

基于共同前缀位置的哺乳动物mtDNA序列系统发育分析

陆超, 王代

辽宁师范大学, 辽宁 大连
Email: 1250332458@qq.com

收稿日期: 2021年2月23日; 录用日期: 2021年3月22日; 发布日期: 2021年3月29日

摘要

本文基于共同前缀标识符, 通过环形前缀树法构建DNA序列的环形前缀树, 得到能够表征序列信息的前缀集, 去除生物序列长度对序列间距离计算的影响的同时并考虑序列对的前缀位置等因素, 创建了一种新的非比方法——基于共同前缀的位置差法, 选取含有80条哺乳动物的mtDNA序列的数据集作为测试对象, 运用邻接(Neighbor-Joining)法构建系统发育树, 快速且准确地得到符合生物学分类的结果。

关键词

哺乳动物, DNA序列, 前缀集, 系统发育分析

Phylogenetic Analysis of Mammalian mtDNA Sequences Based on Common Prefix Positions

Chao Lu, Dai Wang

Liaoning Normal University, Dalian Liaoning
Email: 1250332458@qq.com

Received: Feb. 23rd, 2021; accepted: Mar. 22nd, 2021; published: Mar. 29th, 2021

Abstract

This paper is based on the common prefix identifier. A circular prefix tree of DNA sequences is constructed by the circular prefix tree method to obtain a prefix set that can characterize se-

quence information, which removed the influence of biological sequence length on the calculation of the distance between sequences, and considered the common prefix position of sequence pairs. And other factors, created a new non-comparison method—the position difference method based on the common prefix. The data set containing 80 mammalian mtDNA sequences was selected as the test object, and the phylogenetic tree was constructed using the Neighbor-Joining method, which could fast and accurately get the results that meet the biological classification.

Keywords

Mammal, DNA Sequence, PrefixSet, Phylogenetic Analysis

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

20 世纪 50 年代以来, 随着分子生物学和分子遗传学的兴起, 生物学家根据各种生物间在分子水平上的进化关系, 建立分子进化的系统发育树(Phylogenetic Tree), 估测物种间的亲缘关系, 直观地阐明生物间的进化历程。

系统发育树重建传统上取决于多个或成对的序列比对。但是, 使用基于对对的方法分析大型数据集时会遇到各种限制。高等真核生物的全基因组比对可能超过计算内存[1]。因为存在基因组重排和重复的组合等因素使得不能进行整个基因组的比对, 并且过多的序列信息增加了计算生物学中基因组比较的计算量和时间要求。因此, 必须在初始步骤中确定正在研究的基因组的可比同源片段, 并且迫切需要更快的序列分析算法。为此, 提出了非比对方法克服基于比对方法的局限性[2] [3] [4]。序列比较的非比对方法可以定义为在算法应用的任何步骤中量化序列相似性/相异性的任何方法, 非比对方法不使用或产生比对(残基—残基对应的分配)。它们算法复杂度低且可以抵抗改组和重组等比对方法不能处理的事件。

非比对方法可大致分为两组: 基于定义长度子序列频率的方法(基于词的方法)和评估全长序列之间信息内容的方法(基于信息理论的方法)。还有一些方法不能分类在任何一组中, 包括那些基于匹配词长度的方法, 混沌游戏表示, 迭代映射, 以及 DNA 序列的图形表示, 它以定量的方式捕获基本组成的实质和序列的分布。本文是通过一种新的非比对方法——基于共同前缀的位置差算法(CPPA 算法: Common Prefix Position-difference Algorithm)构建 80 中胎盘类哺乳动物的系统发育树, 能够快速而准确地得到聚类结果, 并且优于分布向量法[5]的结果。

2. 材料与方法

2.1. 材料

2020 年 10 月 22 日从 NCBI 的 Genbank 数据库中下载的 80 条有胎盘类(Placentalia)哺乳动物的线粒体完整基因组, 基因组的平均长度为 1.65×10^3 碱基对。因为它是在大多数多细胞生物中从母亲那里继承而来的, 不是高度保守, 而且突变率不是很高, 所以它对于研究基于线粒体基因组的进化关系非常有效。本文将 Bo 等人(2011) [5]中 80 个物种中含有非 ACGT 的序列进行了同属或同科替换, 保证本方法的有效性, 替换后的数据集详细信息见附录。

