

# Construction and Analysis of the Avian Phylogenetic Tree

Yongfen Zhang, Xun Zhou, Liaofu Luo\*, Lirong Zhang\*

School of Physical Science and Technology, Inner Mongolia University, Hohhot Inner Mongolia  
Email: \*lolfcm@imu.edu.cn, \*pyzlr@imu.edu.cn

Received: May 15<sup>th</sup>, 2017; accepted: May 25<sup>th</sup>, 2017; published: May 31<sup>st</sup>, 2017

---

## Abstract

Based on nucleotide sequences in 394 conserved regions of 47 avian DNA sequences obtained using targeted-next generation DNA sequencing, the k-mer frequency (from 9-mer to 14-mer) was counted. We calculated the distance matrix among 47 avians by k-mer Non-aligned Algorithm (KNA) and constructed the phylogenetic tree by the Neighbor-Joining method. The results showed that the phylogenetic tree is changed with increasing k and stabilized when k equals 12. Then, we compared 12-mer phylogenetic tree with two other trees constructed by Prum and Jarvis respectively and analyzed the evolution and classification of these birds. We found that the three phylogenetic trees are basically same apart from a small part of sister branches on the trees. The consistency revealed that the frequency of 12-mer is a better dynamic variable for measuring evolution of species.

## Keywords

Avian, DNA Sequence, K-Mer, Phylogenetic Tree

---

# 鸟类基因组进化树的构建与分析

张永芬, 周 勋, 罗辽复\*, 张利绒\*

内蒙古大学物理科学与技术学院, 内蒙古 呼和浩特  
Email: \*lolfcm@imu.edu.cn, \*pyzlr@imu.edu.cn

收稿日期: 2017年5月15日; 录用日期: 2017年5月25日; 发布日期: 2017年5月31日

---

## 摘 要

本文选取了47种鸟类的基因组作为研究对象, 根据靶定下一代DNA测序技术(targeted next-generation

\*通讯作者。

DNA sequencing)得到的394个保守片段的DNA序列数据,统计了从9联体到14联体的频数及其频率,基于k联体(k-mer)非联配算法得到了任意两种鸟类之间k联体的距离矩阵,并运用邻接(Neighbor-Joining)法构建了47种鸟类的进化树,发现当k = 12时达到稳定。最后,通过将该树与Prum和Jarvis构建的进化树进行了比较,分析了鸟类的进化和分类。结果发现三个进化树基本一致,只有部分姐妹分支有差别,说明12联体的相对频数是能够较好描述基因组进化的动力学变量。

## 关键词

鸟类, DNA序列, k联体, 进化树

Copyright © 2017 by authors and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

鸟类起源于侏罗纪时代的兽脚亚目,是恐龙的唯一现存后裔。同时鸟类也是四足类脊椎动物中物种最丰富的一个纲,目前大约有 10,000 多个鸟类物种存在于地球的不同环境中。在过去数十年中,人们对鸟类之间物种进化关系的研究有了巨大的进展和重大的突破[1]。研究认为,大约在 6600 万年前小行星撞击地球或是火山爆发而发生了著名的白垩纪 - 第三纪灭绝事件(Artistic depiction of asteroidal impact),这次大范围的物种大灭绝事件造成了恐龙的灭亡与哺乳动物的兴起,同时也为鸟类新物种的爆发创造了条件且提供了新的生存环境。在白垩纪时期,所有现代鸟的祖先发生了第一次分化,出现了一些飞行能力不太强的鸟类祖先,如鸵鸟的祖先。这次分化之后形成的一个姐妹支又出现了一次分化,而产生了现在的鸡、鸭等的祖先[2]。但是,到目前为止仍有许多关于鸟类分类和进化的问题未得到解决。这是因为鸟类在早期进化过程中经历了那次辐射性的物种大爆发事件,而爆发时间很短使得现存鸟类之间的进化关系难以区分。虽然长期以来研究者众多,但是在 2010 年前 2/3 以上的目、科分类级别关系较为混乱。同时,在大爆炸事件前后 1000~1500 万年内形成的新鸟小纲(Neoaves)是占据现存鸟类物种中绝大比例的一支[3] [4] [5] [6],但目前为止仍然无法确定新鸟小纲中一些早期分支的分类情况和具体的进化规律。因此,鸟类进化树的构建和分析有助于了解鸟类的进化历史,对研究鸟类的进化过程和进化规律具有重要的生物学意义,研究中所用的方法对其它类群进化问题的研究同样具有重要的参考价值。

近年来,随着测序技术水平的快速提高,人类获得了更加丰富的鸟类基因组数据,鸟类分子系统发生学的研究也因此进入了快车道。2014 年, Jarvis 等人[3]基于 48 种鸟类全基因组数据,利用系统发育基因组学的方法研究了鸟类的起源与进化。他们重建了一个可信度极高的鸟类进化树(可信度 90%以上),证实了此前备受争议的姐妹群,重新确定了鸟类的进化关系。在这个树中,鸟类物种被分成古颚总目(Palaeognathae)和今颚总目(Neognathae),古颚总目是其余所有现存鸟类的姐妹群,今颚总目由鸡雁小纲(Galloanseres)和新鸟小纲组成。Jarvis 等发现,占现存鸟类 95%物种的新鸟小纲可以分成两大分支,分别被命名为雀类(Passerea)和鸽类(Columbea),这两大分支分别独立趋同进化出了各自的陆生和水生鸟类物种。2015 年, Prum 等人[7]利用靶定下一代 DNA 测序技术,完成了对 198 种鸟类物种的测序,构建了 Zenodo 数据库,从而建立了更加全面的系统发育树。在这个系统发育关系中,鸟类物种同样被分成古颚总目和今颚总目,鸡雁小纲和新鸟小纲组成今颚总目。但是在单源的新鸟小纲内, Prum 等确定了五个主要分支,分别命名为 Strisores, Columbaves, Gruiformes, Aequorlitorinithes 和 Inopinaves,且对它们进行了

