

哺乳动物 *cd59* 基因的进化

龚远英^{①†} 彭旻晟^{①②†} 周伟平^{①③④} 张亚平^{①④*}

(① 中国科学院昆明动物研究所遗传资源与进化国家重点实验室, 昆明 650223; ② 中国科学院研究生院, 北京 100049;
③ 中国科学技术大学分子生物学与细胞生物学系, 合肥 230026; ④ 云南大学省部共建生物资源保护与利用
重点实验室, 昆明 650091)

摘要 采用 PCR 以及 BLAST 方法从 5 种哺乳动物中获得了 *cd59* 基因的编码区序列, 结合 GenBank 中已有的序列, 计算 *cd59* 基因在哺乳动物中的核苷酸替换速率。对非同义替换速率和同义替换速率进行比较的结果显示, *cd59* 在哺乳动物中总体上受到负选择作用; 用 PAML 软件“位点-特异”模型检测到 4 个受到正选择作用的位点, 4 个位点分布于分子表面, 其中 2 个位于功能重要的区域; 此外, 用“支-位点-特异”模型在小鼠通过基因复制后形成的 *cd59a* 和 *cd59b* 上检测正选择引起的加速进化, 并检测到该支系特异的正选择位点 1 个。

关键词 *cd59* 进化 正选择

CD59 是一种在哺乳动物机体中广泛分布的糖基磷脂酰肌醇锚定蛋白(glycosylphosphatidylinositol-anchored proteins, GPI-anchored proteins), 在补体激活的终末阶段, 作为膜攻击复合体(membrane attack complex, MAC)C5b-9 的抑制物, 起到同源限制作用, 在补体免疫反应的调节上行使着重要的功能 [1]。人类的 *CD59* 基因定位于 11p13^[2], 由 4 个外显子和 3 个内含子组成, 其 cDNA 长 387 bp^[3], 包括由 25 个氨基酸残基构成的信号肽, 77 个氨基酸残基构成的成熟蛋白区段以及在添加 GPI 后切除的 26 个氨基酸残基构成的疏水尾(hydrophobic tail)^[4]。实验研究发现, 一些病原微生物以及肿瘤细胞通过不同途径, 利用 CD59 降低补体反应的特性, 从而逃避宿主免疫系统的监视和攻击 [5~8]。一般认为, 参与宿主与病原微生物相互作用的分子由于受到宿主与病原体之间“军备竞赛(arm-race)”的影响, 在进化中很可能因受到达尔文正选择作用而发生加速进化 [9]。另外, 体内 CD59 还与

免疫系统的其他组分存在相互作用 [10,11], 因此具有广泛的免疫调节功能。基因组水平的检测也表明免疫系统经历着显著的正选择作用而快速进化 [12~15]。因此, CD59 在哺乳动物中可能具有复杂的进化过程。

为了弄清 *CD59* 基因在哺乳动物中进化的详细情况, 通过 PCR 扩增以及 BLAST 方法获取了一些物种的 *cd59* 编码序列和一条相关的病毒蛋白编码序列。结合已有的一系列实验结果, 对这些序列进行了系统的分析。

1 材料和方法

1.1 材料

通过对数据库的搜索和文献的追踪, 获得以下物种的序列: 人(Human, *Homo sapiens*), 猩猩(Orangutan, *Pongo pygmaeus*), 食蟹猴(Cynomolgus monkey, *Macaca fascicularis*), 狒狒(Baboon, *Papio sp.*), 非洲绿猴(African green monkey, *Cercopithecus*