2.1.1. 前缀标识符

在有限字母 $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_i\}$ 上指定的长度为 m 符号序列 $P = P\langle 1 \rangle P\langle 2 \rangle \dots P\langle m \rangle$, 其中 $P\langle i \rangle \in \Sigma$ 是序列中的第 i 个符号, 并且被称为位于第 i 个位置。 $P = P\langle i:j \rangle$ 表示从 P 的位置 i 开始到 j 的位置的字符串。与 Σ 上的字符串 S 相关联的前缀树是 t 叉树 $T_p = (N, n_0, S_p)$, $W_p(n)$ 是 S_p 的最小长度唯一子串, 其最左边的符号位于 S 的位置 j 。 $W_p(n) = S\langle j:k \rangle$ 在 S 中仅发生一次, 而 $S\langle j:k-1 \rangle$ 在 S 中至少发生两次。则称 $I(j) = S\langle j:k \rangle$ 为与位置 j 相关的前缀标识符[6]。

本文的 CPPA 算法的前缀树模型是以 Weiner Peter 的[6]前缀标识符理论为基础提出的, 并通过将张欣[7]的相异度量模型中线性前缀树算法改进得到的。

2.1.2. 环形生物序列前缀集

哺乳动物的线粒体基因组是长度约为 1.65×10^3 碱基对的双链 DNA 环[8]。但是从 NCBI (National Center for Biotechnology Information Search database)网站中获取的序列是线性的, 因此将序列首尾进行拼接, 还原为环形序列, 再基于前缀标识符, 将线性前缀树算法[7]改进为环形前缀树算法构建前缀树, 提取前缀集。

对任一序列 $S = S_1, S_2, S_3, \dots, S_{n-1}, S_n, S_i \in \{A, C, G, T\}, 1 \leq i \leq n, n = \text{length}(S)$, 将序列 S 足够长的前 m 个碱基拼接在 S 末端, 构建新序列 $S' = S_1, S_2, S_3, \dots, S_{n-1}, S_n, S_{n+1}, S_{n+2}, \dots, S_{n+m}, m \ll n, n+m = \text{length}(S')$, 对 $\forall i, 1 \leq i \leq n, \exists$ 唯一前缀 $S_i_prefix = S_i, S_{i+1}, \dots, S_{i+l_i}, l_i = \text{length}(S_i_prefix)$, s.t. 当 $l_j = l_i (j \neq i)$ 时, 满足 $S_j_prefix \neq S_i_prefix$, 且至少存在一个 S_k_prefix , 满足 $S_k_prefix = S_i, S_{i+1}, \dots, S_{i+l_i-1}, l_k = l_i - 1$ 。由 $S_i_prefix, 1 \leq i \leq n$ 构成的集合 $I(S)$ 称为序列 S 的前缀集。

2.1.3. 序列间基于共同前缀的位置法

DNA 序列 $S_i, S_j (i \neq j)$ 的长度分别为 n_1, n_2 。首先应用环形前缀树法构建两条序列的前缀树, 得到前缀集 $I(S_i), I(S_j)$ 。再求出两条序列共同前缀集 $I(S_i) \cap I(S_j)$ 的元素在 S_i, S_j 中的标准化位置, 去除序列长度对计算序列间距离的影响。第 t 个共同前缀在 S_i 中的标准化位置定义为:

$$P_i^t = Pr_i^t / n_i \tag{1}$$

Pr_i^t 为第 t 个共同前缀在序列 S_i 中的实际位置。

假设 S_i, S_j 共有 m 个共同前缀, 所有共同前缀在 S_i, S_j 中标准化位置构成如下位置向量:

$$P_i = (P_i^1, P_i^2, \dots, P_i^m), P_j = (P_j^1, P_j^2, \dots, P_j^m)。$$

两条序列的共同前缀的位置做差, 得到位置差向量。考虑到哺乳动物的线粒体基因组是环形的, 第 t 个共同前缀都的位置在序列中的位置如图 1 所示。

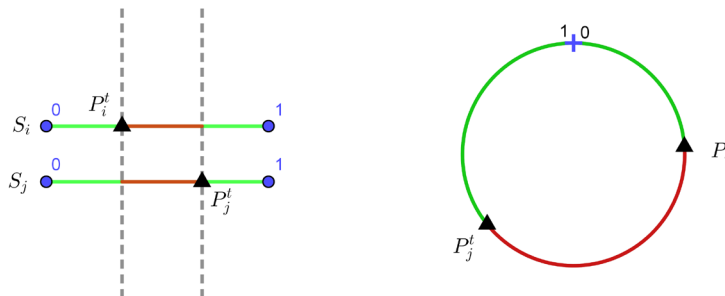


Figure 1. The relative position of the common prefix in the sequence pair
图 1. 共同前缀在序列对中的相对位置