更为详细的分类。建树过程中运用的时间参数以得到的地质化石较准时间为准。

尽管鸟类的基因组信息越来越全面，但在构建鸟类进化树中也遇到了诸多困难。其中一个主要原因是鸟类物种大爆发的时间很短，导致鸟类基因组之间存在大量的基因不完全谱系分选(Incomplete Lineage Sorting of Genes) [8] [9]现象，从而给进化树构建带来不确定性：由单个基因构建出来的进化树很难与由全基因组构建的物种树一致。另外，全基因组上某些区域会有较多的噪音，这些噪音会影响物种进化树的推断[3]。因此，研究者需要从全基因组序列中仔细挑选出适合构树的高质量区域。

本文选取两侧为可变区域的保守序列数据，基于 k 联体的非联配方法，构建了一个由 47 个鸟类物种组成的鸟类基因组进化树。所谓 k 联体的非联配方法就是生物序列数据频次统计算法，也称为 k-mer 频次统计算法。遗传语言-DNA 序列的信息很大程度上反映在其 k-mer 频数中，k-mer 频次统计信息可以用来揭示生物序列中各种子序列的分布规律，它是一种衡量序列相似性的重要工具，这为研究鸟类物种亲缘关系提供了新的角度[10]。我们分别统计了 9 联体到 14 联体的频率，计算了不同基因组之间的距离，构成一个多维的距离矩阵，利用邻接法构建出能反映 47 个鸟类物种之间进化关系的进化树，我们称之为 IMU(Inner Mongolia University 的简称)进化树。在计算过程中发现，随着 k 的增加，IMU 进化树越来越稳定。当 k 增大到 12 时，进化树已经不再随 k 值的增加而变化，由此说明 12 联体对于鸟类基因组具有物种特征性，这也意味着 12 联体频数是能够较好描述进化的动力学变量。同时通过观察 IMU 进化树的拓扑结构，发现 5 种古老鸟类处于树的根部，其余 42 种现代鸟类形成了 Neoaves1(陆鸟)，Neoaves2(陆鸟和水鸟的混合)，Neoaves3(包含猛禽的陆鸟)，Neoaves4(水鸟)等 4 支。最后，将 IMU 进化树与 Prum 和 Jarvis 等人构建的进化树进行了比较，分析了鸟类的进化和分类情况。

## 2. 方法

### 2.1. 数据来源

本文的数据来源于 Zenodo(<https://zenodo.org/record/28343#.VzCY5PmEBFn>)，由 Prum 等[7]人于 2015 年所构建，是由靶定下一代 DNA 测序方法得到的序列数据，其中包含可以代表鸟类所有主要分支的 198 个鸟类和 2 个鳄目类(crocodylians)共 200 个物种的数据。本文通过比较由 Prum 等人研究的 198 种鸟类与 Jarvis 等人研究的 48 种鸟类属的一致性，从而将两种不同的数据名称进行了对应，最终得到了能够代表不同属的 47 种鸟类，如表 1 所示。例如，Prum 等人研究的 198 种鸟中只有一种企鹅，为洪堡企鹅；而 Jarvis 等人研究的 48 种鸟中同为企鹅属的有两种企鹅，分别为帝企鹅和洪堡企鹅；故本文应用洪堡企鹅的数据。因此，由英文常用名和 Prum 等人命名的鸟类一定对应相同的属，但对应的具体鸟类可能不同。为方便与不同进化树进行比较及分析，本文均用英文常用名表述鸟类。具体地，每一种鸟类的 DNA 序列包含由靶定基因组中约 259 个基因所产生的 394 个序列片段，这些片段多数是两侧为可变区域的保守序列，其中每个基因平均约有 1524 个碱基，最后每种鸟类的 DNA 序列约有 390,000 个碱基。

在建树过程中，时间参数依据 Prum 等得到的地质化石较准时间，即文献[7]附录中图 7 所示的时间，由此确定了每一物种的产生时间。

### 2.2. 本文研究的 47 种鸟类名称和记号

为方便与 Prum 和 Jarvis 构建的进化树进行比较，本文对 47 种鸟类的英文常用名、记号、对应于英文常用名的中文常用名[2]、以及 Prum 对鸟类的命名进行了一一对应(如表 1 所示)。字母记号的规则为：前面的大写字母表示多层次分类，之后的两位数字对应物种在其分类中的编号，最后一位小写英文字母代表编号后更加细致的分类，若没有则不写。

