

基因间的同源关系

苏杰¹ 姚杨^{1△} 黄原² 刘凯歌¹

(1 西安医学院第一附属医院 陕西 西安 710077 ;2 陕西师范大学生命科学学院 陕西 西安 710062)

摘要 同源是指从共同祖先的特性遗传下来的通常带有分歧的两个特征之间的关系。同源概念组成了进化基因组学的基础并对功能基因组学有巨大作用,但基于对同源概念的不准确理解,当前对其有诸多模糊表述,因此了解其确切含义具有重要意义。本文就同源、直系同源和旁系同源的概念和性质进行综述。

关键词 :同源 ;直系同源 ;旁系同源

中图分类号 :Q75 Q78 **文献标识码** :A **文章编号** :1673-6273(2012)23-4552-03

Homology between Genes

SU Jie¹, YAO Yang^{1△}, HUANG Yuan², LIU Kai-ge¹

(1 The First Affiliated Hospital of xi'an medical University, Xi'an Shaaxi, 710077, China;

2 College of life sciences of Shaanxi Noemal University, Xi'an, Shaaxi, 710062, China)

ABSTRACT: The homology concept refers to the relationship of two divergent characteristics which were inherited from the common ancestor. It is the solid foundation for evolutionary genomics and plays an important role in functional genomics. However, more and more vague statements and derivative concept about the homology appeared resulting from the misunderstanding of homology. Thus it is becoming increasingly important to understand these terminologies correctly. This paper reviews the concepts and nature of homology, orthologous and paralogous.

Key words: Homology; Orthologs; Paralogs

Chinese Library Classification(CLC): Q75, Q78 **Document code**: A

Article ID: 1673-6273(2012)23-4552-03

把现代基因组学与近代生物学相联系是基因组学的一大突破,不同生命形式的全基因组比对分析也为基因之间的同源关系提出新的前景,通过基因组全序列比对,原则上能够重建每一个基因的进化历史,而基因的进化历史必须以一定形式呈现出来,如:垂直遗传、基因重复、基因丢失、水平转移和融合。其中,垂直遗传与基因重复被认为是基因组进化的首要事件。所以,同源关系、直系同源、旁系同源也成为进化生物学的核心定义。自1970年Fitch^[1]第一次提出同源、直系同源、旁系同源的概念以来,在基因组比较中这些概念被广泛使用,但对其含义的准确理解仍存在争议^[2]。本文就同源、直系同源和旁系同源的概念和性质等进行综述。

1 同源

1.1 同源概念的历史

1843年Richard Owen^[3]提出同源关系的概念:“不同动物中每个形式和功能多样性之下的相同器官。”他从相似物种的角度定义同源关系“某动物中的部分或全部器官与不同动物中的另一部分或全部器官有相似的功能。”他把同源关系归因于所有脊椎动物中相同结构原件的存在,但是忽略了同源器官共同起源的因素。在达尔文的《物种起源》出版之后,同源关系的概念得到了迅速的发展,Huxley^[4]在对达尔文著作的回顾中

首次把同源关系作为一种进化证据提出来。

1.2 正确理解同源概念

基于对同源概念的不准确理解,目前出现了许多模糊的表述,如,David Mount^[5]指出“同源关系是指从共同的祖先基因遗传下来的基因,具有相似序列是检验它们的证据。”尽管Mount正确的把同源作为一个进化事件来陈述,也没有使用“同源程度性”但是,他坚持把序列的相似性作为同源的证据。这样的表述并没有意义,判断同源基因的关键是能否找到它们的共同祖先,而不是他们现在的存在形式。另外,我们经常看到这样的表述^[6]“两个蛋白结构同源”,依据同源的概念我们可以理解成“两个蛋白的结构起源于共同的祖先。”但是两个蛋白三级结构的相似性并不总是表明他们有共同的起源。如果两个蛋白有相同的结构,另外一些统计学表明它们序列间有相似性,才说明它们有同源关系,这种情况,真正涉及同源关系的不是其结构而是序列相似。结构的相似性固然重要,但是我们不能因为两个蛋白的折叠方式一样就确定他们有共同的祖先,活性、结合位点的空间一致及稀有的结构,都能提高判断同源关系的准确性^[7,8]。