S_i, S_j 的第 t 个共同前缀的位置一同放在一个长度为 1 的环形 DNA 中, 则距离差会出现两种结果红色弧线段的长度或绿色弧线段的长度, 因为共同前缀位置之间的距离越小, 则两序列越相似, 因此取两种位置差中较小的数值作为所求的位置差。因此, 两条序列 S_i, S_j 的位置差向量 $position_diff$ 定义为:

$$position_diff = (d_1, d_2, \dots, d_m), \text{ 其中 } d_t = \min\{|P_i^t - P_j^t|, 1 - |P_i^t - P_j^t|\}, t = 1, 2, \dots, m \quad (2)$$

本文采用基于共同前缀的位置差算法来计算进化距离。因为序列间共同前缀越多, 序列越相似, 所以将序列对 S_i, S_j 间的相似性距离定义为:

$$dis(S_i, S_j) = \frac{1}{m} |position_diff| = \frac{1}{m} \sqrt{\sum_{t=1}^m (d_t)^2} = \frac{1}{m} \sqrt{\sum_{t=1}^m (\min\{|P_i^t - P_j^t|, 1 - |P_i^t - P_j^t|\})^2} \quad (3)$$

获得了用于系统发育分析的所有遗传序列之间的距离构成的距离矩阵后, 使用 Matlab2013a 通过邻接法(NJ)绘制系统发育树[9]。

3. 结果

3.1. 系统发育分析

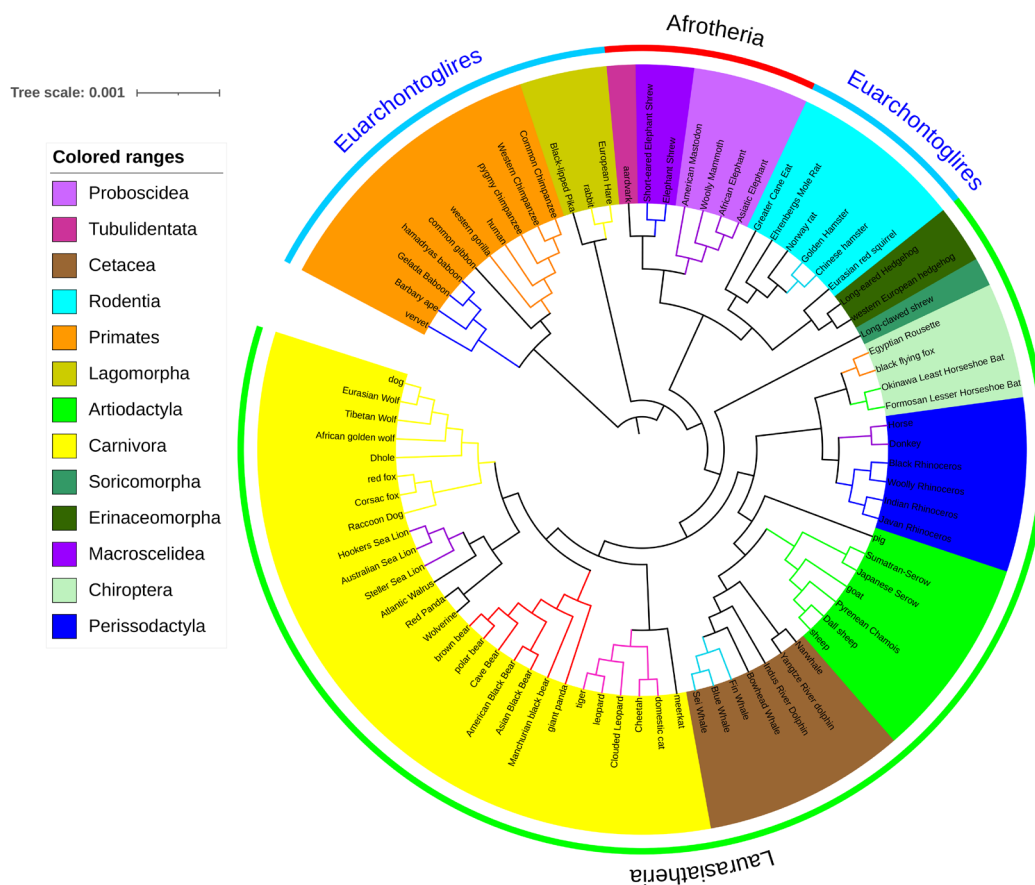


Figure 2. The Phylogenetic tree of 80 mammals by CPPA
图 2. 由 CPPA 法得到的 80 种哺乳动物的系统发育树

80 条哺乳动物线粒体全基因组 mtDNA 序列数据集中, 全部是有胎盘类, 其中非洲兽总目(Afrotheria) 7 条序列, 北方兽类(Boreoeutheria) 73 条序列。北方兽类包含灵长总目(Euarchontoglires) 19 条序列, 劳亚

兽总目(Laurasiatheria) 54 条序列[10] [11] [12]。

从整体上看, 图 2 中 CPPA 法得到的系统发育树中, 非洲兽总目正确聚类, 灵长总目未聚为一个分支, 劳亚兽总目的序列来自 6 个目, 5 个目的 52 条序列聚在一起; 其它目的 2 条序列与灵长总目的啮齿目聚到一起。