因此，198 种鸟类的字母记号具体如下：P01-P09 表示 Palaeognathae, G01-G16 表示 Galloanserae,

**Table 1.** The name and mark of 47 avians  
**表 1.** 47 种鸟类的名称和记号

序号	英文常用名	中文常用名	记号	Prum 对鸟类的命名
1	Common ostrich	非洲鸵鸟	P01	I0453_90121_Struthionidae_Struthio_camelus
2	White-throated tinamou	白喉鸨	P06t	I0432_5886_Tinamidae_Eudromia_elegans
3	Turkey	火鸡	G04g	I0448_52755_Numididae_Numida_malagris
4	Chicken	鸡	G09g	I0374_2419_Phasianidae_Gallus_gallus
5	Peking duck	北京鸭	G12a	I0283_6026_Anatidae_Dendrocygna_viduata
6	Downy woodpecker	绒啄木鸟	NIC22	I0259_3344_Picidae_Picus_canus
7	Carmin bee-eater	红蜂虎	NIC10	I0199_955_Meropidae_Merops_muelleri
8	Zebra finch	斑马雀	NIA49p	I0456_3321_Paridae_Poecile_lugubris
9	Medium-ground finch	中地雀	NIA50p	I0455_1245_Alaudidae_Calandrella_brachydactyla
10	American crow	短嘴鸦	NIA48p	I0197_832_Corvidae_Corvus_albus
11	Golden-collared manakin	金领娇鹀	NIA41p	I0246_2823_Tyrannidae_Tyrannus_albogularis
12	Rifleman	刺鹩	NIA13p	I2829_RIFL_Acanthisittidae_Acanthisitta_chloris
13	Speckled mousebird	斑胸鼠鸟	NIC01	I0410_52764_Coliidae_Colius_colius
14	Bar-tailed trogon	斑尾非洲咬鹃	NIC05	I0190_295_Trogonidae_Apaloderma_aequatoriale
15	Chimney swift	烟囱褐雨燕	NS10a	I0247_2886_Apodidae_Chaetura_brachyura
16	Anna's hummingbird	安氏蜂鸟	NS11a	I0219_1489_Trochilidae_Topaza_pella
17	Common cuckoo	大杜鹃	NC06o	I0423_2497_Cuculidae_Cuculus_optatus
18	Kea	啄羊鹦鹉	NIA07	I0389_14873_Psittacidae_Nestor_meridionalis
19	Budgerigar	虎皮鹦鹉	NIA12	I0240_2217_Psittacidae_Deroptyus_accipitrinus
20	Chuck-will's-widow	卡氏夜鹰	NS02	I0258_3336_Caprimulgidae_Caprimulgus_europaeus
21	Yellow-throated sandgrouse	黄喉沙鸡	NC11c	I0411_53231_Pteroclididae_Pterocles_bicinctus
22	Brown mesite	褐拟鹧	NC09c	I0363_345610_Mesitornithidae_Mesitornis_unicolor
23	Sunbittern	日鹮	NA19	I0396_16818_Eurypyidae_Eurypyga_helias
24	Hoatzin	麝雉	NIM01	I0369_848_Opisthocomidae_Opisthocomus_hoazin
25	Red-crested turaco	红冠蕉鹃	NC01o	I0192_375_Musophagidae_Tauraco_macrorhynchus
26	Mac Queen's bustard	亚洲波斑鸨	NC03o	I0281_5967_Otididae_Ardeotis_kori
27	Pigeon	鸽子	NC16c	I0211_1232_Columbidae_Columba_livia
28	Turkey vulture	红头美洲鹫	NIM02a	I0202_1066_Cathartidae_Cathartes_burrovianus
29	Cuckoo-roller	鹃三宝鸟	NIC03	I0367_449184_Leptosomidae_Leptosomus_discolor
30	Rhinoceros hornbill	马来犀鸟	NIC08	I2828_3250_Bucerotidae_Bucorvus_leadbeateri
31	Peregrine falcon	游隼	NIA04	I0236_2079_Falconidae_Falco_sparverius
32	Bald eagle	白头海雕	NIM07a	I0185_86_Accipitridae_Buteo_jamaicensis
33	White-tailed eagle	白尾海雕	NIM08a	I0268_3790_Accipitridae_Accipiter_superciliosus
34	Red-legged seriema	红腿叫鹤	NIA01	I0421_1173_Cariamidae_Cariama_cristata

## Continued

35	Barn owl	仓鸮	NIM09	I0452_88784_Tytonidae_Tyto_alba
36	American flamingo	美洲火烈鸟	NA01	I0287_6069_Phoenicopteridae_Phoenicopterus_ruber
37	Great-crested grebe	凤头鸕鷀	NA02	I0422_1201_Podicipedidae_Rollandia_rolland
38	Killdeer	双领鸨	NA04	I0212_1289_Charadriidae_Charadrius_hiaticula
39	Grey-crowned crane	灰冠鹤	NG08	I0433_6296_Gruidae_Balearica_regulorum
40	White-tailed tropicbird	白尾鸲	NA20	I0375_2613_Phaethontidae_Phaethon_rubricauda
41	Red-throated loon	红喉潜鸟	NA21	I0449_66928_Gaviidae_Gavia_immer
42	Great cormorant	普通鸬鹚	NA36	I0372_1658_Phalacrocoracidae_Phalacrocorax_brasilianus
43	Little egret	小白鹭	NA39	I0245_2800_Ardeidae_Ardea_herodias
44	Dalmatian pelican	卷羽鹈鹕	NA43	I0391_15014_Pelecanidae_Pelecanus_occidentalis
45	Crested ibis	朱鹮	NA37	I0204_1174_Threskiornithidae_Phersticus_caerulescens
46	penguin	企鹅	NA22	I0451_85124_Spheniscidae_Spheniscus_humboldti
47	Northern fulmar	暴风鹱	NA27	I0404_21828_Procellariidae_Fulmarus_glacialoides

NS01-NS13 表示 Neoaves-Strisores, NC01-NC16 表示 Neoaves-Columbaves, NG01-NG09 表示 Neoaves-Gruiforms, NA01-NA43 表示 Neoaves-Aequorlornithes, NIM01-NIM10 表示与 Neoaves-Inopinaves-Accipitriformes 相关的混合鸟类, NIC01-NIC26 表示 Neoaves-Inopinaves-Coraciimorphae, NIA01-NIA56 表示 Neoaves-Inopinaves-Australaves, NIA13p-NIA56p 中的字母 p 表示 subclass passeriformes。通过分析新命名的字母记号可以直接了解物种分类的从属问题, 方便快速判断所构进化树的准确性。

