

# 演化基因组学初介<sup>\*</sup>

Aaron J. W. HSUEH

Division of Reproductive Biology, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305-5317

**摘要** 近来,人类以及不同模式生物体全基因组测序取得的进展,为生物学研究提供了一个全新的视角。基于演化起源对直向性同源基因和旁向性同源基因序列的比较,可以更好地了解不同基因的生理作用。基因组学研究中的一个重要的方面,是了解全部基因的相互关系,进而最终可以了解基因的功能。本文提供了几个多肽配体及受体的比较基因组分析的范例。对于这个新生领域,本文建议中文名字为“演化基因组学”,以强调演化过程的非“进步”的特性及对不同生物体中所有基因的全面性分析 [动物学报 50(5):856-858,2004]。

**关键词** 演化 基因组学

## Introduction to evolutionary genomics<sup>\*</sup>

Aaron J. W. HSUEH

Division of Reproductive Biology, Department of Obstetrics and Gynecology, Stanford University School of Medicine, Stanford, CA 94305-5317

**Abstract** Recent advances in the sequencing of all genes in human and diverse model organism have provided a new perspective on biological research. Comparison of the sequences of orthologous and paralogous genes based on their evolutionary origins allows a better understanding of the physiological roles of individual genes. One of the most important aspects of genomic research is the understanding of the entire gene repertoire, thus eventually allowing a full understanding of gene functions. This article provides some examples of comparative genomic analyses of polypeptide ligands and their receptors. The Chinese name for the emerging field of “evolutionary genomics” is proposed here in order to emphasize the non-progressive nature of the evolutionary process and the global analyses of the complete gene sets in diverse organisms [Acta Zoologica Sinica 50(5):856-858,2004].

**Key words** Evolution, Genomics

在 20 世纪末 21 世纪初,生命科学经历了一场大革命,人类的基因组序列在 2001 年正式发表了 (Lander et al., 2001; Venter et al., 2001)。虽然已经确定了人类三万个基因的结构及推断出的蛋白质序列,但是我们对至少一半以上的人类基因的功能仍然一无所知。要了解人类基因的功能,一个重要的方法是研究基因生成的历史,这就要借重达尔文的生存竞争、自然淘汰学说。所以,对基因组的研究,不单单包括人类的图谱,各种模式生物,如酵母、线虫、果蝇、河豚、小鼠、水螅、猴子、牛和青蛙等图谱也相继发表或正在进行中。无可置疑,对不同物种直向性同源基因 (orthologs) 和旁向性同源基因 (paralogs) 的研究必将在生物

学上扮演一个重要的角色。

1858 年,达尔文在《Origin of Species》一书中发表了生存竞争、自然淘汰的理论 (Charles, 1858), 1897 年经严复的翻译而传入中国。虽然严复将《Origin of Species》一书译为《天演论》,但可能受日本学者译名的影响,中国学者后来将《天演论》命名为《进化论》; 随后又因为苏联李森科学说的影响,认为生物变异是有方向的发展; 另外,恩格斯也将“进化论”列为 19 世纪自然科学的三大发现之一,这些理论影响更使得“进化论”一词在中国普遍地应用在生物学上。事实上,达尔文的学说并不赞同“进化”的观点。1835 年,达尔文来到加拉帕戈斯群岛 (Galapagos), 这些小岛

2004-05-10 收稿, 2004-06-10 接受

\* 通讯作者 (Corresponding author). E-mail: aaron.hsueh@stanford.edu

© 2004 动物学报 Acta Zoologica Sinica

都是火山岛, 在地质上是新生的。在观察加拉帕戈斯群岛生态时, 他发现, 每个岛上的地莺 (Finch, 亦名达尔文莺) 的口喙形状都有所不同, 从而提出假说, 认为这些地莺都有一个共同的祖先, 从南美洲大陆漂洋过海来到这些不同的小岛上, 日久天长, 因为适应当地的食物而形成了不同形状的口喙, 从而衍生了这些不同种的地莺。正是不同岛上各具特点的地莺给达尔文以启迪, 使他摆脱圣经的创世说, 转而以科学视角来看待自然现象, 从而产生演化论。用达尔文的话说: “加拉帕戈斯群岛的生物决定了我的全部观点, 物种的改变基本上只是适应环境而已。”生物所存在的环境常有改变, 而物种的变异本无“进步”或者方向性的, 所以严复最早的“演化”译名才是正确的。