收稿日期: 2007-05-08; 接受日期: 2007-07-09

国家自然科学基金(批准号: 30621092, 30430110)、国家重点基础研究发展计划(批准号: 2007CB411600)和云南省科技厅资助项目。

†同等贡献

*联系人, E-mail: zhangyp1@263.net.cn

aethiops), 松鼠猴(Squirrel monkey, *Saimiri sciureus*), 夜猴(Owl monkey, *Aotus trivirgatus*), 绒猴(Marmoset, *Callithrix sp.*), 小鼠(Mouse, *Mus musculus*), 大鼠(Rat, *Rattus norvegicus*), 兔(Rabbit, *Oryctolagus cuniculus*), 猪(Pig, *Sus scrofa*), 牛(Cow, *Bos taurus*)和犬(Dog, *Canis familiaris*); 另外获得来自 2 型松鼠猴疱疹病毒(*Saimiri herpesvirus 2*, SaHV2, *Herpesvirus saimiri*)的一条病毒蛋白编码序列。要注意的是, NCBI 的 GenBank 将非洲绿猴与狒狒的序列对换^[16], 本文对此进行了更正。此外, 还采用 PCR 扩增并测序的实验手段获得了黑猩猩(Chimpanzee, *Pan troglodytes*), 白眉长臂猿(Gibbon, *Hylobates hoolock*), 恒河猴(Rhesus, *Macaca mulatta*)以及白臀叶猴(Douc langur, *Pygathrix nemaeus*)4 个灵长类动物的序列。根据人和黑猩猩的 CD59 序列设计了 3 对引物(表 1 和图 1)来对灵长类的第 2~4 个外显子进行 PCR 扩增。PCR 试剂购自 TaKaRa 公司。PCR 产物凝胶回收试剂盒购自上海华舜生物工程有限公司。测序 Kit 为 PE 公司生产的 BigDye™ Terminator, 引物由上海生工生物工程技术有限公司合成。PCR 产物经过纯化后用 ABI3700 自动测序仪进行正负链双向测序。另一方面, 通过 BLAST 的方法从 Ensembl(<http://www.ensembl.org>)公布的基因组数据库中获取了黑猩猩、恒河猴以及负鼠(Opossum, *Monodelphis domestica*)的相关序列, 其中黑猩猩和恒河猴的序列与实验获得的一致。

1.2 进化分析

用 Clustal X 程序^[17]对 *cd59* 基因编码序列进行比对, 再进行手工调整(GenBank 的登录号: 人(Human), NM_203330; 猩猩(Orangutan), CR861075; 食蟹猴(Cynomolgus monkey), AB072017(等位基因A), AB072018(等位基因B); 狒狒(Baboon), L22863; 非洲绿猴(African green monkey), L22862; 松鼠猴(Squirrel monkey), L22859; 夜猴(Owl monkey), L22861; 绒猴(Marmoset), L22860; 小鼠(Mouse), NM_007652(*cd59a*), NM_181858(*cd59b*); 大鼠(Rat), NM_012925; 兔(Rabbit), AF040387; 猪(Pig), NM_214170; 牛(Cow), NM_001037446; 犬(Dog), XM_533156; SaHV2, NC_001350)。将比对结果去除起始和终止密码子后进行下一步分析。使用整合在 MEGA 3.1^[18]中的不均衡指数(Disparity Index)分析^[19]对比对结果进行核苷酸替换同质性检测(重复 8000 次

的 Monte-Carlo 检测)。最适核苷酸替换模型通过 Modeltest 3.7^[20]对多个模型进行比较获得。根据选定的最适替换模型, 使用 DAMBE V.4.5.2^[21]对核苷酸替换(转换/颠换)的饱和性(substitution saturation)进行检测; 使用 MEGA 3.1 构建 NJ 树, 并进行 1000 次 Bootstrap 检验。用 PAML 3.15^[22]来分析 *cd59* 基因在进化过程中选择压力的变化。这一分析使用 3 种模型: (i) “支-特异(branch-specific)”模型^[23,24]。它允许非同义替换和同义替换的比值 ω 在不同支系上有变化, 其中单一比率(one ratio)模型假定所有进化支的 ω 值都是一致的, 而自由比率(free ratio)模型假定各支的 ω 值各不相同。比较这 2 个模型的 2 倍似然率(likelihood ratio test, LRT)对数差的 χ^2 分布可以检验 ω 值在不同支上是否有变化。自由度等于这 2 个模型所包含参数数目的差。(ii) “位点-特异(site-specific)”模型^[25,26]。它假定各进化支有固定的 ω 值, 但不同位点之间可以有变化。采用的 2 对成对比较模型分别是 M1a 和 M2a 以及 M8a 和 M8^[27]。(iii) “支-位点-特异(branch-site-specific)”模型, 它将进化树上的支系分成“前景支(foreground)”和“后景支(background)”两类, 允许 ω 在两类支系上有变化且“前景支” ω 大于“后景支” ω , 再假定在每一类支系上有固定的 ω 值, 但不同位点之间可以有变化^[28]。