### 2.3. 数据预处理

$k$  联体是指在核苷酸序列中连续排列的  $k$  个碱基。由于下载得到的测序数据中含有碱基不确定的位点(占序列总长的 0.56%~5.76%), 如强配对碱基(S), 次黄嘌呤碱基(I), 含氨基碱基(M)等, 为了保证序列长度不变, 同时避免引入虚假的  $k$  联体, 将这些位点用“-”表示。这样, 经过预处理后的序列数据则是由腺嘌呤(A)、胸腺嘧啶(T)、胞嘧啶(C)、鸟嘌呤(G)四种碱基及“-”组成。

本文不考虑含有“-”的  $k$  联体, 故仅包含字母的  $k$  联体最多只有  $4^k$  种。我们从经预处理后的 47 种鸟类的基因组序列资料出发, 统计  $k$  联体( $k = 9 - 14$ )的频数及其相应频率。据此构建进化树。

### 2.4. 核苷酸 $k$ 联体的非联配算法

我们采用  $k$  联体的非联配算法[11] (KNA 算法:  $k$ -mer Non-aligned Algorithm)来计算进化距离。对于任意两保守序列  $\Sigma_i$  和  $\Sigma_j$ , 设其核苷酸  $\sigma$  型  $k$  联体的频率分别为  $\{p_{\sigma i}\}$  和  $\{p_{\sigma j}\}$ , 则两序列之间的距离定义为:

$$d(\Sigma_i, \Sigma_j) = \sum_{\sigma} |p_{\sigma i} - p_{\sigma j}| \quad (i, j = 1, 2, \dots, 47) \quad (1)$$

根据(1)式, 利用  $k$  联体频率可计算出 47 种鸟类两两之间的进化距离值, 从而得到一个  $47 \times 47$  维的距离矩阵。

### 2.5. 用 NJ 算法构建鸟类进化树

邻接法(Neighbor-Joining, 简称为 NJ)是由 Saitou 和 Nei [12]于 1987 年提出的, 是一种基于距离矩阵的快速聚类方法。该算法不需要关于分子钟的假设, 不考虑任何优化标准。其基本思想是进行类的合并时



不仅要求待合并的类是相近的,而且要求待合并的类远离其他的类,从而通过对完全没有解析出的星状进化树进行分解,来不断改善星状树。在递推过程中 NJ 算法不仅考虑一对对的距离,而且考虑了整个树中距离之和的极小化。

特别是,在 NJ 算法的递推过程中要对一对物种(如物种 1 和物种 2)的分支长度求和,记为  $S_{12}$ :

$$S_{12} = \frac{1}{2(N-1)} \sum_{k=3}^N (d_{1k} + d_{2k}) + \frac{1}{2} d_{12} + \frac{1}{(N-2)} \sum_{3 \leq i \leq j \leq N} d_{ij} \quad (2)$$

其中,  $N$  表示物种总数,  $d_{1k}$  和  $d_{2k}$  分别表示物种 1 和 2 与其它物种之间的距离值,  $d_{12}$  表示物种 1 和 2 之间的距离值,  $d_{ij}$  表示除了物种 1 和 2 之外的其他物种两两之间的距离值。当  $S_{12}$  为求得的所有  $S_{ij}$  中最小值时,表示 1 和 2 即为邻近物种。很显然,式(2)中物种 1 和物种 2 是对称的,即两物种替换后  $S_{12}$  值不变。由此说明一个姐妹支上的两个邻近物种是可以上下互换的。

### 3. 结果

#### 3.1. 频数统计结果

对 47 种鸟类基因组中的靶定测序数据,我们计算了从 9-mer 到 14-mer 的 k-mer 的频数分布,如表 2 所示。结果表明,对于多联体而言,频数为 2-4 和频数大于 4 的 k-mer 占 k-mer 总体的比例随着 k 的增大而降低,而频次为 1 的 k-mer 随着 k 的增大而增大。

例如,对于 12 联体,我们计算了暴风鹫(Northern fulmar)、企鹅(penguin)、红头美洲鹫(Turkey vulture)、鸽子(Pigeon)、麝雉(Hoatzin)和非洲鸵鸟(Common ostrich)共 6 种鸟类频数大于 4 的 12-mer 的分布,如图 1 所示。其中,横坐标表示 12-mer 的频数,范围从 5 到 20,纵坐标表示给定频数下 12-mer 的种类数。分析发现,基本满足随着频数的增加对应种类数减小的趋势。通过统计较高频次 12-mer 的分布可深入探究 IMU 树中各姐妹支形成的原因。

#### 3.2. 利用距离矩阵构建 IMU 进化树

本文基于 KNA 算法,我们构建了 47 种鸟类的进化树,称为 IMU(Inner Mongolia University 的简称)树。对所得的 6 个 k-mer ( $k = 9 - 14$ )进化树进行比较,可得出较合理的 IMU 树。经分析发现从 9-mer 进化树开始,5 种古鸟就位于树的根部,与已知的进化关系相符;紧邻其上有一支 NI,树的最上面有一支 NA,这些分枝已经达到稳定。当 k 增加到 12 时,得到 12 联体时的 NJ 进化树,就和 Jarvis 和 Prum 等构建的进化树基本一致。当 k 的值由 12 增加到 13、14 时,发现进化树已不再随 k 的增加而发生任何变化。由此说明,12 联体对于鸟类基因组具有物种特征性。我们得到的 IMU 进化树如图 2 所示,以“P/G”和“N”

