

繁殖基因(*ESR*、*LH*、*LHR*、*PRLR*)编码区 生物信息学分析

张宇, 李祥龙*, 周荣燕, 李兰会, 王志刚

(河北农业大学 动物科技学院, 河北 保定 071000)

摘要: 为了研究各物种 *ESR*、*LH*、*LHR*、*PRLR* 基因进化的特点, 从 GenBank 下载了人、毛猩猩、黑猩猩、猕猴、欧洲兔、野猪、黄牛、大熊猫、小家鼠、仓鼠、褐家鼠 11 个物种的 139 条编码区序列(CDS), 采用生物信息学方法对 4 种基因进行分析。结果表明, 4 种基因均存在较丰富的遗传多样性。其中 *ESR* 基因单位长度多态位点数为 0.330 7, 单位长度突变位点数为 0.399 8, 核苷酸多样性最低, 同义替换率相对最高, 平均遗传距离最小。*LH* 基因单位长度多态位点数(0.484 8)与单位长度总突变位点(0.618 3)均最高, 核苷酸多样性和非同义替换率相对最高。依据 4 种基因遗传距离构建的聚类图的拓扑结构存在差异, 说明 4 种繁殖基因进化速率存在差异, *ESR* 基因进化速率相对较慢, *LH* 基因相对较快。

关键词: *ESR*; *LH*; *LHR*; *PRLR*; 生物信息学

中图分类号: Q811.4 文献标志码: A 文章编号: 1004-3268(2013)01-0114-04

Bioinformatics Analysis of the Coding Regions of *ESR*, *LH*, *LHR* and *PRLR* Genes

ZHANG Yu, LI Xiang-long*, ZHOU Rong-yan, LI Lan-hui, WANG Zhi-gang

(College of Animal Science and Technology, Agricultural University of Hebei, Baoding 071000, China)

Abstract: In order to study the different evolutionary characters of *ESR*, *LH*, *LHR* and *PRLR* genes, 139 sequences of the coding regions (CDS) from *Homo sapiens*, *Pongo abelii*, *Pan troglodytes*, *Macaca mulatta*, *Oryctolagus cuniculus*, *Sus scrofa*, *Bos taurus*, *Ailuropoda melanoleuca*, *Mus musculus*, *Cricetulus griseus* and *Rattus norvegicus* were downloaded from GenBank database and analyzed using bioinformatics methods. The result showed that there was rich genetic diversity of the 4 genes. The number of unit length polymorphic sites of *ESR* was 0.330 7 and the total number of mutation per base pair was 0.399 8. The *ESR* gene has lower genetic diversity and average genetic distance, but higher rate of synonymous substitution. While for the *LH* gene, it had richer genetic diversity and higher rate of non-synonymous substitution. So, it could be inferred that *ESR* may be more conservative and *LH* has higher evolutionary rate.

Key words: *ESR*; *LH*; *LHR*; *PRLR*; bioinformatics

繁殖性状是一个由微效多基因控制的性状, 研究相关基因与繁殖性能的关系有重要意义。雌激素受体(estrogen receptor, *ESR*)能影响雌激素基因在雌性脊椎动物组织的表达与调控, 对胚胎发育过程中乳腺的生长发育以及雌性繁殖周期中卵泡的生长发育

有重要作用^[1]。Rothschild^[2]证实, *ESR* 基因位点存在遗传变异, 其中的一个等位基因最早在梅山猪中发现, 与高繁殖力连锁显著。Hewitt 等^[3]研究表明, 敲除 *ESR* 基因的小鼠不能排卵, 促黄体生成素(LH)调节紊乱, 子宫对雌激素不敏感, 说明 *ESR* 在繁殖中有