1.3 蛋白质结构分析

采用基于 NMR 数据且经过细化修正的人类 CD59 胞外结构区域的结构模型^[29]。采用 BALLView 1.1.1^[30,31]进行显示分析。

2 结果和分析

2.1 序列的获得

黑猩猩、白眉长臂猿、白臀叶猴和恒河猴的 *cd59* 基因的第 2~4 个外显子序列通过普通 PCR 方法获得, 使用的引物对见表 1 及图 1。序列比对结果显示(图 2), 对于 CD59 结构维持必需的 5 对半胱氨酸二硫键以及活性中心 69W 在所有序列中高度保守; 信号肽比成熟蛋白区段保守; 在疏水尾区段, 特别是前半部分, 出现较多的插入-缺失(Indels), 最为显著的是在黑猩猩中出现一段 48 bp 的插入序列。使用软件的默认设置通常均把插入-缺失作为模糊位点予以删除, 尽管已有研究表明插入-缺失也可能受到达尔文正选择作用^[32], 但在本工作中对插入-缺失不作讨论。

表 1 本研究所用引物

名称	正链引物(5'→3')	负链引物(5'→3')
外显子 2	CAGCCCTTGGTTTGAGAC	CAGCCGCCGAGTGTGAC
外显子 3	GGGCTGCCAGGGGACAA	GACAGGGAAACTCTTACAG
外显子 4	CCGTCCCACCCCCATAACTA	CAGCAAGAGAAAGCGGACTACG

图 1 人类 *cd59* 基因的结构

方框代表外显子; 框内数字表示其大小; 阴影标识的表示该部分被翻译, 构成可读框结构. 3 对箭头分别代表 PCR 扩增引物.

2.2 核苷酸替换模型的选择

对 22 条核苷酸序列进行不均衡指数分析, 结果显示 SaHV2 的序列与其他序列在替换方式上存在显著差异, 因此在构建进化树及其分析时应该予以考虑. 通过比较在剔除和保留该序列的情况下所得到的结果, 如果出现不同, 则在分析中采用剔除该序列后所获得的结果. 在使用 Modeltest 3.7 选择最适替换模型时, 阶层式似然率检验(hierarchical likelihood ratio tests, hLRTs) 和 Akaike 信息量准则(Akaike information criterion, AIC) 所得到的结果不一致, 由于 *cd59* 基因序列较短, 为避免更多的参数估计带来的误差, 选用参数较少的模型, 即采用 hLRTs 的结果, 选用 Kimura 二参数(K80)模型为最适模型. 使用 DAMBE V.4.5.2 对核苷酸替换(转换/颠换)的饱和性进行检测, 替换模型采用 K80, 结果如图 3 显示, 表明核苷酸替换并未饱和, 序列可按照常规分析方法进行.

2.3 进化分析

构建 NJ 树的结果表明, 其拓扑结构支持星状系统发育树, 即假设哺乳动物各大类群都是在距今大约 75 Ma 年前同时分化出来的(图 4)^[33]. 该结果中的灵长目支系的拓扑结构与 Goodman 等人提出的拓扑关系相一致^[34]. 根据这种系统发育关系, 把全编码区序列系统发育树的拓扑结构在 TreeView 1.6.6 中进行编辑, 从而获得以负鼠为外群的星状系统发育树. 使用整合在 MEGA 3.1 中的 Pamilo-Bianchi-Li 法^[35,36] 计算所获得哺乳动物的 *cd59* 基因成熟蛋白编码区的每非同义位点的非同义替代数 $Ka = 0.349$ (标准误 $SE = 0.046$), 每同义位点的同义替代数 $Ks = 0.547$ (标准误 $SE = 0.060$), Ka/Ks 显著小于 1(Z 检验, $P < 0.05$), 说明

哺乳动物 *CD59* 基因在进化中总体上体现受到负选择. 但是传统“特定(ad hoc)”方法对达尔文正选择的检测过于严格, 在长期较强负选择和中性漂变的掩盖作用下, 正选择作用往往难以检出; 而使用基于最大似然法估计的正选择检测方法可以提高检测的灵敏性, 可检出处于上述一定掩盖作用下的正选择位点^[37].