数据集中, 劳亚兽总目的序列具体是: 食肉目(carnivora) 27 条序列、鲸偶蹄目 14 条序列、奇蹄目(Perissodactyla) 6 条序列、翼手目(Chiroptera) 4 条序列、鼯形目(Soricomorpha) 1 条序列和猬形目(Erinaceomorpha) 2 条序列, 6 个目都分别聚成一个分支。食肉目的犬形亚目(Caniformia)与猫型亚目(Feliformia)分别聚为一个分支[13], 鲸偶蹄目中的 3 条序列鲸目(cetacea)与偶蹄目(Artiodactyla)的 6 条反刍亚目(Ruminantia)和 1 条猪次目(Suina)正确聚类, 并与交错在一起, 支持鲸偶蹄目的分类[14]。

灵长总目的序列来自 3 个目, 灵长目(Primates)、兔形目(Lagomorpha)和啮齿目(Rodentia)。灵长总目虽然没有形成一个分支, 灵长目和兔形目各自形成自己的分支, 且聚类正确[15]。啮齿目的 6 条序列来自 3 个亚目, 鼠形亚目(Myomorpha) 4 条序列, 松鼠亚目(Sciuromorpha) 3 条序列和豪猪亚目(Hystricomorpha) 1 条序列[16]。每个亚目各自形成单独的分支。

在科的层级上, 含 2 条及以上的科共 16 个, 均正确聚类。

3.2. CPPA 法与分布向量法系统发育树比较

Bo Zhao 和 Rong L. He 等人的分布向量法构建的系统发育树[5], 仅考虑了生物之间的聚类关系, 因此本文在聚类的角度上进行比较。

一、目的层级上, 分布向量法的聚类结果中 27 条食肉目序列被分为两部分, 且猫型亚目与犬形亚目均未聚类正确, 14 条鲸偶蹄目序列被分为鲸目和偶蹄目两部分, 不支持鲸偶蹄目的分类, CPPA 法的食肉目与鲸偶蹄目均聚类正确, 6 条啮齿目序列的聚类 CPPA 法略差于分布向量法。

二、科的层级上, 分布向量法的聚类结果中 4 条犀牛科(Rhinocerotidae)序列、3 条海狮科(Otariidae)序列、7 条犬科(Canidae)、3 条须鲸科(Balaenopteridae)序列、5 条人科(Hominidae)序列、4 条猕猴科(Cercopithecidae)序列和 2 条兔科(Leporidae)序列 7 个科未聚类正确, CPPA 法均聚类正确。