收稿日期: 2012-07-30

作者简介: 张宇(1986-), 女, 河北唐山人, 在读硕士研究生, 研究方向: 动物繁殖调控。E-mail: miya-1104@163.com

* 通讯作者: 李祥龙(1963-), 男, 河北滦南人, 教授, 博士, 博士生导师, 主要从事动物遗传育种研究。

E-mail: lixianglongcn@yahoo.com

重要作用。*LH* 是一种糖蛋白类激素,其作用的发挥必须依赖于促黄体素受体(luteinizing hormone receptor, *LHR*),两者相结合激活 cAMP 系统,引起黄体激素的合成与分泌,从而参与哺乳动物受精卵着床、性腺发育等一系列生理活动的调节^[4-5]。研究表明,*LH* 由 2 个多肽链组成,分别称为 α 亚基和 β 亚基。 α 亚基为促性腺激素共有,而 β 亚基具有激素和种间特异性,决定了该种激素的特异性,本研究选用 β 亚型基因进行分析。*LHR* 是一种在动物生殖调控中发挥重要作用的受体蛋白。敲除 *LHR11* 外显子会严重影响小鼠的性发育成熟^[6],因此,*LHR* 有可能是控制动物繁殖的主效基因之一。催乳素是由垂体前叶细胞合成和分泌的一种蛋白质激素,广泛作用于动物的性腺和乳腺。催乳素要行使其功能必须与催乳素受体(prolactin receptor, *PRLR*)结合,通过受体作用于靶细胞,引起靶细胞的各种生理生化反应^[7]。

目前,在分子水平上对繁殖基因的研究还很有限,远远满足不了从分子水平上研究基因对繁殖性能影响机制的需要。同时也较少有研究者从基因全序列遗传距离的角度进行分析。为此,利用比较基因组学和生物信息学方法研究了 4 种繁殖基因编码区种间变异,以期探明 4 种基因的遗传分化速率,进而为相关繁殖基因的遗传学和进化分析奠定基础。

1 材料和方法

1.1 序列来源

从 NCBI 网站(<http://www.ncbi.nlm.nih.gov/GenBank>)中分别下载了人、毛猩猩、黑猩猩、猕猴、欧洲兔、野猪、黄牛、大熊猫、小家鼠、仓鼠、褐家鼠共 11 个物种同源性较高的 *ESR*、*LH*、*LHR*、*PRLR* 基因的 139 条 CDS 序列(表 1)。

