

哺乳动物朊蛋白基因序列变异及其系统发育意义

管廷龙^{1,2}, 范振鑫², 曾博², 岳碧松², 邹方东^{2*}

(1. 淄博市疾病预防控制中心, 山东淄博 255026; 2. 生物资源与生态环境教育部重点实验室, 四川大学生命科学学院, 成都 610064)

摘要:以朊蛋白基因(PRNP)为研究对象,从 GenBank 中选取代表哺乳动物的 17 目 61 属共 84 个物种 PRNP 的编码区序列,分析了该基因的结构及其在不同层次分类阶元间的进化关系。发现由于在基因内序列重复区存在插入或缺失,PRNP 的编码区长度在哺乳动物不同目的物种间存在差异。在终止密码子的使用上不同目的物种之间存在明显的偏好性, TGA 和 TAG 是哺乳动物 PRNP 最常用的终止密码子,只有奇蹄目的 5 个种和有袋目袋鼠科的 2 个种是使用 TAA 终止。所有 84 条序列呈现了较高的核苷酸多样性(10.54%)。以 PRNP 为分子标记构建的系统进化树总体上是和以往研究一致的,结果表明 PRNP 可以应用于哺乳动物较高分类单元的系统进化研究,但对于能否用于种间和种内不能肯定。最后应用系统进化树对尚未有患病报道的哺乳动物的患病可能性进行了预测。

关键词: 哺乳动物; 朊蛋白基因; 终止密码子; 系统发育

中图分类号: Q959.8; Q75 **文献标识码:** A **文章编号:** 1000-7083(2010)04-0555-05

Prion Protein Gene (PRNP) Variation and its Phylogenetic Implications for Mammals

GUAN Ting-long^{1,2}, FAN Zhen-xin², ZENG Bo², YUE Bi-song², ZOU Fang-dong^{2*}

(1. Zibo Municipal Center for Disease Control and Prevention, Zibo, Shandong Province 255026, China; 2. Key Laboratory of Bio-resources and Eco-environment (Ministry of Education), College of Life Sciences, Sichuan University, Chengdu 610064, China)

Abstract: Eighty-four prion protein gene (PRNP) sequences from 84 mammalian species belonging to 61 genera and 17 orders were analyzed in order to investigate its structure and differentiation among species. The lengths of PRNP coding sequences varied a lot among different species. In addition, TGA and TAG were the most widely used stop codons in mammalian PRNP, while TAA is only found in species of Perissodactyla and Marsupialia. Nucleotide diversity was high (10.54%) among all 84 sequences. The reconstructed PRNP phylogenetic tree for all 84 species matches the results of traditional studies. Our analyses suggest that PRNP could be used as a molecular marker in phylogenetic studies of high level taxonomic units. Finally, we speculate on the possibility of different orders of mammals suffering from transmissible spongiform encephalopathies according to their phylogenetic relationships.

Key words: mammal; prion protein gene; stop codon; phylogeny

传染性海绵状脑病(Transmissible spongiform encephalopathies, TSEs)是由朊蛋白(prion protein, PrP)引起的人和多种哺乳动物以中枢神经退行性变化为主要特征的一种慢性消耗性传染病。例如,人的克雅氏病和库鲁病、绵羊瘙痒症、疯牛病、传染性水貂脑病等约 30 多种(Prusiner, 1997、2001),它们是由朊病毒介导所引起的。朊蛋白是由宿主的朊蛋白基因(prion protein gene, PRNP)所编码,具有致病 PrP^{Sc}和正常 PrP^C两种形态,然而直到现在人们对它们正常生理功能和致病机理依然了解很少。人类的 PRNP 编码为一个含有 253 个氨基酸残基的前体蛋白,其开放阅读框中不含内含子。绝大多数组织中

都有它的表达,在中枢神经系统中表达量最高。朊病毒在种间的传染性差异较大,具体是指一个种属对另一种属产生的朊病毒颗粒的侵染具有抗性的现象,其外在表现为潜伏期的延长甚至不发病。虽然这与朊病毒和宿主的基因序列差异有关,但是也不能简单地以这一标准进行判断(Goldmann *et al.*, 1996; Cleaveland *et al.*, 2001)。PRNP 结构和其编码蛋白的序列是高度保守的。人、仓鼠、小鼠、绵羊、牛、水貂之间 PRNP 的序列相似性均高于 80%。已被测定的 25 种非人灵长类动物和人 PrP 氨基酸同源率为 92.9%~99.6%。发生替代的位置是对 PrP 二级和三级结构提供特殊信息的位置,其数量已足以