另外,在缺乏共同祖先的情况下生物特性也能进化成相同的结构和功能,称为结构或功能的趋同,趋同现象涉及已经在过去分化但是之后又变得非常相似的同源特征或非同源特征。非同源特征也能发生趋同,这种情况,他们更相似但不会变成同源基因^[9]。

1970年Fitch^[10]准确地把同源概念定义为“从共同的祖先基因遗传下来的通常带有分歧的两个特征之间的关系。”该特

作者简介 苏杰(1984-)男,本科,药师,主要研究方向:生物制药。

电话:029-84277733 E-mail: 275152970@qq.com

△通讯作者 姚杨 E-mail: yaoyang1220@sina.com.

(收稿日期 2011-11-20 接受日期 2011-12-17)

征可以表现在基因、结构、功能和行为等水平且都有特定的存在位置,同源是指特征的同源,而不是位置的同源。如A和B是两个特征的存在位点,所以说A和B占据了同源位点比A和B是同源特征的表述更加准确。两个特征或者分享共同的祖先或者不是,这种有或无的关系就是关于同源关系中最基本的事实。需要强调的是,同源的基因并没有相似性的要求,它们可能非常相似也可相异,这完全依赖于它们分歧的时间和进化的速率。另外,同源的基因并没有功能相似的要求,同源是有关进化事件的表述,而与功能无关^[11-12]。

1.3 同源和相似的区别

序列相似性是在进化不可知的情况下的一个数量,相似性不需要由一个共同的进化起源决定,相似程度可以衡量。而同源是描述一个特性的进化历史,除少数例外,一般都无法观测^[13],因此一些文献将“相似比例”错误的表述为“同源比例”。只有当一个基因的不同部分含有不同的进化历史,即基因发生融合时才适用“同源比例”。如:肠激酶的一部分与脂蛋白感受器的一部分是同源的,这两个蛋白都含有不同源的一部分。因此同源关系是帮助解释可观察到的相似性的进化根源的一种推理。但相关片断的比例在他们是否有共同祖先的判断中并没有意义。

2 直系同源基因

2.1 概念

直系同源基因指起源于不同物种最近共同祖先的一些基因。直系同源基因具有相近甚至相同的功能,由相似的途径调控,在不同的物种中扮演相似甚至相同的角色,因此在基因组序列的注释中,是最可靠的选择^[14-16]。直系同源基因的概念包括两个方面。

(1)需要一个共同的祖先基因是直系同源基因概念的核心。

(2)两个参照物中一个最近共同祖先基因的存在,而不是有很多祖先基因。

此概念以假定有共同祖先为前提,所以它的应用对于原核生物的高度水平转移似乎有很大难度(如图1)。

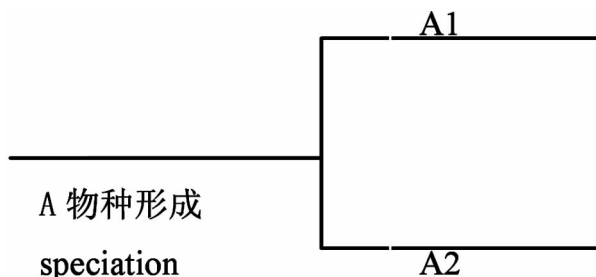


图1 直系同源

Fig.1 Orthologous gens

由于直系同源基因往往有相同或相似的功能,因此有人试图用功能的定义代替系统发生学的定义,认为直系同源基因是不同物种中具有相同或相似功能的基因^[17],另外,直系同源基因在序列上有很大的相似性,有人试图用序列的相似性来定义它,这都是不准确的^[18],并非所有直系同源基因都有相似的功能,也并非不同物种中序列相似性最高的两条序列就一定是直