表 1 不同物种基因的序列来源

| 物种 | 序列号 |
|--------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 人(<i>Homo sapiens</i>) | (<i>ESR</i> , 16) HQ692820. 1, MN__001437. 2, AF051427. 1, AB006590, NM__001214903. 1, NM__001214902. 1, HQ692821. 1, NM__001040276. 1, AK292370. 1, DQ838583. 1, DQ838582. 1, AF0605555. 1, AF051428. 1, AB209620. 1, AB006589. 1, DQ777076. 1 (<i>LH</i> , 30) NM__000894. 2, BC041054. 1, BC106724. 1, BC106723. 1, NM__000737. 3, BC160150. 1, AK291552. 1, NM__033183. 2, BC128603. 1, NM__033142. 1, BC106059. 1, BC051378. 1, NM__033043. 1, BC103971. 1, BC103970. 1, BC103969. 1, BC030994. 1, BC022796. 1, BT007573. 1, BT006890. 1, AY890611. 1, AY893091. 1, AY889866, AY892342. 1, AY892341. 1, BC069526, BC006290. 2, J00117. 1, BC126460. 1, NM__033378. 1, NM__033377. 1 (<i>LHR</i> , 5) NM__000233. 3, BC157028. 1, BC156303. 1, S57793. 1, M73746. 1 (<i>PRLR</i> , 8) AY893481. 1, AY892876. 1, AK313270, AM393574. 1 BC059392. 1, AM393450. 1, AB464070. 1, AY890411. 1 |
| 毛猩猩(<i>Pongo abelii</i>) | (<i>ESR</i> , 1) XM__002824836. 1 (<i>LH</i> , 4) XM002829529, XM002829530. 1, XM002829533. 1, XM002829531. 1 (<i>LHR</i> , 1) XM__002812041 (<i>PRLR</i> , 1) NM__001159791 |
| 黑猩猩(<i>Pan troglodytes</i>) | (<i>ESR</i> , 1) XM__001170310. 1 (<i>LH</i> , 2) XM__001071803. 1, XM__001150167 (<i>LHR</i> , 1) XM__003309005 (<i>PRLR</i> , 1) XM__001150064. 2 |
| 猕猴(<i>Macaca mulatta</i>) | (<i>ESR</i> , 1) XM__001101433. 1 (<i>LH</i> , 2) XM__002801337. 1, NM__001032928 (<i>LHR</i> , 1) XM__001114090. 2 (<i>PRLR</i> , 2) XM__001092241. 2, XM__002804335. 1 |
| 欧洲兔(<i>Oryctolagus cuniculus</i>) | (<i>ESR</i> , 1) XM__002719535 (<i>LH</i> , 4) NM__001082695. 1, AB235913, AB235914. 1, AF362079 (<i>LHR</i> , 1) XM__002709875. 1 (<i>PRLR</i> , 1) NM__001082231. 1 |
| 野猪(<i>Sus scrofa</i>) | (<i>ESR</i> , 2) EF195768. 1 (<i>LH</i> , 1) NM214080. 1 (<i>LHR</i> , 1) NM__214449 (<i>PRLR</i> , 5) AY308825, DQ157757. 1, AY308824. 1, NM__001001868. 1, HM209402. 1 |
| 黄牛(<i>Bos taurus</i>) | (<i>ESR</i> , 2) NM__174051. 3, Y18017. 1 (<i>LH</i> , 2) NM__173930. 1, XM__00178202. 3 (<i>LHR</i> , 4) U20504. 1, NM__174381. 1, AF491303. 1, U87230. 1 (<i>PRLR</i> , 1) NM__174155 |
| 大熊猫(<i>Ailuropoda melanoleuca</i>) | (<i>ESR</i> , 1) M__002914160 (<i>LH</i> , 2) AF488738. 1, AF448455. 1 (<i>LHR</i> , 1) XM__002912442. 1 (<i>PRLR</i> , 1) XM__002919428. 1 |
| 小家鼠(<i>Mus musculus</i>) | (<i>ESR</i> , 4) NM__207707. 1, BC145329. 1, BC141075. 1, NM__010157 (<i>LH</i> , 4) AF333067. 1, AK138425. 1, NM__00497. 2, BC160301. 1 (<i>LHR</i> , 1) NM__013582. 2 (<i>PRLR</i> , 6) BC005555. 1, BC006652, NM011169. 5, D10214. 1, X73372. 1, L13593. 1 |
| 仓鼠(<i>Cricetulus griseus</i>) | (<i>ESR</i> , 1) GU362643. 1 (<i>LH</i> , 1) NM__001246810. 1 (<i>LHR</i> , 1) XM__003499548 (<i>PRLR</i> , 1) XM__003504158. 1 |
| 褐家鼠(<i>Rattus norvegicus</i>) | (<i>ESR</i> , 5) AB190769. 1, AB190770. 1, AF0420 (<i>LH</i> , 5) D00576. 1, NM__001033975. 1, NM__012858. 2, U25653. 1, U25803. 1 (<i>LHR</i> , 2) NM__012978, M61212. 1 (<i>PRLR</i> , 1) NM__001034111. 1 |