收稿日期:2009-11-26 接受日期:2009-12-06

* 通讯作者 Corresponding author, E-mail: fundzou@yahoo.com.cn

确定遗传连锁。研究已经表明人、鼠、羊、牛、鹿的 PRNP 多态性对朊病毒病有连锁关系 (Franziska *et al.*, 1999; Collinge, 2001; Lysek *et al.*, 2005)。

目前针对哺乳动物 PRNP 的研究多集中在与疾病相关方向及序列变异上,对物种间 PRNP 基因的进化和利用 PRNP 序列进行系统发育研究的相对较少。Schatzl 等 (1995) 和 Rheede 等 (2003) 分别应用 PRNP 对 19 种灵长类物种和 26 种哺乳动物进行了初步的系统发育研究;Zhou 等 (2008) 也对包括哺乳动物在内的 83 种动物进行了系统发育的研究。他们的研究表明 PRNP 可以在哺乳动物的一些目、科间进行系统发育研究的应用,但所涉及的哺乳动物的种类都不是很全面,因此有必要运用更多的哺乳动物的物种进行这方面的研究。

在本研究中,我们分析了分属于 17 目 61 属哺乳动物 PRNP 编码序列的结构特征,并初步将其应用于系统发育的研究,探讨系统进化的关系和不同物种患传染性海绵状脑病的可能性之间的关系。

1 实验材料与方法

在本研究中,84 个物种的 PRNP 的编码区序列全部下载于 GenBank (详细信息可向作者索取)。这些物种代表了哺乳动物 17 目的 31 科 61 属。所有 84 条序列用软件 MEGA 3.1 (Kumar *et al.*, 2004) 进行比对,必要时进行人工校正。MEGA 3.1 也被用来进行碱基组成分析。采用 DNASP Ver. 4.10.7 (Rozas *et al.*, 2003) 对序列的保守位点、变异位点、简约信息位点以及核苷酸多样性进行分析。在 DAMBE (Xia & Xie, 2001) 中进行替代饱和检验以探求所用的序列是否饱和 (Xia *et al.*, 2003)。如果 I_{ss} (the observed saturation index) 同 $I_{ss,c}$ (the critical index) 相比较, I_{ss} 不比 $I_{ss,c}$ 小,那么所用的序列即为严重饱和,用这些序列所构建的系统树不能代表真正的系统发育树。

运用贝叶斯 (Bayes) 法构建分子系统进化树。贝叶斯分析采用软件 MrBayes v3.0 (Huelsenbeck & Ronquist, 2001),先在 Modeltest 3.7 (Posada & Crandall, 1998) 软件中基于 AIC (Akaike Information Criterion) 标准下选定最适模型,2 个独立的马尔科夫链蒙特卡罗分析同时进行 (Markov chain Monte Carlo analysis),运行参数为 4 条链同时运行,3 条热链 1 条冷链,以随机树为起始树,运行 1 000 000 代 (generation),每 100 代取样并保存数据一次,保存树的分支

长度。最早的 2500 代作为 conservative burn-in 被丢弃。运算所得最佳模型是 GTR + I + G 模型。运用 Treefinder (Jobb, 2008) 中的最大似然法 (ML) 构建系统进化树,用 bootstrap 1000 次重复随机抽样对 ML 树的拓扑结构可信度进行评估。在所有重建系统发育树的分析中,选用有袋目的 *Monodelphis domestica*、*Macropus eugenii* 和 *Trichosurus vulpecula* 作为外群,因为上述 3 个种是本研究中所有物种中最原始的一类。