以上结果说明由 80 条哺乳动物线粒体全基因组 mtDNA 序列数据集构建的系统发育树中, CPPA 法的聚类结果远好于分布向量法的聚类结果。

4. 结论

本文基于共同前缀标识符提出的 CPPA 法, 首先通过环形前缀树算法获得每条序列的前缀集, 挖掘到能表征序列的定量代表——共同前缀在序列对中所处的标准化位置, 去除了序列长度对序列间距离的影响, 并充分考虑了共同序列个数在计算序列间相似性距离的重要性, 应用 Matlab 运行算法速度快, 且聚类结果准确, 说明 CPPA 法是一种好的非比方法。通过 CPPA 法得到的系统发育树中物种间的进化关系还不够好, 可以考虑影响序列间距离的其它因素进行改进。

致 谢

感谢 Bo 等人为本研究提供了 80 条哺乳动物线粒体序列的 NCBI 检索号。同时, 也要对本文中引用文章的作者表示衷心的感谢。

参考文献

- [1] Saw, A.K., Raj, G., Das, M., et al. (2019) Alignment-Free Method for DNA Sequence Clustering Using Fuzzy Integral Similarity. *Scientific Reports*, 9, Article No. 3753. <https://doi.org/10.1038/s41598-019-40452-6>

-
- [2] Vinga, S. and Almeida, J. (2003) Alignment-Free Sequence Comparison—A Review. *Bioinforma*, **19**, 513-523. <https://doi.org/10.1093/bioinformatics/btg005>
- [3] Zielezinski, A., Vinga, S., Almeida, J. and Karlowski, W.M. (2017) Alignment-Free Sequence Comparison: Benefits, Applications, and Tools. *Genome Biology*, **18**, Article No. 186. <https://doi.org/10.1186/s13059-017-1319-7>
- [4] Bernard, G., Chan, C.X., Chan, Y.-B., *et al.* (2017) Alignment-Free Inference of Hierarchical and Reticulate Phylogenomic Relationships. *Briefings in Bioinformatics*, **20**, 426-435.
- [5] Zhao, B., He, R.L. and Yau, S.S.-T. (2011) A New Distribution Vector and Its Application in Genome Clustering. *Molecular Phylogenetics and Evolution*, **59**, 438-443. <https://doi.org/10.1016/j.ympev.2011.02.020>
- [6] Weiner, P. (1973) Linear Pattern Matching Algorithms. *Proceedings of the 14th Annual IEEE Symposium on Switching and Automata Theory*, USA, 15-17 October 1973, 1-11. <https://doi.org/10.1109/SWAT.1973.13>
- [7] 张欣. 基于后缀树的DNA序列进化树构建研究[D]: [硕士学位论文]. 大连: 辽宁师范大学, 2019。
- [8] Anderson, S., De Bruijn, M.H.L., Coulson, A.R., *et al.* (1982) Complete Sequence of Bovine Mitochondrial DNA Conserved Features of the Mammalian Mitochondrial Genome. *Journal of Molecular Biology*, **156**, 683-717. [https://doi.org/10.1016/0022-2836\(82\)90137-1](https://doi.org/10.1016/0022-2836(82)90137-1).
- [9] Saitou, N. and Nei, M. (1987) The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**, 406-425.
- [10] Murphy, W.J., Eizirik, E., Johnson, W.E., *et al.* (2001) Molecular Phylogenetics and the Origins of Placental Mammals. *Nature*, **409**, 614-618. <https://doi.org/10.1038/35054550>
- [11] Kjer, K.M. and Honeycutt, R.L. (2007) Site Specific Rates of Mitochondrial Genomes and the Phylogeny of Eutheria. *BioMed Central Evolutionary Biology*, **7**, Article No. 8. <https://doi.org/10.1186/1471-2148-7-8>
- [12] Xie, X.H., Yu, Z.G., Han, G.S., Yang, W.F. and Anh, V. (2015) Whole-Proteome Based Phylogenetic Tree Construction with Inter-Amino-Acid Distances and the Conditional Geometric Distribution Profiles. *Molecular Phylogenetics and Evolution*, **89**, 37-45. <https://doi.org/10.1016/j.ympev.2015.04.008>
- [13] Agnarsson, I., Kuntner, M. and May-Collado, L.J. (2010) Dogs, Cats, and Kin: A Molecular Species-Level Phylogeny of Carnivora. *Molecular Phylogenetics and Evolution*, **54**, 726-745. <https://doi.org/10.1016/j.ympev.2009.10.033>
- [14] Zurano, J.P., Magalhães, F.M., Asato, A.E., *et al.* (2019) Cetartiodactyla: Updating a Time-Calibrated Molecular Phylogeny. *Molecular Phylogenetics and Evolution*, **133**, 256-262. <https://doi.org/10.1016/j.ympev.2018.12.015>
- [15] Dos Reis, M., Gunnell, G.F., Barba-Montoya, J., *et al.* (2018) Using Phylogenomic Data to Explore the Effects of Relaxed Clocks and Calibration Strategies on Divergence Time Estimation: Primates as a Test Case. *Systematic Biology*, **67**, 594-615. <https://doi.org/10.1093/sysbio/syy001>
- [16] Steppan, S.J. and Schenk, J.J. (2017) Muroid Rodent Phylogenetics: 900-Species Tree Reveals Increasing Diversification Rates. *PLoS ONE*, **12**, e0183070. <https://doi.org/10.1371/journal.pone.0183070>

