

# DomeTree: a canonical toolkit for mitochondrial DNA analyses in domesticated animals

MIN-SHENG PENG,<sup>\*†<sup>a</sup></sup> LONG FAN,<sup>‡<sup>a</sup></sup> NI-NI SHI,<sup>\*†<sup>a</sup></sup> TIAO NING,<sup>§</sup> YONG-GANG YAO,<sup>†¶</sup> ROBERT W. MURPHY,<sup>\*\*\*</sup> WEN-ZHI WANG<sup>\*</sup> and YA-PING ZHANG<sup>\*†§</sup>

<sup>\*</sup>State Key Laboratory of Genetic Resources and Evolution, Yunnan Laboratory of Molecular Biology of Domestic Animals, and Germplasm Bank of Wild Species, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China, <sup>†</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan 650204, China, <sup>‡</sup>School of Life Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR 999077, China, <sup>§</sup>Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming, Yunnan 650091, China, <sup>¶</sup>Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China, <sup>\*\*\*</sup>Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, Ontario, M5S 2C6, Canada

## Abstract

Mitochondrial DNA (mtDNA) is widely used in various genetic studies of domesticated animals. Many applications require comprehensive knowledge about the phylogeny of mtDNA variants. Herein, we provide the most up-to-date mtDNA phylogeny (i.e. haplogroup tree or matrilineal genealogy) and a standardized hierarchical haplogroup nomenclature system for domesticated cattle, dogs, goats, horses, pigs, sheep, yaks and chickens. These high-resolution mtDNA haplogroup trees based on 1240 complete or near-complete mtDNA genome sequences are available in open resource DomeTree (<http://www.dometree.org>). In addition, we offer the software MitoToolPy (<http://www.mitotool.org/mp.html>) to facilitate the mtDNA data analyses. We will continuously and regularly update DomeTree and MitoToolPy.

**Keywords:** domesticated animal, haplogroup tree, mtDNA, phylogeny, software

Received 20 November 2014; revision received 1 February 2015; accepted 2 February 2015

## Introduction

Mitochondrial DNA (mtDNA) is a useful molecular marker for constructing matrilineal genealogies and tracing evolutionary history of domesticated animals from a matrilineal perspective (Bruford *et al.* 2003; Wang *et al.* 2014). It is also valuable for studies of ancient DNA (Ho & Gilbert 2010) as well as non-human forensic research (Imes *et al.* 2012). Since 2006, the number of complete mtDNA nucleotide sequences (i.e. mitochondrial genomes) has grown quickly due to the improved sequencing technologies and the high-quality commercial kits (Wang *et al.* 2014).

A high-resolution mtDNA phylogeny provides an essential framework for various studies and yet different approaches, such as Bayesian inference and maxi-

mum-likelihood trees, are used to infer historical relationships. The phylogeny as well as the nomenclature of mtDNA lineages fluctuates along with the accumulation of more data even for the same animal such as pig (Wu *et al.* 2007a,b). This phenomenon complicates the comparison of results from different studies especially when this involves different and updated nomenclatorial systems. Predominant in molecular anthropology, a phylogenetic network-like (Bandelt *et al.* 1999, 2000) mtDNA haplogroup tree successfully captures the known mtDNA variation (Torrioni *et al.* 2006). In this system, mtDNA sequences are aligned against a reference sequence and the variants are then scored (Bandelt *et al.* 2014). A haplogroup is a cohort mtDNA lineages derived by descent from the same ancestral mtDNA molecule as revealed by the sharing of a characteristic mutational motif (i.e. strings of variants) (Torrioni *et al.* 2006). Thus, a haplogroup represents a monophyletic clade (Wu *et al.* 2007b) and could be read from the mtDNA haplogroup tree (Torrioni *et al.* 2006). Nowadays, the global

Correspondence: Wen-Zhi Wang, Fax: +86-871-68125502; E-mail: wangwz@mail.kiz.ac.cn and Ya-Ping Zhang, Fax: +86-871-68526521; E-mail: zhangyp@mail.kiz.ac.cn

<sup>a</sup>These authors contributed equally to this work.

human mtDNA haplogroup tree is maintained as PhyloTree (<http://www.phylotree.org>) (van Oven & Kayser 2009). It facilitates the comparison of different mtDNA studies. Some bioinformatic tools are based on this *de facto* standard (e.g. Fan & Yao 2011, 2013; Vianello *et al.* 2013).