2 结果和讨论

2.1 碱基组成分析

以往的研究表明,不同物种间 PRNP 编码区的序列长度变化较大 (Zhou *et al.*, 2008)。在本研究中,所下载的序列都是 PRNP 的编码区,序列长度为 738 bp (*Cercopithecus aethiops* 和 *Cercopithecus diana*) 至 810 bp (*Monodelphis domestica*),其中大部分是完全编码序列,少数是部分编码序列。所有序列经 MEGA 3.1 比对后,共有 925 个位点,其中变异位点 320 个,简约信息位点 240 个。A、T、C、G 的平均含量分别为 23.8%、19%、25.9% 和 31.2%, G + C 含量高于 A + T 含量。密码子第 3 位点表现出明显的碱基组成的不平衡性,C 的含量最高,为 36.2%,而 A 的含量最低,为 14.6%,同时具有明显的 GC 偏倚 ($G + C = 64.2\%$)。转换与颠换比的平均值为 1.8。同时还发现这 84 个物种间的 PRNP 编码区序列表现了很高的遗传多样性,核苷酸多样性达到 10.54%。这个序列片段的替代饱和和检验的结果 ($I_{ss} = 0.418$, $I_{ss,c} = 0.816$; $I_{ss} < I_{ss,c}$) 表明这些序列没有显著饱和 (Xia *et al.*, 2003)。

2.2 终止密码子的多态性

在本研究涉及到的 84 个物种中,采用 TAG 作为终止密码子的物种有 36 个,采用 TGA 的有 29 个,而使用 TAA 作为终止密码子的只有 7 个,另外有 12 个物种由于只是部分序列,没有包含终止密码子那部分序列,因此没有纳入统计。在目这个层次,食肉目、带甲目、兔形目、灵长目、啮齿目这 5 个目的物种都是以 TGA 作为终止密码子,同时原始的有袋目动物负鼠 *Monodelphis domestica* 也是使用 TGA。在由以原先的鲸目和偶蹄目合并而成的鲸偶蹄目 Cetartiodactyla 中,分化最早的猪 *Sus scrofa* 以及形态特化的鲸 *Physeter catodon* 和海豚 *Tursiops truncatus* 使用的是 TGA,而反刍动物都是使用 TAG 作为终止

密码子。在涉及到的物种中,只有 7 个物种是以 TAA 作为终止密码子,分别是奇蹄目的 5 个种和有袋目的袋鼠和袋貂 2 个种。在 Zhou 等(2008)的研究中,他们分析的结果是大部分的科是以 TGA 为终止密码子,只有少部分是以 TAG 或 TAA,这与我们的分析结果有差异。这是由于 Zhou 等(2008)的研究不仅仅包含了哺乳动物,还包括了如雉科、负子蟾科、鲤科和鸭科等非哺乳动物。总之,TGA 和 TAG 是哺乳动物 PRNP 基因的主要终止密码子。同时各个目的终止密码子的使用存在明显的偏好性。

2.3 物种间进化关系分析

由于 ML 树的拓扑结构同贝叶斯树是基本一致的,因此并没有列出 ML 树,其基于 1000 次重复的自举检验置信度标在贝叶斯树上(图 1)。ML 树同贝叶斯树的唯一显著不同是九带犰狳 *Dasypus novemcinctus* 在进化树上所处的位置不同。在贝叶斯树中,*D. novemcinctus* 是与侏食蚁兽 *Cyclopes didactylus* 聚在一起的(图 1),这与 Arnason 等(2008)的研究结果是一致的,在他们以线粒体基因组全序列为分子标记的研究中,*D. novemcinctus* 是与小食蚁兽 *Tamandua tetradactyla* 等聚在一起的;而在 ML 树中,*D. novemcinctus* 是与家兔 *Oryctolagus cuniculus* 聚在一起的(未列出 ML 树)。除此之外,我们用 PRNP 的编码区序列所建的系统进化树总体上是和以往的研究相一致的(如 Arnason *et al.*, 2008),基本反映了哺乳动物的进化关系。在图 1 中,所有灵长目(Primates)的物种聚在一起,但皮翼目 Dermoptera 的 *Cynocephalus variegatus* 插入在了其中,显示出了与灵长目物种不同寻常的很近的亲缘关系,但这与 Arnason 等(2008)的研究结果是一致的,在他们的研究中,*C. variegatus* 同样是与灵长类物种聚集在一起的。在 Arnason 等(2002)的研究中,他们以 60 种哺乳动物的线粒体基因组做为分子标记,其结果中兔形目和树鼯目与鲸偶蹄目的关系更近,而不是与啮齿类和灵长类。Arnason 等(2008)同样基于线粒体基因组分析了更多的物种,结果表明兔形目和树鼯目与啮齿类和灵长类的关系更近,这与我们的分析结果是一致的。在 Zhou 等(2008)的研究中,家猫 *Felis catus* 没有和近缘的家犬 *Canis familiaris* 聚在一起,而是和反刍动物马麝 *Moschus chrysogaster* 聚在一起。但一般来说,家猫和家犬都属于食肉目,而马麝是属于鲸偶蹄目。Zhou 等(2008)发现的是一种意外的进化关系,但并没有对这个现象进行解释。在

