|\epsilon| > 0.20$, $P < 0.05$)。根据酵母基因组数据库(Saccharomyces genome database, <http://www.yeastgenome.org>) ELM1 编码一个丝氨酸/苏氨酸蛋白激酶, 它调控胞质分裂和细胞形态, 是出芽颈部环的一部分。它的相互作用“伙伴”中, ECM33、PER1、PTC1、UME6、SSD1、AXL2、MIH1、ZDS1、GAS1、ASE1、BEM3 编码的蛋白具有出芽生殖相关的活性和功能, 可以看作是 ELM1 有明显相关的一类基因。但是, 除此以外, GET2、BST1、PMR1、ERV14 的活性与囊泡运输相关, NUT1、PIB2、RRM3 具有调控基因表达的功能, HUR1 和 STB5 调节多药抗性和氧化压力, SNZ1 参与维生素 B6 的合成, PKP2 抑制丙酮酸脱氢酶活性, NBP2 与高渗通路相关, 它们与胞质分裂和细胞形态的相关性不是一目了然的。这些观察提示 ELM1 可能与上述多种生物学过程相关。YGL041C 和 YIL141W 是功能未知的两个基因, 即使它们跟 ELM1 具有强的相互作用, 想通过 ELM1 来预测 YGL041C 和 YIL141W 也是有风险的。

2 讨论

由于高等生物的遗传相互作用过于复杂, 目前没办法进行大规模的实验测定。这里我们借助于酵母的知识对人类基因的遗传相互作用进行一些粗略的估计。不过, 在进行估计之前, 需要指出, 由于酵母是单细胞生物, 而高等生物如人类是多细胞的具有复杂组织和器官结构的生物, 遗传相互作用会有差异。一方面, 由于高等生物是多细胞的, 基因可能因它们发挥作用的物理距离变远而使遗传相互作用变疏远; 另一方面高等生物具有高度进化的协调全身的系统——神经系统、内分泌系统和免疫系统,

基因的相互作用变得更复杂和密切。因此在不知道酵母和人类的真实的遗传相互作用差异之前,我们只能作出简略的推断。酵母基因的平均度为 87,如果人类基因组具有相似的相互作用性质,则 2.5 万个人类基因之间估计有 $87 \times 25\ 000$ 个,即约有 210 万个二元相互作用^[10]。实际上,按照酵母目的集的平均度 229 来估计,相互作用约有 600 万个。而对于人类基因网络,平均最短路径估计大致是 $\lg(25\ 000) \approx 4.4$ 。而按照我们现在的计算结果,在酵母中,最短路径甚至比 $\lg(n)$ 的估计值还短。由于这是全基因组的相互作用图谱,相对可靠。这些估计告诉我们相互作用蕴含了巨大的复杂性^[11],这方面的一个明显的例证就是常见的慢性疾病,比如高血压,被认为是多个基因的联合作用^[12]。即使我们知道了单个易感基因的贡献,它们的相互作用问题仍然会带来巨大的挑战。

[参考文献]

[1] Gondo Y. Trends in large-scale mouse mutagenesis: from genetics to functional genomics[J]. *Nat Rev Genet*, 2008, 9: 803-810.
 [2] Lehner B, Crombie C, Tischler J, Fortunato A, Fraser A G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways

[J]. *Nat Genet*, 2006, 38: 896-903.
 [3] Tong A H, Lesage G, Bader G D, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network[J]. *Science*, 2004, 303: 808-813.
 [4] Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear E D, Sevier C S, et al. The genetic landscape of a cell[J]. *Science*, 2010, 327: 425-431.
 [5] Boone C, Bussey H, Andrews B J. Exploring genetic interactions and networks with yeast[J]. *Nat Rev Genet*, 2007, 8: 437-449.
 [6] Albert R. Scale-free networks in cell biology[J]. *J Cell Sci*, 2005, 118(Pt 21): 4947-4957.
 [7] Barabási A L. Scale-free networks: a decade and beyond[J]. *Science*, 2009, 325: 412-413.
 [8] Barabási A L, Bonabeau E. Scale-free networks[J]. *Sci Am*, 2003, 288: 60-69.
 [9] Levine A J, Oren M. The first 30 years of p53: growing ever more complex[J]. *Nat Rev Cancer*, 2009, 9: 749-758.
 [10] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome [J]. *Nature*, 2004, 431: 931-945.
 [11] Battle A, Jonikas M C, Walter P, Weissman J S, Koller D. Automated identification of pathways from quantitative genetic interaction data[J]. *Mol Syst Biol*, 2010, 6: 379.
 [12] Lee M S, Choi T Y, Shin B C, Kim J I, Nam S S. Cupping for hypertension: a systematic review [J]. *Clin Exp Hypertens*, 2010, 32: 423-425.

[本文编辑] 孙 岩

· 书 讯 ·

《核与辐射突发事件知识百问》已出版

《核与辐射突发事件知识百问》由蔡建明、李雨主编,第二军医大学出版社出版,ISBN 978-7-5481-0116-1,定价 28.00 元。

核与辐射突发事件是一种意外发生的涉及核与辐射,对社会的稳定与公众的健康和安全、对环境、对国家和私人财产等具有重大危害的大事件。公众如果缺乏对核与辐射突发事件处置常识的了解,一旦发生核与辐射突发事件,即使该事件未造成人员伤亡,也会造成重大社会恐慌,演变成重大社会事件。日本福岛核电站事故表明,对公众进行必要的核辐射突发事件与辐射恐怖事件应对宣传是必要的。

《核与辐射突发事件知识百问》从原子与辐射、辐射的生物效应及其对健康的影响、核恐怖活动与核武器、辐射防护与核意外急救等几个方面向公众宣传相关知识,以提高公众有关的防范意识和应对能力。

本书由第二军医大学出版社发行科发行,全国各大书店均有销售。

通讯地址:上海市翔殷路 800 号,邮编:200433

邮购电话:021-65344595,65493093

<http://www.smmup.com>