由于正选择可能仅作用于进化的某一阶段, 即 *CD59* 基因在哺乳动物某些支系的进化中可能呈现“插曲式”的进化模式. 采用“支-特异”模型进行检测, 结果未能有效检出受到正选择的支系, 说明 ω 值在各支系上的分布并非差异显著. 另外, 正选择可能仅仅作用于部分位点, 用“位点-特异”模型进行检测, 其中 M8a 和 M8 以及 M1a 和 M2a 2 组模型分别检出 16 个和 4 个潜在的正选择位点, 并且后一组结果包含于前一组的结果之中. 为了减少假阳性问题, 采用较为保守的 M1a 和 M2a 一组的结果. 如图 2 所示, 4 个位点(36A, 78D, 82R 和 104Q) 均位于成熟蛋白区段. 在结构模型上对其进行分析^[29], 结果显示(图 5) 4 个位点均分布于分子表面, 可能参与分子间的相互作用. 值得注意的是氨基酸 Asp(D), Arg(R) 和 Gln(Q) 是常见的电子受体和供体, 这些位点的改变常会影响分子间的相互作用, 因此位点 78D, 82R 和 104Q 可能对 *CD59* 在不同物种中功能产生影响. 其中, 78D 和 82R 位于 α 融合蛋白(76F~85E) 内(图 2), 这段 α 融合蛋白被认为结构上接近于活性中心 69W, 在与补体分子 C8 和 C9 结合时起到辅助作用^[29], 其中位点 82R 的定点突变可以显著影响 *CD59* 的结合能力^[38]. 而通过对哺乳动物 *cd59* 序列的比较, 发现 76F~85E α 融合蛋白在物种间存在较大的差异, 联系其功能以及潜在正选择位点检测结果, 可以推断 76F~85E α 融合蛋白可能在一定程度上决定 *CD59* 分子对异源补体分子结合活性大小, 而