The haplogroup tree-based strategy has been employed for pigs (Wu *et al.* 2007b), cattle (Achilli *et al.* 2008, 2009; Bonfiglio *et al.* 2010, 2012), horses (Achilli *et al.* 2012), chickens (Miao *et al.* 2013), sheep (Lancioni *et al.* 2013) and dogs, goats and yaks (Shi *et al.* 2014). Nevertheless, the established phylogeny including the haplogroup nomenclature (e.g. Wu *et al.* 2007b) is neglected in subsequent researches. Some recently released mtDNA sequences have not been analysed under such context. Different haplogroup/clade/group nomenclature has been proposed (e.g. Yu *et al.* 2013; Cannon *et al.* 2015). The generated confusion and chaos complicates the comparison of different studies. Therefore, we present an integrated bioinformatic approach to provide (i) the standardized mtDNA phylogeny with haplogroup nomenclature system; and (ii) the related software for data analysis, to address issues in the mtDNA studies of domesticated animals.

## Materials and methods

### Data sets

We only considered animals (including domesticates and their wild ancestors) with at least 20 complete or near-complete mtDNA genomes deposited in GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>; Accessed on: June 1st, 2014). The data set includes a total of 1342 mtDNA sequences after scrutiny of data quality (Shi *et al.* 2014). Eight sequences (Table S1, Supporting information) fell outside the macro-clades (i.e. macro-haplogroups) containing other samples of dogs (Thalman *et al.* 2013), pigs (Wu *et al.* 2007b) and yaks (Wang *et al.* 2010), respectively. More than 400 variants (data are not shown) were scored out for the sequence EU442884 of Mongolian wolf (*Canis lupus chanco*). Thus, a total of nine sequences (Table S1) with remote relationships to others as suggested before (Shi *et al.* 2014) were removed in subsequent analyses. For 194 sequences with potential errors (Shi *et al.* 2014), we (i) rescued 101 sequences (Table S2, Supporting information) by disregarding error-prone sites in next-generation sequencing (Table S3, Supporting information); and (ii) yet excluded the remaining 89 flawed sequences (Table S4, Supporting information). Ultimately, we used 1240 complete or near-complete mtDNA genome sequences.

### mtDNA haplogroup trees

Sequences were aligned with SeqMan Pro of DNASTAR LASERGEN 7.1.0 (DNASTar Inc., Madison, WI) against reference sequences (Table 1). The mtDNA sequence alignments were deposited in Dryad (doi:10.5061/dryad.cc5kn). Sequences with different start positions were reassembled in terms of the corresponding reference sequences and then were realigned. Mutations were scored relative to the related reference sequences (Table 1). Difficult-to-align regions and error-prone sites in next-generation sequencing were excluded from further consideration. We followed the rules proposed by Bandelt and Parson (Bandelt & Parson 2008) to score length variants. We employed the matching or near-matching strategy (Yao *et al.* 2004) to screen haplogroup-specific mutations defined in available mtDNA haplogroup trees for each kind of domesticated animals (Wu *et al.* 2007b; Achilli *et al.* 2008, 2009, 2012; Bonfiglio *et al.* 2010, 2012; Lancioni *et al.* 2013; Miao *et al.* 2013; Shi *et al.* 2014) in the scored variants of every sequence. And then we could allocate each of sequences into specific haplogroup. As a result, the phylogenetic status for each of sequences could be determined so that the trees could be updated by incorporating all of sequences. The trees were checked by NETWORK 4.611 (<http://www.fluxus-engineering.com/sharenet.htm>) (Bandelt *et al.* 1999). The mtDNA haplogroup trees for domestic animals were transcribed into the extensible stylesheet language format (xls) as adopted by rCRS-oriented version of PhyloTree ([http://phylotree.org/rCRS-oriented\\_version.htm](http://phylotree.org/rCRS-oriented_version.htm)) (van Oven & Kayser 2009) and were then deposited in DOMETREE (<http://www.demotree.org>). The hierarchical haplogroup nomenclature systems were updated from previous studies (Table 1) while referring to the rules used in the phylogeny of human mtDNA (van Oven & Kayser 2009) in order to avoid chaos and confusion.

### Application software

We developed MITOTOOLPY (<http://www.mitotool.org/mp.html>) written in Python language to handle data analyses. It was based on BIOPYTHON (<http://www.biopython.org/>) (Cock *et al.* 2009) and licensed under GPL v3. The pairwise sequence alignment was based on CLUSTALW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) (Larkin *et al.* 2007), as the results were similar to commercial software SEQMAN PRO of which codes were not available. The scoring system for mtDNA haplogroup determination was employed from our previously presented MITOTOOL (Fan & Yao 2013). And the software fixed the issues of haplogrouping (Bandelt *et al.* 2012). MITOTOOLPY has been and will be updated synchronously with DOMETREE.