另一方面, 最近对 genomics 这个新词也有不同的翻译。genome 一词原是由 chromosome 一词衍生, 来代表生物体内所有的基因; 随后又有 proteome, transcriptome 等新创词来说明生物体内所有的蛋白质以及所有的转录物质。现在国内大部分生物学者将 genome 译为“基因组”, 但是这个译词不能强调 genomics 学科希望研究生物体内所有基因的“整体”概念。研究 genome 的价值, 绝大部分原因是因为我们知道了一个生物体内所有的基因, 所以“整体”的概念是非常重要的。若依照原来染色体词尾的翻译法, 应该将 genomics 称为“基因体学”, 而 proteomics 及 transcriptomics 就该称为“蛋白体学”及“转录体学”; 同样, 在研究激素与受体时, 配体体 (Ligandome) 和受体体 (Receptome) 等概念也就应运而生 (Ben-Shlomo et al., 2003)。因为只有知道了所有的基因, 有了这样整体的概念之后, 生物学者才可以用基因来解释所有的生命现象。

下面就以基于“基因体学”这个整体概念来寻找 GDF-9 受体的过程为例: 生长分化因子-9 (Growth differentiation factor-9, GDF-9) 是来自卵子的生长因子, 属于转化生长因子- $\beta$  超家族, 这个超家族还含有转化生长因子- $\alpha$  (Transforming growth factor- $\alpha$ , TGF- $\alpha$ )、激活素 (Activin) 和骨形成蛋白 (Bone morphogenetic proteins, BMPs) 等成员。TGF- $\beta$  家族的成员通过与 I 型和 II 型丝氨酸/苏氨酸激酶受体结合形成复合物, 激活 Smad 转录因子, 从而启动信号通路 (Kingsley, 1994; Wrana et al., 1994; Mathews and Vale, 1993)。基于序列的多样性, 这些受体被分成两类: I 型受体

和 II 型受体。II 型受体也称作激活素受体样激酶 (Activin receptor-like kinases, ALKs), 包括 ALK1-7 等 7 个成员。I 型受体包括 ACTR-2、ACTR-2B、BMPR-2、AMHR-2 和 TGFBR 等 5 个成员 (Vitt et al., 2001)。有关 TGF 蛋白和激活素的研究显示, 这些配体在与 I 型受体结合之前, 一般结合各自的 II 型受体, 导致 II 型受体磷酸化, 然后引起转录因子 Smads 的磷酸化 (Kingsley, 1994)。为了发现 GDF-9 的受体, 我们搜索了人类整个基因组, 并分析比较了 GDF-9 与 TGF 超家族其他配体成员的序列 (Vitt et al., 2001)。从演化的角度上看, GDF-9 与 BMP 蛋白具有更相近的同源性 (Vitt et al., 2002; Newfeldt et al., 1999)。在研究丝氨酸激酶受体的亚基因体时, 我们发现没有其他未知的 II 型受体存在, “基因体”研究的观念表示: GDF-9 必须结合已知存在的这 5 个 I 型受体, 引起 II 型受体磷酸化, 产生信号转导。经实验研究显示: BMPR 在 GDF-9 在颗粒细胞中信号转导过程中是必须的, BMPR 胞外结构域直接与 GDF-9 发生作用并阻止了 GDF-9 的功能 (Vitt et al., 2002)。同理, 我们发现人类基因体没有其他未知的 II 型 ALK 受体, 所以可以推断: 当 GDF-9 与 BMPR 结合后, 一定会与 I 型受体中 7 个已知成员之中的某些或某个成员结合。由此, 我们最后的研究可以集中于 7 个已知的 I 型受体上, 结果显示, GDF-9 可以通过 I 型受体 ALK5 以及 Smad2 和 Smad3 蛋白介导发生作用 (Mazerbourg et al., 2003)。用同样的亚基因体的观念, 我们进一步预测松弛素 (Relaxin) 和胰岛素样蛋白-3 (Insulin-like 3 protein, INSL-3) 只会作用在富含亮氨酸重复的 G 蛋白偶联受体 (Leucine-rich repeat-containing, Gprotein-coupled Receptors, LGRs) 中的 LGR7 和 LGR8 两个受体上 (Hsu et al., 2000, 2002)。这些研究验证了亚基因体的观念以及不同基因有共同祖先的演化学说。因为旁向性同源基因 (Paralogs) 经常保留相近的功能, 对特定基因功能研究的结果往往可以推广到这个基因的旁向性同源基因上, 所以亚基因体可以应用到对任何基因的研究上。

