

Phylogeny and Evolution of Multiple Common Carp (*Cyprinus carpio* L.) Populations Clarified by Phylogenetic Analysis Based on Complete Mitochondrial Genomes

Chuanju Dong^{1,2} · Jian Xu¹ · Baosen Wang¹ · Jianxin Feng³ · Zsigmond Jeney⁴ · Xiaowen Sun¹ · Peng Xu¹

Received: 4 January 2015 / Accepted: 6 May 2015 / Published online: 9 June 2015
© Springer Science+Business Media New York 2015

Abstract Common carp (*Cyprinus carpio* L.) is one of the oldest, most widely farmed commercially important freshwater fish in the world. However, many undetermined phylogenetic relationships and origins of common carp lineages remain, which are obstacles to conservation and genetic breeding of this species. Phylogenetic analyses based on molecular tools are helpful to distinguish the origin of species, understand and clarify their evolutionary history, and provide a genetic basis for selective breeding. In this study, we demonstrated a method to extract complete mitochondrial genome sequences from whole-genome resequencing data using the Illumina platform. The complete mitochondrial genome sequences of 26 individuals representing nine strains were obtained and subjected to a phylogenetic analysis. We reconstructed the phylogenetic topologies of the nine strains and analyzed the haplotypes. Results from both analyses suggested that the genome sequences

belonged to two distinct subspecies from Europe and East Asia. We also estimated the time of divergence of the nine strains, which was up to 100 KYA. The phylogenetic results clarified the breeding history of Songpu mirror carp and suggest that this species may be hybrid of paternal European mirror carp and maternal Xingguo red carp. The results also support a previous hypothesis that koi may have originated from or have close ancestry with Oujiang color carp in China.

Keywords Common carp · Mitochondrial genome · Cyt *b* · D-loop · Phylogenetics

Introduction

Common carp (*Cyprinus carpio* L.) is a widespread freshwater fish that originated in Eurasia. Common carp have a cultural history and domestication of several thousand years and have become one of the most important edible and popular aquarium fish, comprising a large number of strains worldwide (Zhou et al. 2003b).

Because of long-term geographic isolation, selection, and breeding, common carp have developed distinct body shape, skin color, scale pattern, body size, growth rate, and stress tolerance phenotypes to adapt to various environments, providing abundant germplasm for further genetic selection. China has the greatest number of common carp populations and domesticated strains in the world. The renowned strains in China include Hebao red carp (HB), Xingguo red carp (XC), Oujiang color carp (OJ), and Songpu mirror carp (SP). Intensive hybrid breeding of common carp was conducted over the last half century in China, which generated many hybrid strains. Jian carp (JC) is one of the most successful hybrid strains and has contributed significantly to the rapid growth of common carp. However, aquaculture of the hybrid strains has caused

Electronic supplementary material The online version of this article (doi:10.1007/s10126-015-9639-7) contains supplementary material, which is available to authorized users.

✉ Xiaowen Sun
sunxw@cafs.ac.cn

✉ Peng Xu
xupeng77@gmail.com

¹ CAFS Key Laboratory of Aquatic Genomics & Beijing Key Laboratory of Fishery Biotechnology, Centre for Applied Aquatic Genomics, Chinese Academy of Fishery Sciences, 100141 Beijing, China

² College of Life Science and Technology, Shanghai Ocean University, Shanghai 201306, China

³ Henan Academy of Fishery Sciences, Zhengzhou 450044, China

⁴ National Agricultural Research and Innovation Centre (NARIC), Research Institute for Fisheries and Aquaculture (HAKI), Szarvas 5540, Hungary

significant adverse impacts. Escape, artificial propagation, and release of the hybrids into natural waters have changed the genetic profiles of many wild populations and have caused serious management, conservation, and breeding consequences. Phylogenetic analysis is a widely used approach to protect and conserve germplasm and is used to clarify the phylogenetic relationships and reveal the obscured breeding histories of populations and strains. Phylogenetic studies benefit genetic breeding and germplasm conservation efforts.