1.2 分析方法

先利用 BioEdit 软件分别对 35 条 *ESR* 序列、57 条 *LH* 序列、19 条 *LHR* 序列、28 条 *PRLR* 序列的编码区进行比对分析,编辑并选取共有的编码区序列(长度分别为 1 273、427、1 069、851 bp)进行比较,利用 DanSP 5.0 软件对其进行遗传多样性分析。分别计算出各基因的单位长度多态位点总数、突变位点总数、种间核苷酸歧异度(D_{xy})和净遗传距离(Da)。再利用 MEGA 5.0 软件分别计算各基因中的所有序列的平均遗传距离(使用 Kimura-2-parameter model 模型)、同义替换速率均值、非同义替换速率均值及其之间比例[使用 Nei-Gojori method(Jukes-Cantor)模型],并用 UPGMA 方法进行种间聚类分析。

2 结果与分析

2.1 繁殖基因的多态位点、总突变位点及其多样性

表 2 列出了 *ESR*、*LH*、*LHR* 和 *PRLR* 基因的单位长度多态位点数、单位长度总突变位点数以及核苷酸多样性。由表 2 可以看出,*ESR* 基因具有相对最低的单位长度多态位点数(0.330 7)、单位长度总突变位点数(0.399 8)和核苷酸多样性(0.095 07),而 *LH* 基因具有相对最高的单位长度多态位点数(0.484 8)、单位长度总突变位点数(0.618 3)和核苷酸多样性(0.122 09)。结果表明,*ESR* 基因进化速率相对较慢,而 *LH* 基因进化速率相对较快^[8]。

表 2 不同基因多态位点及核苷酸多样性

| 基因 | 序列数/条 | 序列长度/bp | 单位长度多态位点数 | 单位长度总突变位点数 | 核苷酸多样性 |
|-------------|-------|---------|-----------|------------|----------|
| <i>ESR</i> | 35 | 1 273 | 0.330 7 | 0.399 8 | 0.095 07 |
| <i>LH</i> | 57 | 427 | 0.484 8 | 0.618 3 | 0.122 09 |
| <i>LHR</i> | 19 | 1 069 | 0.366 7 | 0.459 3 | 0.103 53 |
| <i>PRLR</i> | 28 | 851 | 0.371 3 | 0.437 1 | 0.120 66 |

2.2 繁殖基因同义替换和非同义替换

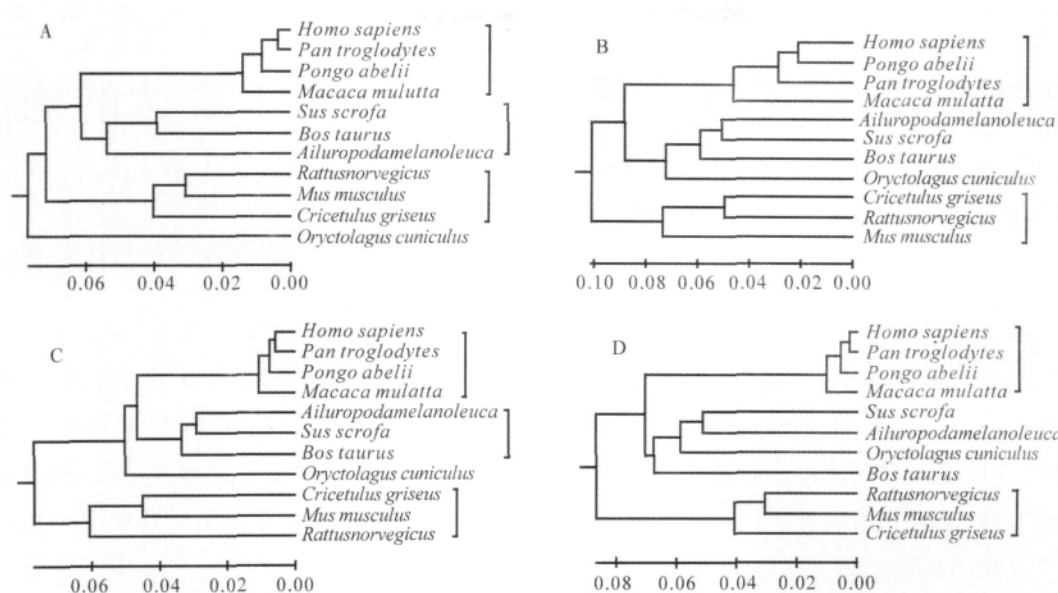
核苷酸替代是基因组进化的动力之一,从其替代发生的效率来看,不造成氨基酸改变的称为同义替换,导致氨基酸变化的称为非同义替换。同义替换速率可近似地表示分子的中性进化速率^[9]。在所分析的 35 条 *ESR* 序列、57 条 *LH* 序列、19 条 *LHR* 序列、28 条 *PRLR* 序列中,同义替换概率 $[k(s)]$ 分别为 0.416 57、0.124 75、0.072 86、0.155 54,非同义替换概率 $[k(a)]$ 分别为 0.037 72、0.142 68、0.121 13、0.124 20。数据表明,*ESR* 的同义替换速率要远远高于其他 3 种基因, $k(a)/k(s)$ 值为 0.090 55,远小于 1。*LH* 的非同义替换概率相对最