系同源基因。它们相似与否,决定于它们进化的分歧率。它们也没有功能相似性的要求,同源是关于进化事件而不是功能的。

2.2 直系同源基因的性质

(1)直系同源基因是相互的。如图2中,Y是HA1的直系同源基因,反之亦然。

(2)直系同源基因是不可传递的。在图2中,HA1是Y的直系同源基因,HB也是Y的直系同源基因,但HA1和HB却不是直系同源的关系,而是一对旁系同源基因。

(3)直系同源基因不一定是一对一的关系,它可以是一对多或多对多的关系,因此认为某物种中每条基因在另一物种中都有一条直系同源基因与之对应,是错误的,这仅是一种情况。如图2中,HA1,HA2,HA3是Y的直系同源基因,这显然是一对多的关系,HA1,HA2,HA3和WA1WA2这两组基因互为直系同源,则这显然是多对多的关系。

3 旁系同源

旁系同源基因是指由一个共同祖先基因通过基因重复得到的基因^[19-20]。它的定义并不包括“旁系同源基因必须存在于同一个基因组内”或者“关于重复时间导致旁系同源基因出现”的说法。一些重复事件发生在生命进化的最早阶段,但各自的基因仍然是旁系同源基因。如图3,B1和B2是旁系同源基因。

有人把旁系同源定义为“在同一物种内由于基因复制而分离的同源基因”,这个说法是不全面的^[21],正确的定义为“由基因重复而产生的同源特性,即指由一个共同祖先基因通过重复而来的基因。”它可以存在于同一个物种内,也可以存在于不同物种中,这就引出了另两个概念,“内旁系同源”和“外旁系同源”,它们是根据复制事件与物种发生事件发生的先后顺序而提出的。对于一个给定的世系,在物种分化后发生基因重复事件产生的基因称为“内旁系同源”,反之称为“外旁系同源”。

我们使用直系同源与旁系同源描述基因产物之间的关系,通常喜欢把现有的词加以改造来定义新的环境,而不是再创造一个新词汇,这个术语起源于进化生物学领域,并且涉及与物种或基因重复相关的序列差异,但都没有涉及功能方面,所以正确使用这两个概念需要掌握产生生物功能差异进化途径的详细知识,Jense^[22]指出:“决定直系同源还是旁系同源最好的方法就是把任一基因追溯到它们结合的地方。”直系同源和旁系同源分别是由于祖先基因的垂直遗传和基因重复进化而来的。所以,把任何一对基因通过追溯到他们的结合点,或者是一个颠倒的“Y”(直系同源基因)或者是一条水平线(旁系同源基因)。

如图2所示:发生了两次物种事件和一次基因重复事件,形成了三个物种的八个基因,“人类”基因和线虫基因都是酵母菌Y的直系同源基因,它们都是由第一次物种分化而形成的。反之亦然,“人类”的基因HA1,HA2,HA3和线虫的基因WA1,WA2互为直系同源基因,因为他们是由二次物种分化而形成的,即:在一个物种中因基因重复产生了一套基因,而它们又都与另一个物种中的基因为直系同源关系,则由基因重复产生的一套基因。HA1,HA2,HA3互为内旁系同源基因,WA1和WA2也互为内旁系同源基因。如前所述,即基因重复发生在物种形