**Table 2.** The proportion of k-mer with different frequency

**表 2.** 不同频数 k-mer 的占比

	k 联体总数范围	频数为 1 的比例(%)	频数为 2-4 的比例(%)	频数大于 4 的比例(%)
9 联体	339,155~388,500	17.59~20.75	53.18~55.92	23.48~29.21
10 联体	334,966~387,994	53.98~58.62	39.75~43.64	1.64~2.42
11 联体	330,896~387,507	82.61~85.04	14.86~17.20	0.077~0.14
12 联体	326,894~387,039	94.57~95.46	4.73~5.42	0.015~0.041
13 联体	323,140~386,587	98.88~98.56	1.31~1.65	0.0037~0.03
14 联体	319,475~386,113	99.46~99.59	0.40~0.52	0.0013~0.025

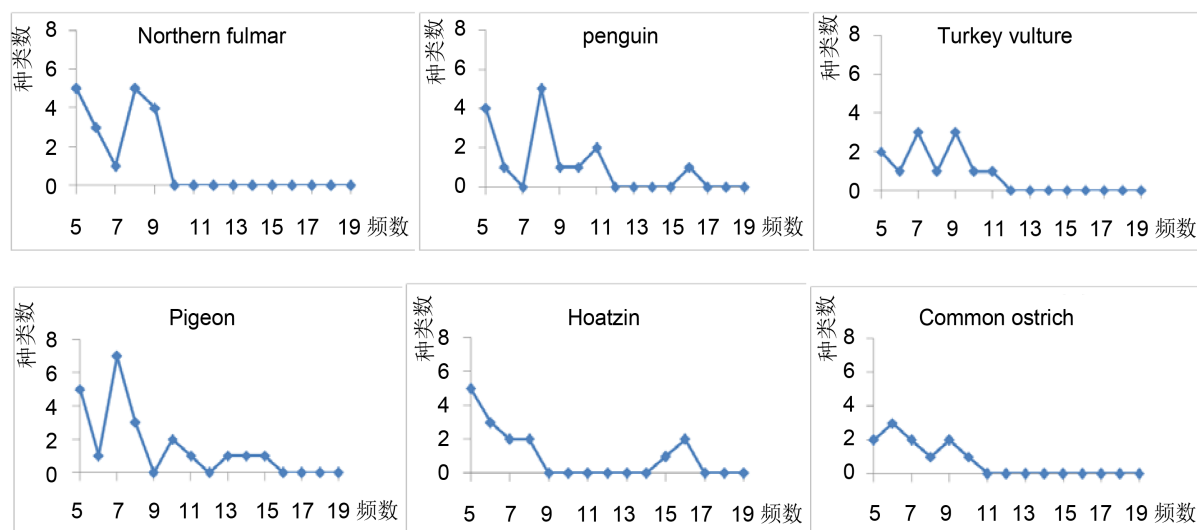


Figure 1. The distribution of 12-mer frequency in six kinds of avian DNA sequences  
图 1. 6 种鸟类 DNA 序列的 12 联体频数分布

开头分别代表了古鸟和新鸟,可分为五部分,分别为古老鸟,新鸟小纲的 1、2、3、4 共五支,各自代表了古老鸟、陆鸟、水陆混合鸟类、陆鸟和水鸟,图中分别以紫、黑、蓝、红、绿色的线表示。其中,古老鸟类(Palae/Gallo)、猛禽(Neoaves3 中的 NIM02a、NIA04、NIM07a、NIM08a、NIA01、NIM09)、水鸟(Neoaves4)的分类非常明确。此外,物种形成时间取自文献[7]中的补充图 7。

## 4. 讨论

### 4.1. Prum 和 Jarvis 等构建进化树的比较

由 Prum 等人构建的进化树[7]和 Jarvis 等人构建的进化树[3]是目前最为可信的两棵鸟类生命之树。通过对这两棵进化树的拓扑结构进行比较,我们分析了它们之间的异同。

两树一致把鸟类分为古颚总目和今颚总目,且鸡雁小纲和新鸟小纲共同组成今颚总目。两树也同样认为蜂鸟科,雨燕科和夜鹰科形成姐妹分支,其拓扑结构为(蜂鸟科 + 雨燕科) + 夜鹰科(三者都属于夜鹰目)。沙鸡目,拟鹑目和鸽形目形成姐妹分支,其拓扑结构为(沙鸡目 + 拟鹑目) + 鸽形目。美洲火烈鸟(火烈鸟科)和凤头鹮(鹮科)形成姐妹分支。日鸕目,鸕蒙目和潜鸟目形成姐妹分支,其拓扑结构为(日鸕目 + 鸕蒙目) + 潜鸟目。对于鸽形目,蕉鹑目和鸕形目,两树都认为它们是近亲,然而物种的拓扑结构有些差异, Jarvis 等人认为(鸕形目 + (蕉鹑目 + 鸕形目)), Prum 等人认为(蕉鹑目 + (鸕形目 + 鸕形目))。

对于水生鸟类,两树一致认为鸕形目和企鵝目有密切关系。鸕鹚科,鸕科,鹭科和鸕鹚科形成连续的姐妹分支,其拓扑结构为(鸕鹚科 + 鸕科) + 鹭科 + 鸕鹚科(四者都属于鸕形目),且鸕形目与鸕形目-企鵝目进化支形成姐妹群。对于陆生鸟类,两树一致的认为美洲鹭科与鸕科是近亲,形成姐妹分支(两者同属于鸕形目),在此我们将其称为陆鸟群 1。鸕形目,鼠鸟目,鸕鹚目,咬鸕目,犀鸟目,鸕形目和佛法僧目形成连续的姐妹分支,其拓扑结构为(鸕形目 + 佛法僧目) + 犀鸟目 + 咬鸕目 + 鸕鹚目 + 鼠鸟目 + 鸕形目,在此我们将其称为陆鸟群 2。雀形目,鸕形目,隼形目和叫鸕目形成连续的姐妹分支,其拓扑结构为(雀形目 + 鸕形目) + 隼形目 + 叫鸕目,在此我们将其称为陆鸟群 3。在新鸟小纲中,两树一致的认为陆生鸟类的共同祖先是顶端食肉动物。