我们重建的系统进化树中,家猫与犬科动物和鼬科动物聚在一起构成一个单系的食肉目类群(贝叶斯后验概率 1.00),这与传统分类是一致的(如 Arnason *et al.*, 2008)。我们分析了 GenBank 可以获取的家猫的所有 5 条 PRNP 序列,除 Zhou 等(2008)的研究中使用的序列 AF003087 外,其他序列(包括我们使用的 EU588730)彼此之间非常相似且都与其他食肉目动物的序列具有很高的相似性,相反序列 AF003087 则与反刍动物的序列相似性极高,因此该序列很可能不是来源于家猫真实的序列。另外,犬蝠 *Cynopterus sphinx* 在我们重建的系统进化树中是鲸偶蹄目的姐妹群,这与 Cabria 等(2006)以 13 个线粒体上的蛋白编码基因和 9 个核基因片段作为分子标记的研究结果一致,而根据 Rheede 等(2003)的研究,犬蝠和刺猬 *Erinaceus europaeus*、毛猬 *Hylomys suillus* 等猬科 Erinaceidae 物种聚在一起。因此,我们不能凭 PRNP 最终确认犬蝠在物种进化树中的位置。综合来看,PRNP 可以作为高级分类单元,如目、科间系统发育分析的分子标记,但由于我们目前能够获取的资源有限,每个物种我们仅选择了一条序列,同时在各个属中物种也并不齐备,因此我们不能肯定 PRNP 能否较好地应用于种内以及种间进化关系的研究。

已经有报道称在鹿科、牛科、猫科、鼬科的一些物种以及人类发现的患有传染性海绵状脑病中(图 1),以反刍动物患病的占多数(Prusiner, 1997、2001),同时种属间的屏障是由宿主和受主之间 PRNP 的同源程度决定(Prusiner, 1991; Schatzl *et al.*, 1995)。虽然 PRNP 结构和其编码蛋白的序列是高度保守的,但不同物种间的 PRNP 序列不尽相同。例如,人 PRNP 与有袋类动物 PRNP 相比其氨基酸序列的相似度仅约 70%,与鸡 PRNP 相比其相似度则更低(约 30%)(Gabriel *et al.*, 1992),猪 PRNP 基因蛋白编码区与其他种哺乳动物 PRNP 序列的全部相似性为 77% ~ 88%(Martin *et al.*, 1995)。而这种相似性可以在一定程度上从系统进化树上反映出来。此外,动物传染性海绵状脑病的易感性受宿主基因组调节,特别是受 PRNP 多态性及其等位基因型的控制,其中绵羊瘙痒症的发生和易感性极易受到 PrP 密码子第 136、154、171 位多态性的影响(O'Doherty *et al.*, 2001)。灵长动物和其他哺乳动物 PrP 序列的比对表明 90 ~ 130 间的密码子对朊病毒由一种动物向另一种动物传染具有明显的影响(Schatzl