**Table 1** Summary of (near-)complete mtDNA sequences analysed in this work

Animals	Related species	No. of sequences	Reference sequence	Haplogroup nomenclature
Cattle and Aurochs	<i>Bos taurus</i>	258	V00654	I, P, Q, R, and T
	<i>B. indicus</i>	9		
	<i>B. primigenius</i>	2		
Yak and wild yak	<i>B. grunniens</i>	70	GQ464259	A–D
Dog and grey wolf	<i>Canis lupus</i>	424	EU789787	A–F
Horse and Przewalski's horse	<i>Equus caballus</i>	230	JN398377	A–R
	<i>E. przewalskii</i>	6		
Chicken and Red Junglefowl	<i>Gallus gallus</i>	65	AP003321	A–I and X–Z
Pig and wild boar	<i>Sus scrofa</i>	100	EF545567	A, D, and E
Sheep	<i>Ovis aries</i>	45	AF010406	A–E
Goat	<i>Capra hircus</i>	31	GU068049	A–C

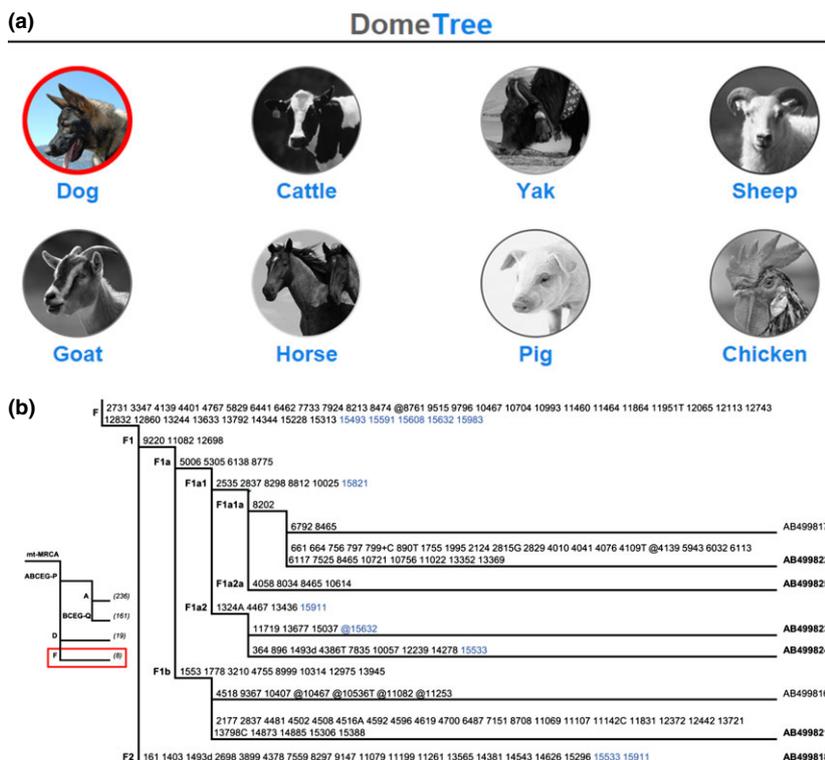
**Results**

*The updated mtDNA haplogroup trees*

By analysing 1240 complete or near-complete published mtDNA genomes (Table 1), we updated the mtDNA haplogroup trees of cattle, dogs, goats, horses, pigs, sheep, yaks and chickens and deposited them in DomeTree (Fig. 1a). Monophyly of each major clades/haplogroups described in previous studies (Table 1) was confirmed. The diagnostic motif (in both control- and/or coding-region) for each (sub)haplogroup was characterized and mapped on each branch. All trees were made available

for viewing with common web browsers such as Internet Explorer and Mozilla Firefox. It was convenient to use the 'Find' command of the browsers to search for specific haplogroups or mutations.

The incorporation of recently released sequences into our analyses defined some new (sub)haplogroups. In one case, sequences AB499818, AB499821–AB499825 of the extinct Japanese Honshu wolf (*Canis lupus hodophilax*) clustered with Kishu (AB499816) and Husky (AB499817) in haplogroup F (Fig. 1b; <http://www.dometree.org/trees/dog.htm>), which previously was only defined by control region information (Pang *et al.* 2009). In another case, sequence EF375877 of the Lanyu pig from Lanyu



**Fig. 1** Snapshots of DOMETREE (a) and haplogroup F branch from the mtDNA haplogroup tree of dog (b).

Islet (Wu *et al.* 2007a) rooted at the base of the branch for haplogroup A1, and, accordingly, was used to define new subhaplogroup A1c (<http://www.dometree.org/trees/pig.htm>).