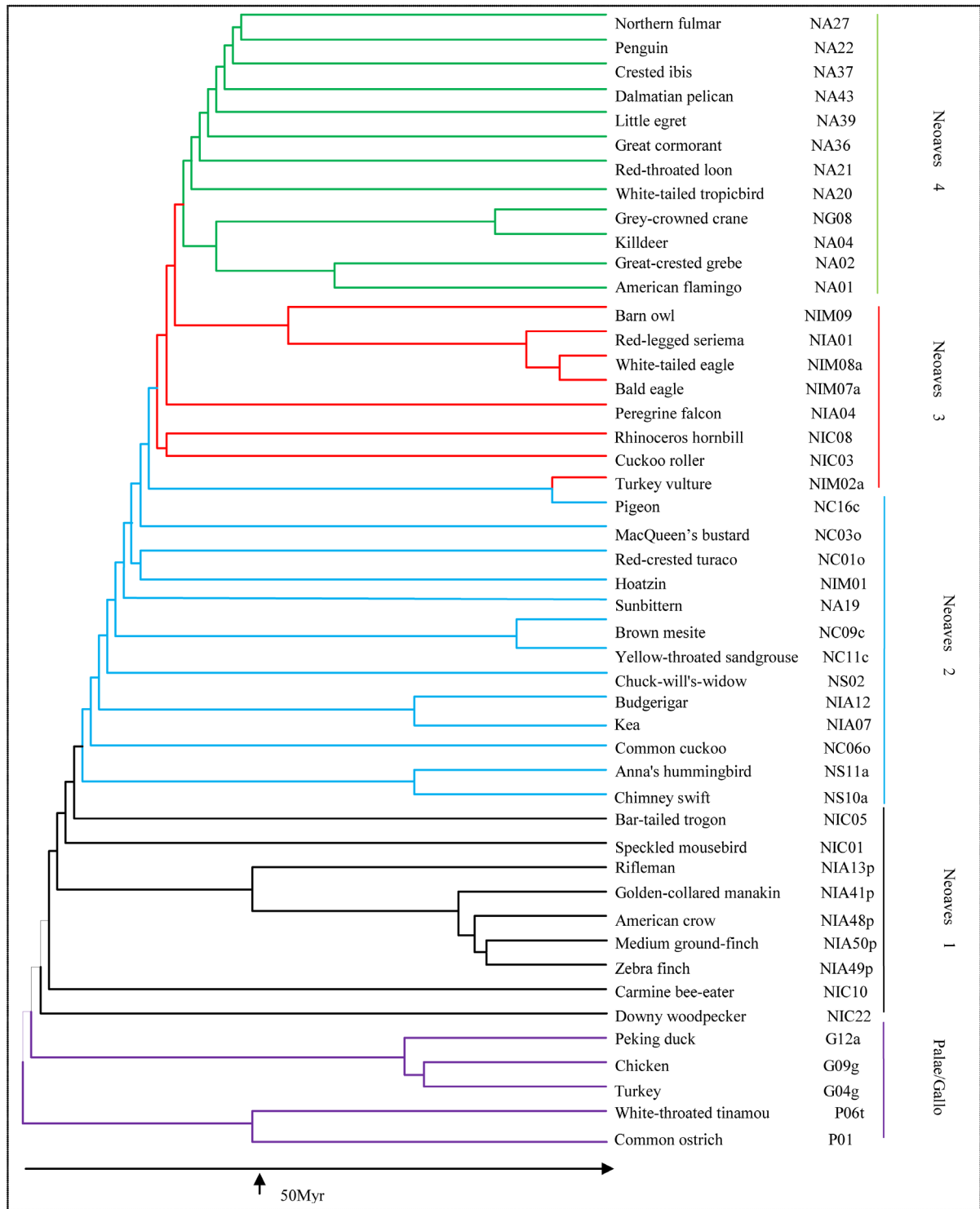


Figure 2. The Phylogenetic tree of 47 avians by KNA algorithm (IMU tree)

图 2. 由 KNA 算法得到的 47 种鸟类的进化发生树(IMU 树)

对于两树拓扑结构的差异，可归纳为以下四点：其一，两树虽一致的认为麝雉应该自成一目，但是 Jarvis 等人认为麝雉与鸨形目，鹤形目形成姐妹群；而 Prum 等人构建的进化树将其放置在陆鸟的基部，



作为其余所有陆鸟的姐妹群，而且认为鹤形目应该自成一分支，与所有的陆鸟和水鸟形成姐妹分支。其二，Jarvis 等人认为火烈鸟科-鸕鷀科与鸽形目是近亲，且紧邻鸡雁小纲; Prum 等人认为火烈鸟科-鸕鷀科与岸鸟(鸕鷀形目等)是近亲，紧邻鹤形目，紧邻鸡雁小纲的是鹰形目。其三，Jarvis 等人认为(蜂鸟科 + 雨燕科) + 夜鹰科)进化支与(鸕鷀形目 + (蕉鹃目 + 鸨形目)进化支形成姐妹群; Prum 等人认为(沙鸡目 + 拟鹑目) + 鸽形目)进化支与(蕉鹃目 + (鸕鷀形目 + 鸨形目)进化支形成姐妹群。其四，在整个陆生鸟类这个大进化支内，Jarvis 等人认为拓扑结构为(陆鸟群 1 + 陆鸟群 2) + 陆鸟群 3)，而 Prum 等人认为拓扑结构为(陆鸟群 1 + (陆鸟群 2 + 陆鸟群 3))。