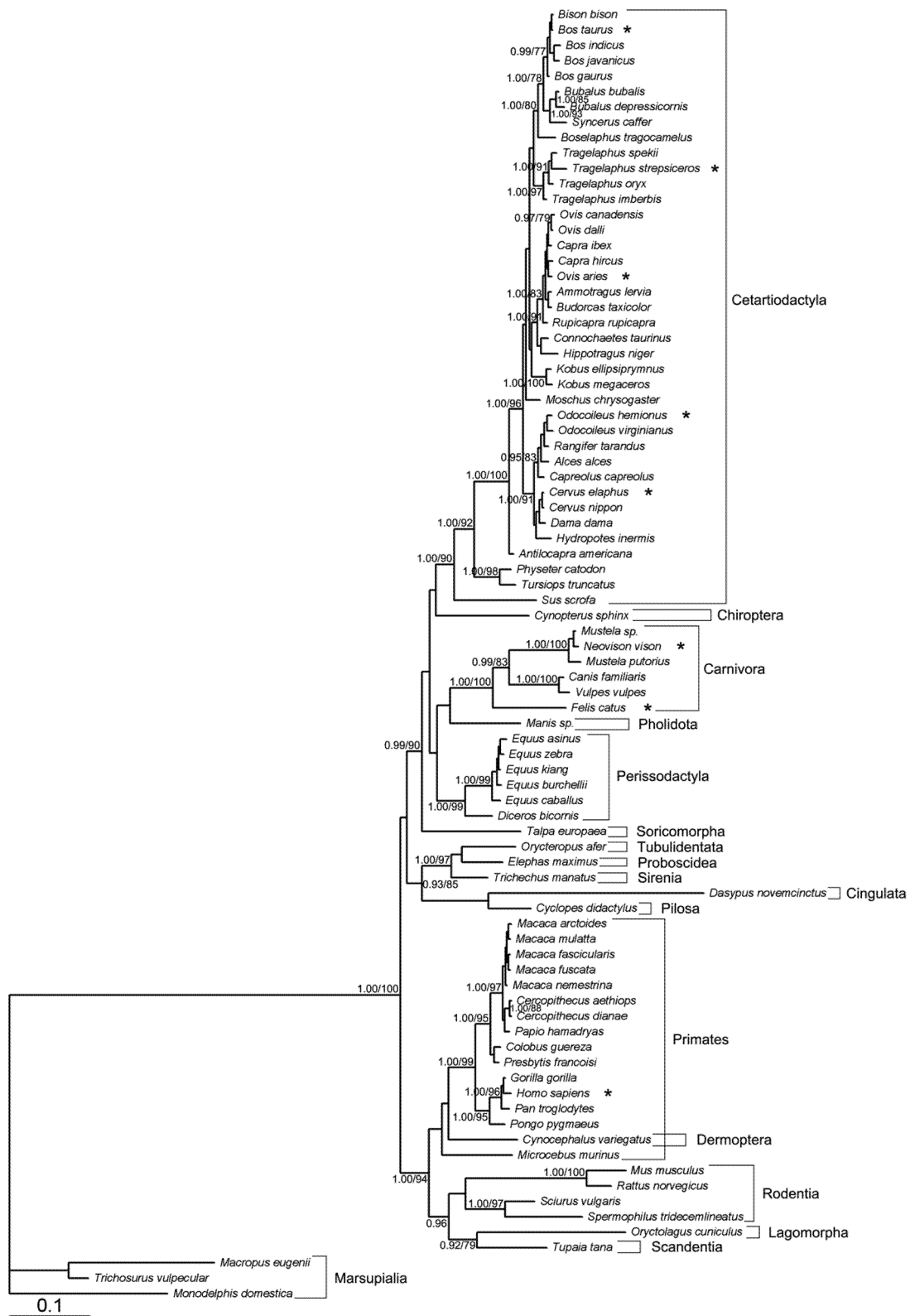


图 1 基于 PRNP 序列编码区构建的 84 种哺乳动物的系统进化树。数字分别代表的是贝叶斯分析的后验概率和 ML 分析的自举检验置信度。只有大于或等于 0.90 或 75 才在图中标明

Fig. 1 The phylogenetic tree of 84 mammalian species based on PRNP coding sequences. Numbers represent node support values inferred from Bayesian posterior probability and ML bootstrap analyses, respectively. Only values equal or greater than 0.90 or 75 are shown

注 Note: * 患朊粒病的典型动物 Typical animals with prion diseases

et al., 1995)。而且 Collinge 等(2001)研究证实人与小鼠可能存在朊病毒易感性相关基因,这对于鉴定新型人类克雅氏病有着重要的意义。从图 1 可以看出,大部分报道已经患有传染性海绵状脑病的物种都聚在一起,它们的亲缘关系较近。但有相当一部分同患有传染性海绵状脑病的物种亲缘关系较近的,还没有报道出现患病的情况,那么可以推测这些亲缘关系较近,但还没有发现患病的物种,存在患病的可能性且可能性较高。

3 参考文献

Amason U, Adegoke JA, Bodin K, *et al.* 2002. Mammalian mitogenomic relationships and the root of the eutherian tree[J]. *Proc Natl Acad Sci USA*, 99: 8151 ~ 8156.

Amason U, Adegoke JA, Gullberg A, *et al.* 2008. Mitogenomic relationships of placental mammals and molecular estimates of their divergences [J]. *Gene*, 421: 37 ~ 51.

Cabria MT, Rubines J, Gomez-Moliner B, *et al.* 2006. On the phylogenetic position of a rare Iberian endemic mammal, the Pyrenean desman (*Galemys pyrenaicus*) [J]. *Gene*, 375: 1 ~ 13.