附录

Table 1. Description of mitochondrial genome of 80 mammals

表 1. 80 条哺乳动物线粒体序列

NO	NCBI	Common Name	Super Orders	Order	Family
1	AB096867	Elephant Shrew	Afrotheria	Macroscelidea	
2	NC_004026	Short-eared Elephant Shrew	Afrotheria	Macroscelidea	
4	DQ316068	Asiatic Elephant	Afrotheria	Proboscidea	Elephantidae
3	AJ224821	African Elephant	Afrotheria	Proboscidea	Elephantidae
5	NC_007596	Woolly Mammoth	Afrotheria	Proboscidea	Elephantidae
6	NC_009574	American Mastodon	Afrotheria	Proboscidea	Mammutidae
7	Y18475	aardvark	Afrotheria	Tubulidentata	
9	NC_004028	European Hare	Euarchontoglires	Lagomorpha	Leporidae
8	AJ001588	rabbit	Euarchontoglires	Lagomorpha	Leporidae
10	NC_011029	Black-lipped Pika	Euarchontoglires	Lagomorpha	Ochotonidae
18	NC_009747	vervet	Euarchontoglires	Primates	Cercopithecidae
14	NC_002764	Barbary ape	Euarchontoglires	Primates	Cercopithecidae
17	Y18001	hamadryas baboon	Euarchontoglires	Primates	Cercopithecidae
19	FJ785426	Gelada Baboon	Euarchontoglires	Primates	Cercopithecidae
12	D38114	western gorilla	Euarchontoglires	Primates	Hominidae
15	V00662	human	Euarchontoglires	Primates	Hominidae
13	D38116	pygmy chimpanzee	Euarchontoglires	Primates	Hominidae
11	D38113	Common Chimpanzee	Euarchontoglires	Primates	Hominidae
20	GU112744	Western Chimpanzee	Euarchontoglires	Primates	Hominidae
16	X99256	common gibbon	Euarchontoglires	Primates	Hylobatidae
23	EU660217	Chinese hamster	Euarchontoglires	Rodentia	Cricetidae
26	NC_013276	Golden Hamster	Euarchontoglires	Rodentia	Cricetidae
22	X14848	Norway rat	Euarchontoglires	Rodentia	Muridae
21	AJ238588	Eurasian red squirrel	Euarchontoglires	Rodentia	Sciuridae
25	NC_005315	Ehrenberg's Mole Rat	Euarchontoglires	Rodentia	Spalacidae
24	NC_002658	Greater Cane Eat	Euarchontoglires	Rodentia	Thryonomyidae
28	AF533441	goat	Laurasiatheria	Artiodactyla	Bovidae
27	AF010406	sheep	Laurasiatheria	Artiodactyla	Bovidae
33	NC_012096	Japanese Serow	Laurasiatheria	Artiodactyla	Bovidae
30	FJ207534	Sumatran-Serow	Laurasiatheria	Artiodactyla	Bovidae
31	FJ207538	Pyrenean Chamois	Laurasiatheria	Artiodactyla	Bovidae
32	NC_039432	Dall sheep	Laurasiatheria	Artiodactyla	Bovidae
29	AJ002189	pig	Laurasiatheria	Artiodactyla	Suidae
57	NC_009691	Red Panda	Laurasiatheria	Carnivora	Ailuridae
45	NC_027956	African golden wolf	Laurasiatheria	Carnivora	Canidae
44	NC_008434	red fox	Laurasiatheria	Carnivora	Canidae