#### Phylogenetic analyses of mtDNA data

In terms of mtDNA haplogroup trees, we developed MitoToolPy for data analyses, which included the following functions: (i) score variants relative to a reference sequence; (ii) classify haplogroups (haplogrouping); and (iii) check for potential errors due to missing diagnostic mutations as well as excessive private mutations. We applied MITOTOOLPY to 34 novel sequences released by GenBank (Table S5, Supporting information) and further confirmed the results manually. This trial found MITOTOOLPY to be practical and efficient. Moreover, MITOTOOLPY handled mtDNA fragments (e.g. control region) and served as a convenient 'barcoding tool' for some strains or breeds.

#### Discussion

The updated mtDNA phylogeny including its haplogroup nomenclature can serve as a starting point for future mtDNA analyses for cattle, dogs, horses, pigs, yaks and chickens. This is necessary for avoiding conflicts and confusion when comparing different mtDNA studies, especially now when next-generation sequencing accelerates the accumulation of mitochondrial genome data (Lippold *et al.* 2011; Horsburgh *et al.* 2013). The highly resolved, updated phylogenies often highlight the multiple matrilineal backgrounds of domesticated animals. For example, the mtDNA phylogeny of dogs and grey wolves indicates that haplogroup F (Fig. 1b) is a genetic relic of extinct Japanese Honshu wolves or their ancestors in modern Kishus and Huskies. Similarly, the haplogroup A1c in the Lanyu pig probably traces back to local wild boar. Similar scenarios exist for haplogroups P, Q and R in cattle (Achilli *et al.* 2008, 2009; Bonfiglio *et al.* 2010). Sporadic introgression from wild ancestors or local domestication events best explained these patterns (Wang *et al.* 2014). Greater effort, such as studies of ancient DNA, can shed light on such observations (Larson & Burger 2013). In addition, our results support that there are no signals of recombination neither in the phylogeny explored nor on maternal inheritance (Bandelt *et al.* 2005).

We will simultaneously update both DOMETREE and MITOTOOLPY at least every year. Updates will reconcile issues of phylogeny and haplogroup nomenclature due to novel sequence data. Further, we plan to incorporate additional domesticated animals into this system as their data become accessible. We hope that DOMETREE and

MITOTOOLPY will facilitate and foster future work on the mtDNA of domesticated animals.

#### Acknowledgements

This work was supported by the 973 program (2013CB835200 and 2013CB835204). M-S.P. thanks the support from the Youth Innovation Promotion Association, Chinese Academy of Sciences. R.W.M. thanks the support from the Visiting Professorship for Senior International Scientists from the Chinese Academy of Sciences.

#### References

- Achilli A, Olivieri A, Pellecchia M *et al.* (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology*, **18**, R157–R158.
- Achilli A, Bonfiglio S, Olivieri A *et al.* (2009) The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS ONE*, **4**, e5753.
- Achilli A, Olivieri A, Soares P *et al.* (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences, USA*, **109**, 2449–2454.
- Bandelt H-J, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *International Journal of Legal Medicine*, **122**, 11–21.
- Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.
- Bandelt H-J, Macaulay V, Richards M (2000) Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA. *Molecular Phylogenetics and Evolution*, **16**, 8–28.
- Bandelt H-J, Kong Q-P, Parson W, Salas A (2005) More evidence for non-maternal inheritance of mitochondrial DNA? *Journal of Medical Genetics*, **42**, 957–960.
- Bandelt H-J, van Oven M, Salas A (2012) Haplogrouping mitochondrial DNA sequences in Legal Medicine/Forensic Genetics. *International Journal of Legal Medicine*, **126**, 901–916.
- Bandelt H-J, Kloss-Brandstätter A, Richards M-B, Yao Y-G, Logan I (2014) The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies. *Journal of Human Genetics*, **59**, 66–77.
- Bonfiglio S, Achilli A, Olivieri A *et al.* (2010) The enigmatic origin of bovine mtDNA haplogroup R: sporadic interbreeding or an independent event of *Bos primigenius* domestication in Italy? *PLoS ONE*, **5**, e15760.
- Bonfiglio S, Ginja C, De Gaetano A *et al.* (2012) Origin and spread of *Bos taurus*: new clues from mitochondrial genomes belonging to haplogroup T1. *PLoS ONE*, **7**, e38601.
- Bruford M-W, Bradley D-G, Luikart G (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, **4**, 900–910.
- Cannon MV, Brandebourg TD, Kohn MC *et al.* (2015) Mitochondrial DNA sequence and phylogenetic evaluation of geographically disparate *Sus scrofa* breeds. *Animal Biotechnology*, **26**, 17–28.
- Cock P-J, Antao T, Chang J-T *et al.* (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422–1423.
- Fan L, Yao Y-G (2011) MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion*, **11**, 351–356.