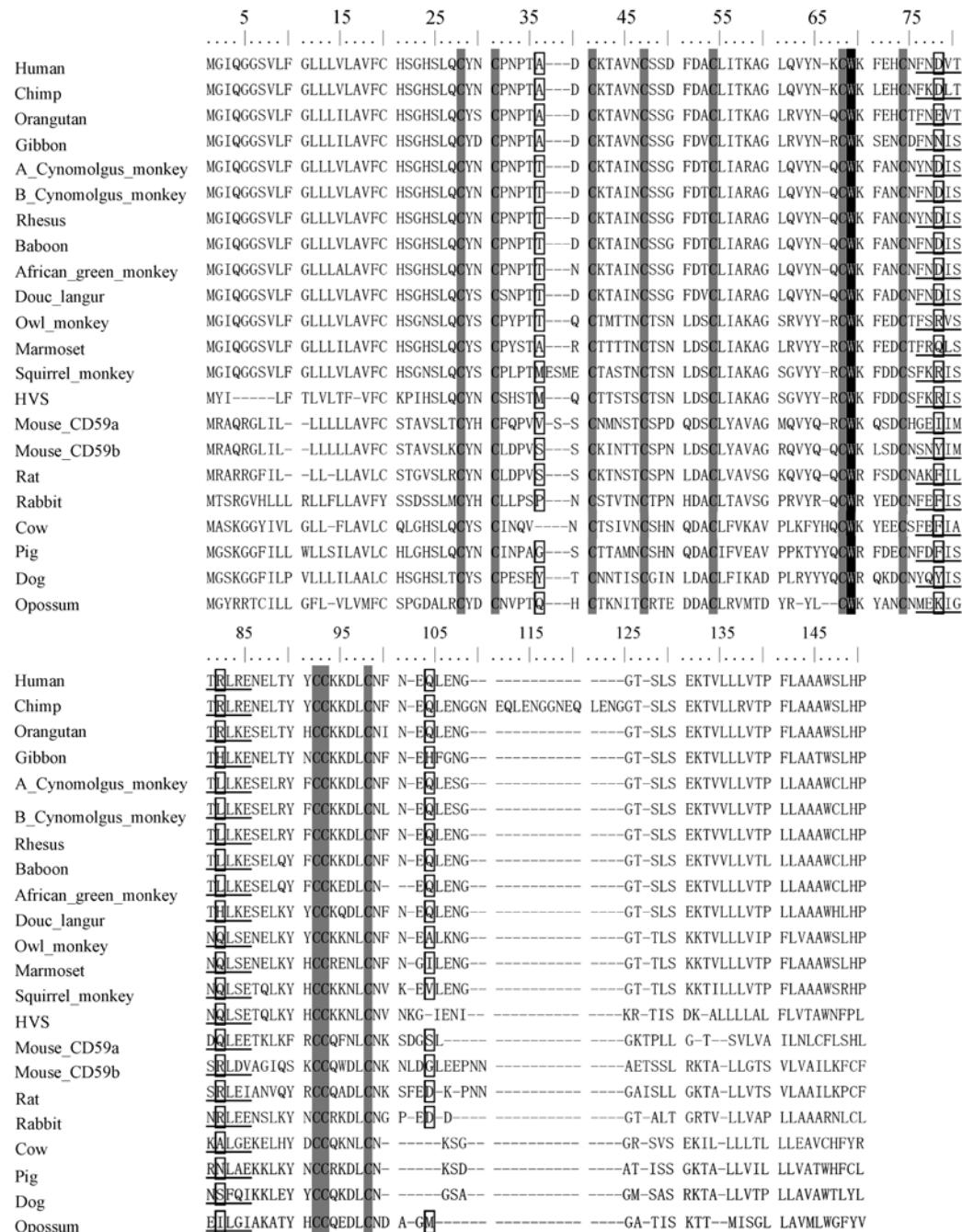


图 2 CD59 的氨基酸序列比对

其中 1~25 位点为信号肽, 26~107 位点为成熟蛋白区段, 108~150 位点则构成疏水尾区段。灰底背景表示 5 对半胱氨酸残基; 黑底背景表示 CD59 的活性中心 69W; 下画线表示 α 螺旋(76F~85E); 矩形框表示“位点-特异”模型检出的潜在正选择位点(95%以上显著性水平); “—”表示一个氨基酸缺失。

α 旋在序列上的差异很可能是受到达尔文正选择加速进化而非选择压力放松的结果。在涉及 CD59 关于肿瘤治疗以及异源移植及其过敏性反应研究时, 可以考虑对 76F~85E α 螺旋进行改造, 或设计一定的相互作用物以调节 CD59 分子与补体分子结合反应的活性。

另外检出的 2 个潜在正选择位点 36A 和 104Q, 特别是 104Q, 其分布不在与 C8 和 C9 的结合界面范围内, 猜测其可能参与和其他免疫组分的相互作用。

cd59 在小鼠中发生复制事件, 为了探讨该事件对 *cd59* 基因进化的影响, 使用“支-位点-特异”模型对

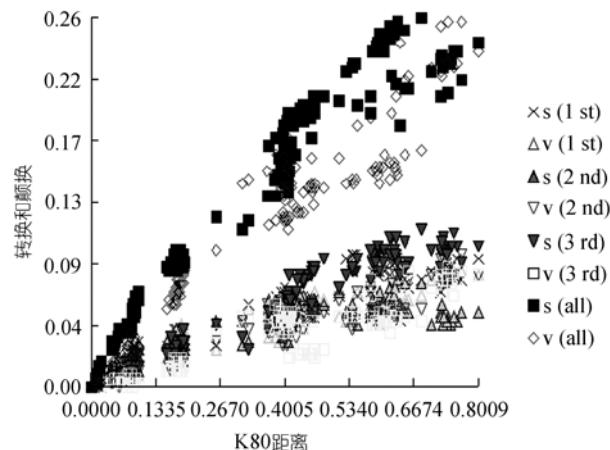


图 3 转换-颠换饱和性分析结果(K80 距离)