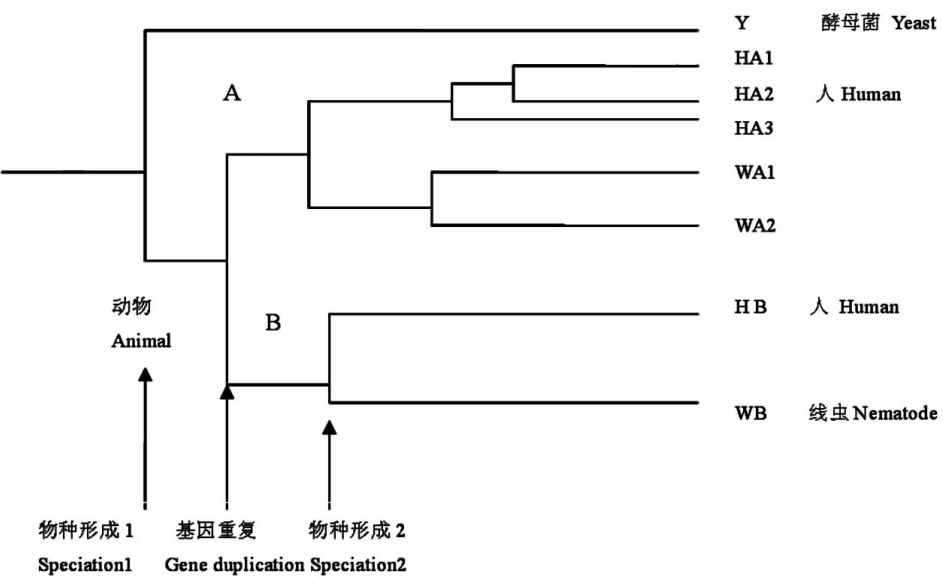


图 2 3 个物种 8 个基因的同源关系
Fig.2 Homologous relationships of eight genes of three species

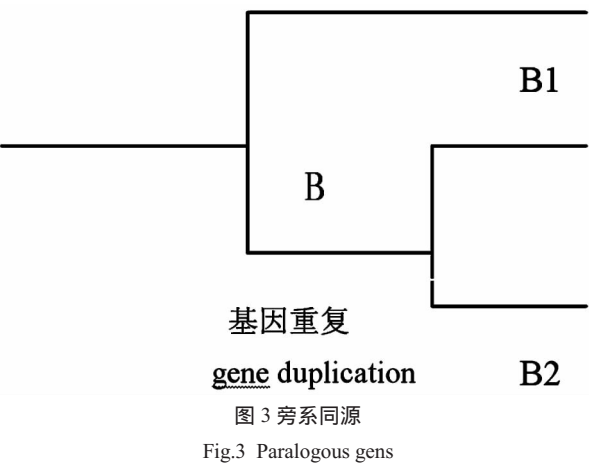


图 3 旁系同源
Fig.3 Paralogous gens

成之后,由于形成晚,所以相似度一般较高,相应地,HA1和HB是一对外旁系同源基因,因为基因重复产生A和B的过程发生在“人”和“线虫”的物种形成之前。当进化中出现基因横向转移,趋同进化,基因丢失或基因融合时,以上方法对于正确确定基因间的同源关系是不够的。如,分析在高度分化的蛋白中结构蛋白功能的相关性,并不能从以上方法中获益。

4 直系同源与旁系同源概念的推广

直系同源,旁系同源的概念不仅适用于基因,它还可以应用于几个基因,甚至单个核苷酸。如,通过内在基因的复制而产生的染色体片段是同源的基因范围,在两个物种中,在他们共同祖先之后仍持续完整的序列也能被叫做直系同源基因范围^[23]。原则上说,可以把直系同源的概念推广到任何单独的领域,从任何长度的核酸序列到单个碱基,但基本的描述仍然相同,“两个物种的基因组元件是从他们最近的共同祖先的基因组中相同的基因元件遗传而来的,则它们被认为是同源的。”更广的范围来说,只要每个基因组中碱基对相应的直系同源和旁系同源关系被确定,我们就能推测出基因组每个位点的进化关系。但这个目的,只有在近缘基因组间才能实现,在远缘物种间似

乎不能实现。

5 小结

同源关系指出了任何有共同祖先的实体间的一种关系,而不是有关进化事件的一个详述。该实体特别是基因,被称为同源基因^[24]。直系同源与旁系同源的概念是进化学与功能基因组学最基本概念的核心部分,由此派生的概念如内旁系同源和外旁系同源,虽然还没有广泛使用,但有助于准确且简洁的描述基因功能的多样性及进化过程。直系同源是指与物种事件相关的基因,旁系同源是与基因重复相关的基因。把物种事件与基因重复相关联的是水平转移和基因丢失,从而把它们放在一个复杂的关系网中。直系同源与旁系同源之间的明显不同主要用于基因进化分类的构建和一个新基因组序列功能的可靠注释。