- Fan L, Yao Y-G (2013) An update to MitoTool: using a new scoring system for faster mtDNA haplogroup determination. *Mitochondrion*, **13**, 360–363.
- Ho S-Y, Gilbert M-T (2010) Ancient mitogenomics. *Mitochondrion*, **10**, 1–11.
- Horsburgh K-A, Prost S, Gosling A *et al.* (2013) The genetic diversity of the Nguni breed of African cattle (*Bos spp.*): complete mitochondrial genomes of haplogroup T1. *PLoS ONE*, **8**, e71956.
- Imes D-L, Wictum E-J, Allard M-W, Sacks B-N (2012) Identification of single nucleotide polymorphisms within the mtDNA genome of the domestic dog to discriminate individuals with common HVI haplotypes. *Forensic Science International. Genetics*, **6**, 630–639.
- Lancioni H, Di Lorenzo P, Ceccobelli S *et al.* (2013) Phylogenetic relationships of three Italian merino-derived sheep breeds evaluated through a complete mitogenome analysis. *PLoS ONE*, **8**, e73712.
- Larkin M-A, Blackshields G, Brown N-P *et al.* (2007) CLUSTAL W and CLUSTAL X version 2.0. *Bioinformatics*, **23**, 2947–2948.
- Larson G, Burger J (2013) A population genetics view of animal domestication. *Trends in Genetics*, **29**, 197–205.
- Lippold S, Matzke N-J, Reissmann M, Hofreiter M (2011) Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evolutionary Biology*, **11**, 328.
- Miao Y-W, Peng M-S, Wu G-S *et al.* (2013) Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity*, **110**, 277–282.
- Pang J-F, Kluetsch C, Zou X-J *et al.* (2009) mtDNA data indicate a single origin for dogs south of Yangtze River, less than 16,300 years ago, from numerous wolves. *Molecular Biology and Evolution*, **26**, 2849–2864.
- Shi N-N, Fan L, Yao Y-G, Peng M-S, Zhang Y-P (2014) Mitochondrial genomes of domestic animals need scrutiny. *Molecular Ecology*, **23**, 5393–5397.
- Thalmann O, Shapiro B, Cui P *et al.* (2013) Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, **342**, 871–874.
- Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends in Genetics*, **22**, 339–345.
- van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, **30**, E386–E394.
- Vianello D, Sevini F, Castellani G *et al.* (2013) HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Human Mutation*, **34**, 1189–1194.
- Wang Z, Shen X, Liu B *et al.* (2010) Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *Journal of Biogeography*, **37**, 2332–2344.
- Wang G-D, Xie H-B, Peng M-S, Irwin D, Zhang Y-P (2014) Domestication genomics: evidence from animals. *Annual Review of Animal Biosciences*, **2**, 65–84.
- Wu C-Y, Jiang Y-N, Chu H-P *et al.* (2007a) The type I Lanyu pig has a maternal genetic lineage distinct from Asian and European pigs. *Animal Genetics*, **38**, 499–505.
- Wu G-S, Yao Y-G, Qu K-X *et al.* (2007b) Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biology*, **8**, R245.
- Yao Y-G, Kong Q-P, Wang C-Y, Zhu C-L, Zhang Y-P (2004) Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. *Molecular Biology and Evolution*, **21**, 2265–2280.
- Yu G, Xiang H, Wang J, Zhao X (2013) The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences. *Journal of Animal Science and Biotechnology*, **4**, 9.

---

M.-S.P., W.-Z.W. and Y.-P.Z. designed research. M.-S.P., L.F., N.-N.S. and T.N. performed the analysis. W.-Z.W., N.-N.S. and M.-S.P. constructed the website. L.F., N.-N.S. and M.-S.P. coded the software. M.-S.P., Y.-G.Y., R.W.M. and Y.-P.Z. wrote the manuscript.

---

## Data Accessibility

The DomeTree is freely available on the web at <http://www.dometree.org>. The MitoToolPy and user manual are accessible on <http://www.mitotool.org/mp.html>.

The mtDNA sequence alignments were deposited in Dryad (doi:10.5061/dryad.cc5kn).

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1** Sequences showing remote relationships with the others.

**Table S2** Rescued sequences generated by next-generation sequencing platforms.

**Table S3** Disregarded error-prone sites in next-generation sequencing.

**Table S4** Flawed sequences data excluded in the analyses.

**Table S5** Analyses of 34 sequences released recently.