s: 转换; v: 颠换. 1st, 2nd, 3rd 分别表示密码子的第 1, 2, 3 位. all 表示整个密码子

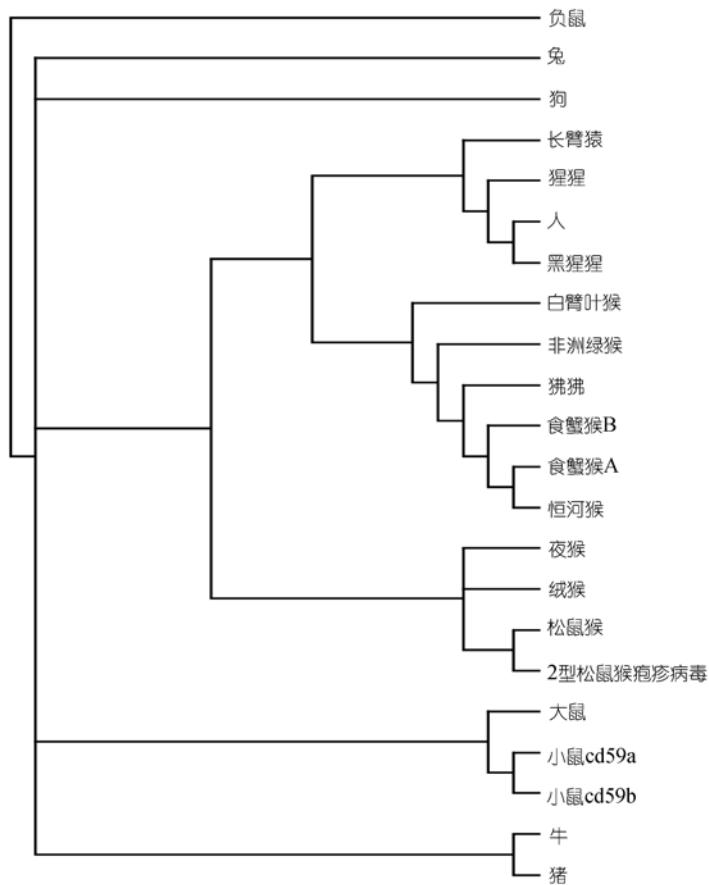


图 4 哺乳动物星状系统发育树

大鼠、小鼠的各条序列以及小鼠 $cd59$ 基因发生复制前的共同祖先序列(“前景支”)进行检测。结果仅在小鼠 $cd59a$ 和 $cd59b$ 支系(作为一个“前景支”)上检测到潜在的正选择位点 90Y(95%以上显著性水平), 说明相

比于整个“后景支”, 小鼠的 2 个 $cd59$ 拷贝具有更高的 ω 和更高的进化速率。这两个拷贝在小鼠大多数组织中都具有表达活性, 但产物活性相差 6 倍; 且两者在雄性生殖系统中存在时空表达差异, 说明两个拷贝

可能行使着不同的功能^[39]。因此, 小鼠中的cd59基因在复制后, cd59a和cd59b实现功能分化并在达尔文正选择的作用下加速进化, 而特异性的潜在正选择位点90Y对两者产物功能差异的产生可能具有重要的影响。