高。由此推测,*ESR* 基因在进化过程中可能承受了强烈的负选择压力^[10],这是由于 *ESR* 基因在繁殖性能上功能较为保守引起的。*LH* 和 *LHR* 基因的 $k(a)/k(s)$ 值均略大于 1,*PRLR* 的 $k(a)/k(s)$ 的比值略小于 1。说明 *PRLR* 基因受到了净化选择作用,*LH* 和 *LHR* 基因受到了正选择的作用^[11]。

2.3 不同基因的遗传分化

比较使用 Kimura-2-parameter 模型计算的 4 种基因物种间的平均遗传距离,*ESR* 的平均遗传距离最小(0.107 21),*LH* 的平均遗传距离最大(0.144 60)。表明 *ESR* 的各物种的遗传分化程度较小,较为保守,而 *LH* 分化程度较大。利用 MEGA 5.0 软件 UPGMA 方法构建的聚类图层次大体相似,11 个物种能较明显地聚为 3 组,并且均能较好地反映出各物种之间的亲缘关系。

依据 4 个基因遗传距离构建的聚类图可知(图 1),不同的组群有不同的分化程度。第 1 组均包括人、黑猩猩、毛猩猩和猕猴。除了 *LH* 基因是人与毛猩猩先聚合外(图 1-B),其他 3 个基因都按照人与黑猩猩先聚合,再和毛猩猩聚合,最后和猕猴聚合的顺序聚类(图 1-A、C、D)(其中 *ESR*、*LHR* 和 *PRLR* 基因人和黑猩猩的遗传距离分别为 0.006 63、0.011 79 和 0.004 85)。表明人与黑猩猩的亲缘关系最近,与毛猩猩和猕猴亲缘关系次之,毛猩猩和黑猩猩的 *LH* 基因在进化过程中与其他基因有所差异。聚类图的第 2 组中包括野猪、黄牛和大熊猫,其中 *LH* 基因和 *LHR* 基因的拓扑结构非常接近,都是野猪和大熊猫先聚合,再与黄牛聚合(图 1-B、C)。说明这些物种的 *LH* 基因和 *LHR* 基因具有相似的分化程度。4 种基因聚类图的第 3 组均包括仓鼠、小家鼠和褐家鼠。*ESR* 基因和 *PRLR* 基因在第 3 组中的聚类层次相同,都是褐家鼠和小家鼠先聚合,再和仓鼠聚合(图 1-A、D)。说明在这 3 个物种中 *ESR* 基因和 *PRLR* 基因的进化速率一致。*LH* 基因是褐家鼠先和仓鼠聚合,再和小家鼠聚合(图 1-B),*LHR* 基因是仓鼠与小家鼠聚合再和褐家鼠聚合(图 1-C)。表明小家鼠、褐家鼠和仓鼠的 *LH*、*LHR* 基因进化速率上存在差异。4 种基因构建的系统发育树拓扑结构存在差异,同样说明这 4 种基因序列的进化速率有所差异^[12]。系统发育树树枝的长度取决于基因进化的速率和进化的时间。由于 4 个基因都选取同样的物种,因此,4 个基因进化的时间相同。*ESR* 构建的进化树树枝最短,*LH* 树枝最长,说明 *ESR* 进化速率相对较慢,*LH* 相对较快。