参考文献(References)

[1] Walter M Fitch. Distinguishing Homologous from Analogous Proteins [J]. Systematic Biology,1970,19(2):99-113

[2] Leszek P. Pryszcz, Jaime Huerta-Cepas, et al.MetaPhOrs: orthology and paralogy predictions from multiple phylogenetic evidence using a consistency-based confidence score [J]. Oxford Journals,2010,39(5): 32-34

[3] Walter M. Fitch . Homology: a personal view on some of the problems [J]. Trends in Genetics,2000,16(5):227-231

[4] Henryk Szarski. The concept of homology in the light of the comparative anatomy of vertebrates[J]. The Quarterly Review of Biology,1949,24(2):124-127

[5] Caleb Webber, Chris P. Genes and homology[J]. Curr Biol, 2004,4(9): 322-323

[6] Roy A Jensen. Orthologs and paralogs-we need to get it right [J]. Genome Biol,2001,2(8):21-24

[7] Alan Boyden. Homology and Analogy. A Critical Review of the Meanings and Implications of these Concepts in Biology[J]. American Midland Naturalist,1947,37(3):648-650

(下转第 4587 页)

杂志 2010,7(30) 264-266

Qu Shen, Liu Qing. Nuclear transcription factor FOXO1 and glucose, lipid metabolism [J]. International Journal of Endocrinology and Metabolism,2010,7(30):264-266

[14] Kamagate A, Qu S, Perdomo G, et al. Foxo1 mediates insulin-dependent regulation of hepatic VLDL production in mice[J]. J Clin Invest, 2008,118(6):2347-2364

[15] Wende AR, Huss JM, Schaefer PJ, et al. PGC-1 α coactivates PDK4 gene expression via the orphan nuclear receptor ERR α : a mechanism for transcriptional control of muscle glucose metabolism[J]. Mol Cell Biol,2005,25(24):10684-10694

[16] Li X, Monks B, Ge Q, et al. Akt/PKB regulates hepatic metabolism by directly inhibiting PGC-1 α transcription Coactivator[J]. Nature,2007,21,447(7147):1012-1016

[17] Barthel A. FOXO proteins in insulin action and metabolism. Trends Endocrinol Matab,2005,16:183-189

[18] Gross DN, Wan M, Birnbaum MJ. The role of FOXO in the regulation of metabolism[J]. Curr Diab Rep,2009,9 (3):208-214

[19] Kamagate A, Qu S, Perdomo G, et al. FoxO1 mediates insulin-dependent regulation of hepatic VLDL production in mice[J]. J Clin Invest, 2008,118(6):2347-2364

[20] Galbo T, Olsen GS, Quistorff B, Nishimura E. Free Fatty Acid-Induced PP2A Hyperactivity Selectively Impairs Hepatic Insulin Action on Glucose Metabolism[J]. PLoS One,2011,6(11):e27424

[21] Altomonte J, Cong L, Harbaran S, et al. Foxo1 mediates insulin action on ApoC- and triglyceride metabolism [J]. J Invest,2004,114 (10):1493-1503

[22] Kamagate A, Qu S, Perdomo G, et al. Foxo1 mediates insulin-dependent regulation of hepatic VLDL production in mice[J]. J Clin Invest,2008,118(6):2347-2364

[23] 卢忠燕,甘立霞,缪洪明,等. FOXO1在胰岛 β 细胞中的表达及对增殖凋亡功能的影响[J]. 中国生物化学与分子生物学报,2009,25 (1) :50-56

Lu Zhong-Yan, Gan Li-Xia, Miu Hong-Ming, et al. The influence of Foxo1 in the expression and proliferation of islet β cell [J]. Chinese Journal of Biochemistry and Molecular Biology,2009,25(1):50-56