Continued

40	NC_023958	Corsac fox	Laurasiatheria	Carnivora	Canidae
42	U96639	dog	Laurasiatheria	Carnivora	Canidae
58	NC_011218	Tibetan Wolf	Laurasiatheria	Carnivora	Canidae
56	NC_009686	Eurasian Wolf	Laurasiatheria	Carnivora	Canidae
47	GU063864	Dhole	Laurasiatheria	Carnivora	Canidae
60	NC_013700	Raccoon Dog	Laurasiatheria	Carnivora	Canidae
39	EF551003	tiger	Laurasiatheria	Carnivora	Felidae
50	NC_005212	Cheetah	Laurasiatheria	Carnivora	Felidae
41	U20753	domestic cat	Laurasiatheria	Carnivora	Felidae
38	EF551002	leopard	Laurasiatheria	Carnivora	Felidae
54	NC_008450	Clouded Leopard	Laurasiatheria	Carnivora	Felidae
51	NC_045900	meerkat	Laurasiatheria	Carnivora	Herpestidae
55	NC_009685	Wolverine	Laurasiatheria	Carnivora	Mustelidae
48	NC_004029	Atlantic Walrus	Laurasiatheria	Carnivora	Odobenidae
52	NC_008418	Hooker's Sea Lion	Laurasiatheria	Carnivora	Otariidae
49	NC_004030	Steller Sea Lion	Laurasiatheria	Carnivora	Otariidae
53	NC_008419	Australian Sea Lion	Laurasiatheria	Carnivora	Otariidae
46	NC_011117	Manchurian black bear	Laurasiatheria	Carnivora	Ursidae
37	EF212882	giant panda	Laurasiatheria	Carnivora	Ursidae
43	AF303109	American Black Bear	Laurasiatheria	Carnivora	Ursidae
34	AF303110	brown bear	Laurasiatheria	Carnivora	Ursidae
35	AF303111	polar bear	Laurasiatheria	Carnivora	Ursidae
59	NC_011112	Cave Bear	Laurasiatheria	Carnivora	Ursidae
36	DQ402478	Asian Black Bear	Laurasiatheria	Carnivora	Ursidae
61	AJ554051	Bowhead Whale	Laurasiatheria	Cetacea	Balaenidae
66	NC_006929	Sei Whale	Laurasiatheria	Cetacea	Balaenopteridae
64	NC_001601	Blue Whale	Laurasiatheria	Cetacea	Balaenopteridae
63	NC_001321	Fin Whale	Laurasiatheria	Cetacea	Balaenopteridae
67	NC_007629	Yangtze River dolphin	Laurasiatheria	Cetacea	Lipotidae
62	AJ554062	Narwhale	Laurasiatheria	Cetacea	Monodontidae
65	NC_005275	Indus River Dolphin	Laurasiatheria	Cetacea	Platanistidae
68	NC_023122	black flying fox	Laurasiatheria	Chiroptera	Pteropodidae
71	NC_007393	Egyptian Rousette	Laurasiatheria	Chiroptera	Pteropodidae
69	NC_005433	Formosan Lesser Horseshoe Bat	Laurasiatheria	Chiroptera	Rhinolophidae
70	NC_005434	Okinawa Least Horseshoe Bat	Laurasiatheria	Chiroptera	Rhinolophidae
72	X88898	western European hedgehog	Laurasiatheria	Erinaceomorpha	
73	NC_005033	Long-eared Hedgehog	Laurasiatheria	Erinaceomorpha	
76	NC_001788	Donkey	Laurasiatheria	Perissodactyla	Equidae
75	NC_001640	Horse	Laurasiatheria	Perissodactyla	Equidae

Continued

77	NC_012681	Woolly Rhinoceros	Laurasiatheria	Perissodactyla	Rhinocerotidae
78	NC_012682	Black Rhinoceros	Laurasiatheria	Perissodactyla	Rhinocerotidae
79	NC_012683	Javan Rhinoceros	Laurasiatheria	Perissodactyla	Rhinocerotidae
74	X97336	Indian Rhinoceros	Laurasiatheria	Perissodactyla	Rhinocerotidae
80	AB061527	Long-clawed shrew	Laurasiatheria	Soricomorpha	Soricidae