A、B、C、D 分别为 *ESR*、*LH*、*LHR*、*PRLR* 基因根据物种间的核苷酸歧异度构建的聚类关系

图 1 11 个物种依据 4 个基因遗传距离构建的聚类关系

3 结论

研究结果表明,4 种基因在所研究的 11 个物种中的核苷酸多样性不同,*ESR* 基因的单位长度核苷酸多态位点数、单位长度总突变位点数和核苷酸多态性均明显低于其他 3 种基因,而 *LH* 基因的单位长度核苷酸多态位点数、单位长度总突变位点和核苷酸多态性均相对高。*ESR* 基因的非同义替换与同义替换的比值远远低于 1,*PRLR* 基因比值略小于 1,*LH* 和 *LHR* 基因的比值均大于 1。*ESR* 基因的平均遗传距离最小,*LH* 基因的平均遗传距离最大。4 种基因构建进化树拓扑结构存在差异。4 种繁殖基因在进化过程中,*ESR* 基因的各物种的遗传分化程度较小,进化较为保守。*LH* 基因在各物种的遗传分化程度较大,进化速率相对较快。

参考文献:

- [1] Rothschild M F. Identification of quantitative trait loci and interesting candidate genes in the pig: progress and prospect[J]. *Wegalp*, 1998, 26(7): 403-409.
- [2] Rothschild M F. Analysis of RFLPs (restriction fragment length polymorphisms) on estrogen receptor in pigs[C]// *Proc 5th world Congr Genet Applied Livestock Prod*. 1994, 21: 225-228.
- [3] Hewitt S C, Korach K S. Oestrogen receptor knockout mice: roles for oestrogen receptors A and B in reproductive tissues [J]. *Reproduction*, 2003, 125 (2): 143-149.
- [4] 孙明亮, 马金成. *FSHR* 与 *LHR* 的研究进展[J]. *黑龙江动物繁殖*, 2008, 16(4): 1-3.
- [5] Dufau M L, Tsai-Morris C H, Hu Z Z. Structure and regulation of the luteinizing hormone receptor gene[J]. *The Journal of Steroid Biochemistry and Molecular Biology*, 1995, 53(1/6): 283-291.
- [6] 梁鸿雁, 周虚, 李存锦, 等. 成年公猪睾丸及附睾中促黄体素受体 mRNA 表达的研究[J]. *黑龙江畜牧兽医*, 2009(7): 10-12.
- [7] 王怀禹. *PRLR* 基因多态性与动物繁殖性能的关系[J]. *中国农学通报*, 2009, 25(12): 1-5.
- [8] 赵亮. 太湖新银鱼线粒体 D-loop 和 *Cyt b* 片段序列结构与进化速率比较[J]. *动物学杂志*, 2010, 45(2): 27-38.
- [9] 唐萍. 叶绿体基因组进化的速率和方式[J]. *生物学通报*, 2010, 45(6): 8-10.
- [10] 申欣. 文昌鱼线粒体基因组特征分析及分子标记探讨[J]. *海洋科学*, 2010, 35(7): 7-13.
- [11] 李易. 基因进化的同义与非同义替代计算及统计检验的比较分析[J]. *曲靖师范学院学报*, 2006, 25(6): 1-8.
- [12] 王利利. *Cyt b* 和 *rRNA* 基因进化速度对构建系统进化树的影响[J]. *安徽农业科学*, 2011, 39(9): 5101-5103.