参 考 文 献

- 1 Davies A, Lachmann P J. Membrane defense against complement lysis: the structure and biological properties of CD59. *Immunol Res*, 1993, 12: 258—275
- 2 Philbrick W M, Palfree R G, Maher S E, et al. The CD59 antigen is a structural homologue of murine Ly-6 antigens but lacks interferon inducibility. *Eur J Immunol*, 1990, 20: 87—92[DOI]
- 3 Petranka J G, Fleenor, D E, Sykes K, et al. Structure of the CD59-encoding gene: further evidence of a relationship to murine lymphocyte antigen Ly-6 protein. *Proc Nat Acad Sci USA*, 1992, 89: 7876—7879[DOI]
- 4 Fletcher C M, Harrison R A, Lachmann P J, et al. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure*, 1994, 2: 185—199[DOI]
- 5 Albrecht J C, J Nicholas, K R Cameron, et al. Herpesvirus Saimiri has a gene specifying a homologue of the cellular membrane glycoprotein CD59. *Virology*, 1992, 190: 527—530[DOI]
- 6 Rother R P, S A Rollins, W L Fodor, et al. Inhibition of complement-mediated cytolysis by the terminal complement inhibitor of herpesvirus saimiri. *J Virol*, 1994, 68: 730—737
- 7 Giddings K S, Zhao J, Sims P J, et al. Human CD59 is a receptor for the cholesterol-dependent cytolsin intermedilysin. *Nat Struct Mol Biol*, 2004, 11: 1173—1178[DOI]
- 8 Donev R M, Cole D S, Sivasankar B, et al. p53 regulates cellular resistance to complement lysis through enhanced expression of CD59. *Cancer Res*, 2006, 66: 2451—2458[DOI]
- 9 Sawyer S L, Wu L I, Emerman M, et al. Positive selection of primate TRIM5α identifies a critical species-specific retroviral restriction domain. *Proc Nat Acad Sci USA*, 2005, 102: 2832—2837[DOI]
- 10 Blok V T, Gelderman K A, Tijsma O H, et al. Cytokines affect resistance of human renal tumor cells to complement-mediated injury. *Scand J Immunol*, 2003, 57: 591—599[DOI]
- 11 Naderi S, Hormann P, Seiter S, et al. CD2-mediated CD59 stimulation in keratinocytes results in secretion of IL-1 alpha, IL-6, and GM-CSF: Implications for the interaction of keratinocytes with intraepidermal T lymphocytes. *Int J Mol Med*, 1999, 3: 609—614
- 12 Bustamante C D, Fledel-Alon A, Williamson S, et al. Natural selection on protein-coding genes in the human genome. *Nature*, 2005, 437: 1153—1157[DOI]
- 13 Nielsen R, Bustamante C, Clark A G, et al. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol*, 2005, 3: e170[DOI]
- 14 Voight B F, Kudaravalli S, Wen X, et al. A map of recent positive selection in the human genome. *PLoS Biol*, 2006, 4: e72[DOI]
- 15 Wang E T, Kodama G, Baldi P, et al. Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc Natl Acad Sci USA*, 2006, 103: 135—140[DOI]
- 16 Fodor W L, Rollins S A, Bianco-Caron S, et al. Primate terminal complement inhibitor homologues of human CD59. *Immunogenetics*, 1995, 41: 51[DOI]
- 17 Thompson J D, Gibson T J, Plewniak F, et al. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, 1997, 25: 4876—4882[DOI]
- 18 Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform*, 2004, 5: 150—163[DOI]
- 19 Kumar S, S R Gadagkar. Disparity Index: A simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. *Genetics*, 2001, 158: 1321—1327
- 20 Posada D, Crandall K A. Modeltest: testing the model of DNA substitution. *Bioinformatics*, 1998, 14: 817—818[DOI]
- 21 Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. *J Hered*, 2001, 92: 371—373[DOI]
- 22 Yang Z. PAML: A program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*, 1997, 13: 555—556
- 23 Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol*, 1998, 15: 568—573
- 24 Yang Z, Nielsen R. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J Mol Evol*, 1998, 46: 409—418[DOI]
- 25 Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, 1998, 148: 929—936
- 26 Yang Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J Mol Evol*, 2000, 51: 423—432
- 27 Yang Z, W S W Wong, R Nielsen. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol*, 2005, 22: 1107—1118[DOI]
- 28 Yang Z, R Nielsen. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*, 2002, 19: 908—917
- 29 Huang Y X, Smith C A, Song H B, et al. Insight into the human CD59 complement binding interface toward engineering new therapeutics. *J Biol Chem*, 2005, 280: 34073—34079[DOI]

- 30 Moll A, Hildebrandt A, Lenhof H P, et al. BALLView: An object-oriented molecular visualization and modeling framework. *J Comput Aided Mol Des*, 2005, 19: 791—800[\[DOI\]](#)
- 31 Moll A, Hildebrandt A, Lenhof H P, et al. BALLView: A tool for research and education in molecular modeling. *Bioinformatics*, 2006, 22: 365—366[\[DOI\]](#)
- 32 Podlaha O, Zhang J Z. Positive selection on protein-length in the evolution of a primate sperm ion channel. *Proc Nat Acad Sci USA*, 2003, 100: 12241—12246[\[DOI\]](#)
- 33 Graur D. Molecular phylogeny and the higher classification of eutherian mammals. *Trends Ecol Evol*, 1993, 8: 141—147[\[DOI\]](#)
- 34 Goodman M, Porter C A, Czelusniak J, et al. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phyl Evol*, 1998, 9: 585—598[\[DOI\]](#)
- 35 Li W H. Unbiased estimation of rates of synonymous and nonsynonymous substitution. *J Mol Evol*, 1993, 9: 96—99[\[DOI\]](#)
- 36 Pamilo P, Bianchi N O. Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. *Mol Biol Evol*, 1993, 10: 271—281
- 37 Yang Z. Inference of selection from multiple species alignments. *Curr Opin Genet Dev*, 2002, 12: 688—694[\[DOI\]](#)
- 38 Qin X, Miwa T, Aktas H, et al. Genomic structure, functional comparison, and tissue distribution of mouse Cd59a and Cd59b. *Mamm Genome*, 2001, 12: 582—589[\[DOI\]](#)
- 39 Bodian D L, Davis S J, Morgan B P, et al. Mutational analysis of active site and antibody epitopes of the complement-inhibitory Glycoprotein, CD59. *J Exp Med*, 1997, 185: 507—516[\[DOI\]](#)