“演化学”及“基因体学”应该算是 19 世纪及 20 世纪生物学登峰造极的两大支柱。在 21 世纪, 这两门学派将进行整合, 成为“演化基因体学”, 使我们能够藉由演化的观念以更进一步了解生物的基因及功能。另一方面, 演化论的验证不再只是依

靠稀有的化石标本及粗略的形态变化,而是能够从物种所有基因及 DNA 中用统计方法找出物种变异的轨迹。上面的例子是应用“演化”和“基因体”的观念来寻找激素受体的几个典型的成功尝试,同以往寻找受体的常规手段相比,省时省力,而且一切清晰明了。知道每一个基因的演化历史是了解基因功能的一个捷径,对于这样一个正在开启的重要学科,我们应该有正确的译名。“演化”及“基因体”两个学科的整合将开启生物学上的新篇章,我们可以预见:在未来二、三十年内,中国的生物学者会利用演化及基因体的范例,对生物学做出更大的贡献。

**致 谢** 本文在本实验室张键 (Dr. Jian V. Zhang) 博士的协助下完成,特此感谢。

#### 参考文献 (References)

- Ben-Shlomo I, Yu Hsu S, Rauch R, Kowalski HW, Hsueh AJ, 2003. Signaling receptome: a genomic and evolutionary perspective of plasma membrane receptors involved in signal transduction. *Sci. STKE*. (187): RE9.
- Darwin C, 1858. *The Origin of Species*, New American Library, New York.
- Hsu SY, Kudo M, Chen T, Nakabayashi K, Bhalla A, van der Spek PJ, van Duin M, Hsueh AJ, 2000. The three subfamilies of leucine-rich repeat-containing G protein-coupled receptors (LGR): identification of LGR6 and LGR7 and the signaling mechanism for LGR7. *Mol. Endocrinol.* 14 (8): 1 257 - 1 271.
- Hsu SY, Nakabayashi K, Nishi S, Kumagai J, Kudo M, Sherwood OD, Hsueh AJ, 2002. Activation of orphan receptors by the hormone relaxin. *Science* 295 (5555): 671 - 674.
- Kingsley DM, 1994. The TGF $\beta$  superfamily: new members, new receptors, and new genetic tests of function in different organisms. *Genes Dev.* 8: 133 - 146.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McEwan K, 2001. Initial sequencing and analysis of the human genome. *Nature* 409: 860 - 921.
- Mathews LS, Vale WW, 1993. Molecular and functional characterization of activin receptors. *Receptor* 3: 173 - 181.
- Mazerbourg S, Klein C, Roh J, Kaivo-Oja N, Mottershead DG, Korchynski O, Ritvos O, Hsueh AJ, 2003. Growth differentiation factor-9 (GDF-9) signaling is mediated by the type I receptor ALK5. *Mol. Endocrinol.* Dec. 18 (Epub ahead of print)
- Newfeld SJ, Wisotzkey RG, Kumar S, 1999. Molecular evolution of a developmental pathway: phylogenetic analyses of transforming growth factor-beta family ligands, receptors and Smad signal transducers. *Genetics* 152: 783 - 795.
- Venter JC, Adams MD, Myers EW, Li PW, Mural, RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides, P, Ballew RM, Huxon DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, 2001. The sequence of the human genome. *Science* 291: 1 304 - 1 351.
- Vitt UA, Hsu SY, Hsueh AJ, 2001. Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules. *Mol. Endocrinol.* 15: 681 - 694.
- Vitt UA, Mazerbourg S, Klein C, Hsueh AJ, 2002. Bone morphogenetic protein receptor type I is a receptor for growth differentiation factor-9. *Biol. Reprod.* 67: 473 - 480.
- Wrana JL, Attisano L, Wieser R, Ventura F, Massague J, 1994. Mechanism of activation of the TGF $\beta$  receptor. *Nature* 370: 341 - 347.