Cleaveland S, Laurenson MK, Taylor LH. 2001. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence[J]. *Phil Trans R Soc Lond B*, 356: 991 ~ 999.

Collinge J. 2001. Prion diseases of humans and animals: their causes and molecular basis[J]. *Annu Rev Neurosci*, 24: 519 ~ 550.

Franziska W, Weidenhofer G, Schneider R, *et al.* 1999. Analysis of 27 Mammalian and 9 Avian PrPs Reveals High Conservation of Flexible Regions of the Prion Protein[J]. *J Mol Biol*, 289: 1163 ~ 1178.

Gabriel JM, Qesch B, Kretschmar H, *et al.* 1992. Molecular cloning of a candidate chicken prion protein[J]. *Proc Natl Acad Sci USA*, 89: 9097 ~ 9101.

Goldmann W, Hunter N, Somerville R, *et al.* 1996. Prion phylogeny re-

visited[J]. *Nature*, 382: 32 ~ 33.

Jobb G. , 2008. TREEFINDER version of October 2008. Munich, Germany[OL]. Distributed by the author at www.treefinder.de.

Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief [J]. Bioinform*, 5: 150 ~ 163.

Lysek DA, Schorn C, Nivon LG, *et al.* 2005. Prion protein NMR structures of cats, dogs, pigs and sheep[J]. *Proc Natl Acad Sci USA*, 3: 640 ~ 645.

Martin T, Hughes S, Hughes K, *et al.* 1995. Direct sequencing of PCR amplified pig PrP genes[J]. *Biochim Biophys Acta*, 1270(2-3): 211 ~ 214.

O'Doherty E, Aherne M, Ennis S, *et al.* 2001. Prion protein gene polymorphisms in pedigree sheep in Ireland[J]. *Res Vet Sci*, 70(1): 51 ~ 56.

Prusiner SB. 1991. Prion biology and diseases[J]. *Harvey Lect*, 87: 85 ~ 114.

Prusiner SB. 1997. Prion diseases and the BSE crisis[J]. *Science*, 278: 245 ~ 251.

Prusiner SB. 2001. Neurodegenerative Diseases and Prions[J]. *N Engl J Med*, 344:1516 ~ 1526.

Rheede TV, Smolenaars MMW, Madsen O, *et al.* 2003. Molecular Evolution of the Mammalian Prion Protein[J]. *Molecular Biology and Evolution*, 20: 111 ~ 121.

Schatzl HM, Costa MD, Taylor L, *et al.* 1995. Prion Protein Gene Variation Among Primates[J]. *Journal of Molecular Biology*, 245: 362 ~ 374.

Xia X, Xie Z. 2001. DAMBE: data analysis in molecular biology and evolution[J]. *J Hered*, 92: 371 ~ 373.

Xia X, Xie Z, Salemi M. *et al.* 2003. An index of substitution saturation and its application[J]. *Mol Phylogenet. Evol*, 26: 1 ~ 7.

Zhou R, Li X, Li L, *et al.* 2008. Evolution and Differentiation of the Prion Protein Gene (PRNP) among Species[J]. *Journal of Heredity*, 99: 647 ~ 652.

(上接第 554 页)

Rintamäki PT, Stone JR, Lundberg SA. 2003. Seasonal and diurnal body-mass fluctuations for two nonhoarding species of *parus* in Sweden modeled using path analysis[J]. *Auk*, 120(3): 658 ~ 670.

Rowan W, 1925. Relation of light to bird migration and developmental changes[J]. *Nature*, 115: 494 ~ 495.

Summers-Smith JD. 1989. A history of the status of the tree sparrow *Passer montanus* in the British Isles[J]. *Bird Study*, 36(1): 23 ~ 31.

Svensson M, Rintamäki PT, Birkhead TR, *et al.* 2007. Impaired hatching

success and male-biased embryo mortality in tree sparrows[J]. *J Ornithol*, 148: 117 ~ 122.

Zhang SP, Zhang GM, Xu JL. 2008. Habitat use of urban tree Sparrows in the process of urbanization: Beijing as a case study[J]. *Front Biol China*, 3(3): 308 ~ 314.

Zheng WH, Li M, Liu JS, *et al.* 2008. Seasonal acclimatization of metabolism in Eurasian tree sparrows (*Passer montanus*) [J]. *Comp Bio Physiol*, 151(A): 519 ~ 525.