## 4.2. IMU 进化树与 Prum 和 Jarvis 等构建进化树的比较分析

Jarvis 和 Prum 等人构建的进化树都是基于序列联配方法构造的联配进化树，而 IMU 进化树是基于非联配方法构造的。因此，IMU 进化树为解析 Jarvis 和 Prum 等构建进化树之争提供了一个新的观察角度。通过三树比较，它们的拓扑结构基本一致，但存在一些差异，分析如下：

一，从进化树整体上看，三树一致把鸟类分为古颚总目和今颚总目，且鸡雁小纲和新鸟小纲共同组成今颚总目，鸡雁小纲由鸡形目和雁形目构成，古颚总目和鸡雁小纲位于进化树的基部。

二，在 IMU 进化树中，鸕鷀形目与蜂鸟科-雨燕科进化支形成姐妹群，这与 Jarvis 等人构建的进化树和 Prum 等人构建的进化树都不同。在 IMU 进化树中，鸕鷀形目和鹤形目形成姐妹群，这与 Jarvis 等人构建的进化树相同。然而，在 IMU 树中鸕鷀形目-鹤形目进化支与火烈鸟科-鸕鷀科进化支形成姐妹分支，但在 Prum 等人构建的进化树，鸕鷀形目与火烈鸟科-鸕鷀科进化支形成姐妹分支，鹤形目自成一分支。麝雉是一个较为特殊的物种，它在三树中的位置各不相同，IMU 进化树认为麝雉与红冠蕉鹃(蕉鹃目)是近亲，Jarvis 等人构建的进化树中麝雉与鸕鷀形目，鹤形目形成姐妹分支，Prum 等人构建的进化树将其放置在陆鸟的基部，与所有陆鸟形成姐妹群。

三，在 IMU 进化树中，仓鸨，红腿叫鹤，白尾海雕，白头海雕和游隼这些食肉类猛禽形成了连续的姐妹群，而另外两个进化树中分别位于陆鸟基部。同时，鸨形目，佛法僧目，雀形目，鼠鸟目和咬鹃目形成连续的陆鸟姐妹群，这个陆鸟姐妹群不同于在另外两个进化树中处于进化树的中下端，且紧邻鸡雁小纲，这个结果与 Slack[13]得到的相近。对于水生鸟类，三树的拓扑结构是一致的，新奇的是在 IMU 进化树中水鸟处于进化树的最顶端，其另外两树不同。

四，在 IMU 进化树中，鸨形目与夜鹰科形成了姐妹分支，日鳶目与拟鹑目-沙鸡目进化支形成姐妹群，这些关系都是独特且新奇的。特别需要指出的是鸨科和美洲鸨科，在 IMU 树中，它们形成了姐妹分支，而在另外两树中，鸨科与沙鸡科-拟鹑科进化支形成了姐妹群。

以上结果说明部分鸟类在 IMU 进化树中与在另外两树中的分类有差异，为了探究其原因，本文以鸨科分类为例，从序列数据出发研究了其保守序列中频数大于 4 的 12-mer 类型(如表 3 所示)。在 IMU 树中鸽子与红头美洲鸨形成姐妹分支，而在另外两树中鸽子与黄喉沙鸡、褐拟鹑形成姐妹分支。对比分析发现，鸽子与红头美洲鸨的 12-mer 高频次类型具有高度相似性(如表 3 所示，具有相同颜色、下划线等标记的 12-mer 相同)，这应该是导致在 IMU 树中鸨科和美洲鸨科形成姐妹分支的主要原因。我们猜测这些重复性极高的 12 联体同时出现在两类鸟的基因组中可能说明它们有某些相似的功能。这是一个鸟的分类研究中值得注意的问题。

## 5. 结论

本文利用靶定下一代 DNA 测序技术得到的 394 个保守片段的 DNA 序列数据，基于 k 联体的非联配算法构建了 47 种鸟类的 k(k = 9 - 14)联体 NJ 进化树，我们称之为 IMU 进化树。结果显示随着 k 增加到