[24] Eduardo R. Roparll, Jose R. Pauli, Dennys E. Cintral, et al. Acute exercise modulates the FOXO1/PGC-1 α pathway in the liver of diet-induced obesity rates [J]. The Journal of Physiology,2009,587 (9): 2069-2076

[25] Avid c. Bouck, PetEr SHu, Jimmy cuI, AnAng SHELAt, and tAOS-HEng cHEn. A High-content Screen identifies inhibitors of nuclear Export of Forkhead transcription Factors[J]. J Biomol Screen,2011,16 (391):394-404

[26] Paola Dongiovanni, Luca Valenti, Anna Ludovica Fracanzani. Iron Depletion by Deferoxamine Up-Regulates Glucose Uptake and Insulin Signaling in Hepatoma Cells and in Rat Liver[J]. The American Journal of Pathology,2008,172(3):738-746

[27] Bai L. Iron overload and liver disease[J]. Journal of Southern Medical University,2002,22(4):370-371

[28] Claudio T. De Souza, Marisa J.S. Frederico, Gabrielle da Luz, et al. Acute exercise reduces hepatic glucose production through inhibition of the Foxo1/HNF-4 α pathway in insulin resistant mice [J]. J Physiol 2010,588(12),2239-2253

(上接第 4554 页)

[8] Alan Boyden. Homology and analogy: a century after the definitions of "homologue" and "analogue" of Richard Owen [J]. The Quarterly Review of Biology,1943,18(3):228-230

[9] Eugene V. Koonin. Orthologs, Paralogs, and Evolutionary Genomics[J]. Genetics,2005,39:309-338

[10] Walter M. Fitch. Homology: a personal view on some of the problems[J]. Trends in Genetics,2000,16(5):227-231

[11] Mario G. de Pinna. Concepts and Tests of homology in the cladistic paradigm[J]. Cladistics,1991,7:67-94

[12] Peter Gibling. Graps, surfaces and homology[M]. Cambridge,2010:42-45

[13] Eugene V. Koonin. An apology for orthologs-or brave new memes[J]. Genome Biology,2001,2(4):51-54

[14] Kriston L. McGary, Tae Joo Park, John O. Woods, et al. Systematic discovery of nonobvious human disease models through orthologous phenotypes[J]. PNAS,2011,108(43):6544-6549

[15] J. Muller. P. Julien. S. Powell, et al. Extending the evolutionary genealogy of genes with enhanced non-supervised orthologous groups, species and functional annotations [J]. Nucleic Acids Research,2011,38 (1):190-195

[16] Gogarten JP, Olendzenski L. Orthologs, paralogs and genome comparisons [J]. Current Opinion in Genetics & Development,1999,9(6):

630-636

[17] Li Li, Christian J, Stoeckert Jr, David S Roos. OrthoMCL: Identification of Ortholog Groups For Eukaryotic Genomes [J]. Genome Research,2003,13:2178-2189

[18] Arcady R Mushegian, et al. Large-scale Taxonomic Profiling of Eukaryotic Model organisms: A Comparison of Orthologous Proteins Encoded by the Human, Fly, Nematode, and Yeast Genomes [J]. Genome Research,1998,8:590-598

[19] Erik L. L. Sonnhammer. Orthology, paralogy and proposed classification for paralog subtypes[J]. Trends in Genetics, 2002,12,18(12):619-620

[20] S Yokoyama. W T Starmer and R Yokoyama. Paralogous origin of the red and green-sensitive visual pigment genes in vertebrates [J]. Molecular Biology and Evolution,2010,10(3):527-538

[21] Eugene V. Koonin. Orthologs, Paralogs, and Evolutionary genomics [J]. Annual Review of Genetics,2005,39:309-338

[22] Andrey Alexeyenko. Overview and comparison of ortholog databases [J]. Drug Discovery Today: Technologies,2006,(2):137-143

[23] Romain A. How confident can we be that orthologs are similar, but paralogs differ[J]. Trends in Genetics,2009,25(5):210-216

[24] Mado Remm. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons [J]. Journal of Molecular Biology,2001,14(5):1041-1052