研究论文

基于编码序列、基因间序列和氨基酸序列构建的系统发生关系比较

杨林源[1], 李宏[1], 赵小庆[1], 刘国庆[2]

1. 内蒙古大学物理科学与技术学院, 呼和浩特 010021;
2. 内蒙古科技大学数理与生物工程学院, 包头 014010

摘要：

以7种古菌、46种细菌和10种真核生物的基因组为样本,考虑碱基间的短程关联和长程关联作用,得到编码序列的密码对和基因间序列的三联体对中不同位点的二核苷酸频率,据此构建了基于编码序列和基因间序列的系统发生关系。无论是基于编码序列还是基因间序列对信息进行聚类,古菌或真核均被聚在一支上,表明聚类参数的选择是合适的;与基于氨基酸序列构建的系统发生关系进行两两比较,发现大部分硬壁菌的编码序列与基因间序列之间,以及编码序列与氨基酸序列之间的进化都存在较大差异。通过分析认为,只有综合考虑这三类序列的进化信息,才可能得到更自然的系统发生关系。

关键词： 系统发生关系　 二核苷酸频率　 碱基长程关联　 编码序列　 基因间序列

Comparison of Phylogenetic Relationships Constructed by Using Protein-Coding Sequences, Intergenic Sequences and Amino Acid Sequences

YANG Linyuan[1], LI Hong[1], ZHAO Xiaoqing[1], LIU Guoqing[2]

1. School of Physical Science and Technology, Inner Mongolia University, Hohhot 010021, China;
2. School of Mathematics, Physics and Biological Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China

Abstract:

Considering the effect of short and long range correlation of base pairs, dinucleotide frequencies in different sites either of codon pairs of protein coding sequences or triplet pairs of intergenic sequences were calculated for 7 archaea, 46 bacteria and 10 eukaryote genomes. Phylogenetic relationships were constructed respectively for protein coding sequences and intergenic sequences. Archaea or eukaryotes were clustered one independent branch by the two kinds of sequences. It shows that the selected parameters are suitable. Compared with the phylogenetic relationship which was constructed by amino acid sequences, it was found that the evolutions of protein coding, intergenic and amino acid sequences have distinct differences for most firmicutes. We propose that the evolution information of three kinds of sequences should be combined together so that a more natural phylogenetic relationship can be obtained.

Keywords: Phylogenetic relationship　 Dinucleotide frequency　 Long rang base correlation　 Coding sequence　 Intergenic sequence

通讯作者： 李宏,电话：(0471)4996004,E-mail：ndlihong@imu.edu.cn

作者简介：

作者Email: ndlihong@imu.edu.cn

参考文献：

1. Zuckerkandl E, Pauling L. Molecules as documents of evolutionary history. J Theoret Biol, 1965, 8(2): 357~366
2. Woese CR. Bacterial evolution. Microbiol Rev, 1987, 51(2):221~271
3. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. Proc

扩展功能

本文信息
- Supporting info
- PDF(979KB)
- [HTML全文]
- 参考文献[PDF]
- 参考文献

服务与反馈
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- 引用本文
- Email Alert
- 文章反馈
- 浏览反馈信息

本文关键词相关文章
- 系统发生关系
- 二核苷酸频率
- 碱基长程关联
- 编码序列
- 基因间序列

本文作者相关文章
- PubMed

Natl Acad Sci USA, 1977, 74(11): 5088~5090

4. Sun JD, Xu Z, Hao BL. Whole-genome based archaeal phylogeny and taxonomy: A composition vector approach. Chinese Sci Bull, 2010, 55(22): 2323~2328

5. Brown JR, Doolittle WF. Archaea and the prokaryote- to-eukaryote transition. Microbiol Mol Biol R, 1997, 61(4):456~502

6. Woese CR. Interpreting the universal phylogenetic tree. Proc Natl Acad Sci USA, 2000, 97(15): 8392~8396

7. Dong XZ. Archaea: Transition from prokaryotes to eukaryotes? Microbiol China, 1999, 26(6): 426~430

8. Burge C, Campbell AM, Karlin S. Over and under- representation of short oligonucleotides in DNA sequences. Proc Natl Acad Sci USA, 1992, 89(4): 1358~1362

9. Kariin S, Burge C. Dinucleotide relative abundance extremes: A genomic signature. Trends Genet, 1995, 11(7):283~290

10. Qi J, Wang B, Hao BL. Whole proteome prokaryote phylogeny without sequence alignment: A K-string composition approach. J Mol Evol, 2004, 58(1): 1~11

11. Hao BL, Qi J. Prokaryote phylogeny without sequence alignment: From avoidance signature to composition distance. J Bioinform Comput Biol, 2004, 2(1): 1~19

12. Qi J, Luo H, Hao BL. CVTree: A phylogenetic tree reconstruction tool based on whole genomes. Nucleic Acids Res, 2004, 32(Web Server Issue): W45~W47

13. Karlin S, Ladunga I, Blaisdell BE. Heterogeneity of genomes: Measures and values. Proc Natl Acad Sci USA,1994, 91(26): 12837~12841

14. Wang FP, Li H. Codon-pair usage and genome evolution. Gene, 2009, 433(1-2): 8~15

15. Liu GQ, Li H. The correlation between recombination rate and dinucleotide bias in Drosophila melanogaster. J Mol Evol, 2008, 67(4): 358~367

16. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. Proc Natl Acad Sci USA, 1990, 87(12): 4576~4579

17. Li W, Fang W, Ling L, Wang J, Xuan Z, Chen R. Phylogeny based on whole genome as inferred from complete information set analysis. J Biol Phys, 2002, 28(3): 439~447

18. Kurtz S, Narechania A, Stein JC, Ware D. A new method to compute k-mer frequencies and its application to annotate large repetitive plant genomes. BMC Genomics, 2008, 9: 517

19. Romá F, Ramirez-Pastor AJ, Riccardo JL. Configurational entropy in k-mer adsorption. Langmuir, 2000, 16(24): 9406~9409

20. Wolf YI, Rogozin IB, Grishin NV, Tatusov RL, Koonin EV. Genome tree constructed using five different approaches suggest new major bacterial clades. BMC Evol Biol, 2001, 1: 8

21. Karlin S, Mrázek J. Compositional differences within and between eukaryotic genomes. Proc Natl Acad Sci USA, 1997, 94(19): 10227~10232

本刊中的类似文章

文章评论

反馈人		邮箱地址	
反馈标题		验证码	9461