**Table 3.** The 12-mers with a frequency greater than 4 in 4 avians  
**表 3.** 4 种鸟类中频数大于的 4 的 12 联体

频数	5	6	7	8	9-10	>10
Pigeon	AAAGAAAAGA ACCTCCTCCTTT CTGCAGCAGCAG CTTCTTCTTCTT TTCTTCTTCTG	<u>CTCCTCTTCCTC</u>	TTTTTTTTTTTT TGGGATGGGATG GGGATGGGATGG GGATGGGATGGG GATGGGATGGGA ATGGGATGGGAT AAAAAAAAAAAA A	<u>TGGTGGTGGTGG</u> <u>GTGGTGGTGGTG</u> <u>GGTGGTGGTGGT</u>	<u>TCCTCCTCCTCC</u> <u>CTCCTCCTCCTC</u>	<u>CCCTCCTCCTCT</u> <u>GCTGCTGCTGCT</u> <u>CTGCTGCTGCTG</u> <u>TGCTGCTGCTGC</u>
Turkey vulture	TCCTTCTCCTCC CCCAGCAGCAGC	<u>CTCCTCTTCCTC</u>	<u>TGGTGGTGGTGG</u> <u>GTGGTGGTGGTG</u> <u>GGTGGTGGTGGT</u>	<u>GCTGCTGCTGCT</u>	<u>CTGCTGCTGCTG</u> <u>CTCCTCCTCCTC</u> <u>CCTCCTCCTCCT</u> <u>TCCTCCTCCTCC</u>	<u>TGCTGCTGCTGC</u>
Brown mesite	CAGAAGCTGAAG CAGCAGCAGCAG CAGCTGGAGGAG CCAGCAGCAGCT TCCTCCTCCTCC CCTCCTCCTCCT CTCCTCCTCCTC GGTGGTGGTGGT GTGGTGGTGGTG TCCTCCTCCTCC TTTTTTTTTTTT	CTCCTCTTCCTC TTTTTTTTTTTT	CCCAGCAGCAGC	GGTGGTGGTGGT GTGGTGGTGGTG	CTGCTGCTGCTG GCTGCTGCTGCT TGCTGCTGCTGC	TGCTGCTGCTGC
Yel- low-thr oated sandgro use	GGTGGTGGTGGT G	TGGTGGTGGTG G		TGCTGCTGCTGC	GCTGCTGCTGCT CTGCTGCTGCTG	

注：表中用相同颜色和下划线等标记的两个 12-mer 是相同的，通过比较鸽子与红头美洲鹫的 12-mer 高频次类型，发现它们具有高度相似性。

12 时进化树达到稳定，由此表明 12 联体对于鸟类基因组具有物种特征性，这也意味着 12 联体频数是能够较好描述进化的动力学变量。同时通过观察 IMU 进化树的拓扑结构，发现其明确的分为 5 支。最后，比较了 IMU 进化树与 Prum 和 Jarvis 等构建进化树的异同，分析了鸟类的进化和分类情况，发现三树的各个姐妹分支基本一致。但处于三个树根部的新鸟分类还有明显不同。这可能暗示发生物种大爆发的过程中尚存在某些有待探索的规律。如果核苷酸频数是好的动力学变量，我们就可以写出关于这些变量的时间进化方程，进而把基因组进化方程量子化[14]。这样就可以用量子随机性来解释新鸟的大爆发式的产生。这是一个有趣的有待进一步研究的问题。

## 致 谢

感谢基金项目：国家自然科学基金资助项目(批准号：61462068 和 31106188)和内蒙古自治区自然科学基金资助项目(批准号：2014MS0103)对本论文的大力支持。感谢 Prum 等人为本研究提供了鸟类及其对应化石时间数据。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 (References)

- [1] Zhang, G., Li, C., Li, Q., *et al.* (2014) Comparative Genomics Reveals Insights into Avian Genome Evolution and Adaptation. *Science*, **346**, 1311-1320. <https://doi.org/10.1126/science.1251385>
- [2] 胡皓夫. 用基因组学解码鸟类进化史[J]. 科学, 2015, 67(2): 21-25.
- [3] Jarvis, E.D., Mirarab, S., Aberer, A.J., *et al.* (2014) Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science*, **346**, 1320-1331. <https://doi.org/10.1126/science.1253451>
- [4] Hedges, S.B., Parker, P.H., Sibley, C.G., *et al.* (1996) Continental Breakup and the Ordinal Diversification of Birds and Mammals. *Nature*, **381**, 226-229. <https://doi.org/10.1038/381226a0>
- [5] Cooper, A. and Penny, D. (1997) Mass Survival of Birds across the Cretaceous-Tertiary Boundary: Molecular Evidence. *Science*, **275**, 1109-1113. <https://doi.org/10.1126/science.275.5303.1109>
- [6] Simon, Y.W.H. and Matthew, J.P. (2009) Accounting for Calibration Uncertainty in Phylogenetic Estimation of Evolutionary Divergence Times. *Systematic Biology*, **58**, 367-380. <https://doi.org/10.1093/sysbio/syp035>
- [7] Prum, R.O., Berv, J.S., Dornburg, A., *et al.* (2015) A Comprehensive Phylogeny of Birds (Aves) Using Targeted Next-Generation DNA Sequencing. *Nature*, **526**, 569-573. <https://doi.org/10.1038/nature15697>

- [8] Suh, A., Paus, M., Kiefmann, M., *et al.* (2011) Mesozoic Retroposons Reveal Parrots as the Closest Living Relatives of Passerine Birds. *Nature Communications*, **2**, Article No. 443. <https://doi.org/10.1038/ncomms1448>
- [9] Matzke, A., Churakov, G., Berkes, P., *et al.* (2012) Retroposon Insertion Patterns of Neoavian Birds: Strong Evidence for an Extensive Incomplete Lineage Sorting Era. *Molecular Biology and Evolution*, **29**, 1497-1501. <https://doi.org/10.1093/molbev/msr319>
- [10] 刘红梅, 刘国庆. 基于 k-mer 组分信息的系统发生树构建方法[J]. 生物信息学, 2013, 11(2): 100-104.
- [11] Luo, L.F. (2004) Theoretic-Physical Approach to Molecular Biology. Shanghai Scientific & Technical Publishers, Shanghai, 388-402.
- [12] Saitou, N. and Nei, M. (1987) The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology and Evolution*, **4**, 406-425.
- [13] Slack, K.E., Janke, A., Penny, D., *et al.* (2003) Two New Avian Mitochondrial Genomes (Penguin and Goose) and a Summary of Bird and Reptile Mitogenomic Features. *Gene*, **302**, 43-52.
- [14] Luo, L.F. (2015) Quantum Theory on Genome Evolution. <https://doi.org/10.1101/034710>

期刊投稿者将享受如下服务:

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [hjcb@hanspub.org](mailto:hjcb@hanspub.org)