Many studies have been performed to illustrate the origin and phylogenetic relationships among common carp strains, employing various genetic tools, including microsatellites, random amplified polymorphic DNA, amplified fragment length polymorphism, restriction fragment length polymorphism, and mitochondrial DNA (mtDNA) sequence analyses (Bartfai et al. 2003; Cheng et al. 2010; Mabuchi et al. 2008; Zhou et al. 2003a). Some studies have provided insight into the phylogenetic relationships among specific strains and wild populations. All of these studies indicate that all common carp populations belong to the European and East Asian subspecies. Genetic differentiation into the two subspecies of *Cyprinus carpio carpio* in Europe and *Cyprinus carpio haematopterus* in East Asia is clearly evident, and the existence of a third subspecies (*C. carpio rubrofusca*) that may have diverged from *C. carpio haematopterus* in China has been suggested (Kohlmann and Kersten 2013; Zhou et al. 2004a). However, many phylogenetic relationships and the breeding history remain uncertain in common carp. For example, the origin and phylogeny of koi (KC), the most popular cultured ornamental carp in the world, are long-term controversies. Balon (1995) believed that KC originated from pond carp in Japan, China, or Germany. Froufe et al. (2002) suggested that KC originated from European common carp. A phylogenetic analysis sequences conducted by Wang and Li (2004b) using mtDNA found that KC may be derived from OJ inhabiting Zhejiang province, China. The relationships among the three major domestic strains in South China, such as XC, HB, and OJ, are also controversial (Wang and Li 2004a). Many hybrid common carp strains occur in China. Multiple rounds of hybridization and genetic introgression are employed during hybrid breeding. These hybrid strains have ambiguous genetic backgrounds when breeding history is not well recorded. These phylogenetic uncertainties require the development of comprehensive genetic tools and thorough sampling to resolve the obscure origins and relationships among common carp. A substantial investment has been made in common carp genetics and genomics, which has greatly expanded the available genetic resources and tools. These resources include a large number of genetic markers (Ji et al. 2012; Kongchum et al. 2010; Zhang et al. 2008), genetic and physical maps (Cheng et al. 2010; Xu et al. 2011c; Zhang et al. 2013), expressed sequence tags and RNA-seq (Moens et al. 2007; Williams et al. 2008; Xu et al. 2011a), microRNA (Yan

et al. 2012), a bacterial artificial chromosome (BAC) library (Li et al. 2011), and BAC-end sequences (Xu et al. 2011b). The complete common carp genome has been recently published and released (Xu et al. 2014). With these molecular tools and resources, phylogenetic analyses will provide a better understanding of common carp phylogeny and comprehensively unveil obscure relationships among strains.

MtDNA is the most popular genetic tool for phylogenetic studies. However, only partial mtDNA sequences were used in most studies because of budget and technology concerns. The cytochrome *c* oxidase subunit II, cytochrome *b* (Cyt *b*), and the displacement loop (D-loop) regions are the most commonly used sequences for phylogenetic analyses (Kohlmann and Kersten 2013; Zhou et al. 2004b). However, different genes and regions in the mitogenome have various evolutionary rates, which may generate significant differences during phylogenetic analyses. Many scientists have suggested analyzing multiple genes or gene clusters rather than a single gene or genome region in the mitogenome, as this approach increases the possibility of revealing the actual evolutionary phylogeny of a particular species (Cunningham 1997; Wiens 1998b). Higher resolution and accuracy of phylogenetic relationships are gained when more mtDNA sequences are used (Wiens 1998a). Complete mitochondrial genome (CMG) sequences are desired for an accurate phylogenetic study. For example, the CMGs of teleost species have been extensively used to reinvestigate the main patterns of higher teleostean phylogenies and have resolved persistent controversies over higher-level teleost relationships. The CMGs from ten phyllostomid bat species were sequenced using the Illumina platform and were used to investigate the phylogenies of leaf-nosed bats and resolve their evolutionary history (Botero-Castro et al. 2013). Ten new CMGs from pulmonates were sequenced and used to evaluate their relationship through phylogenetic analyses (White et al. 2011).

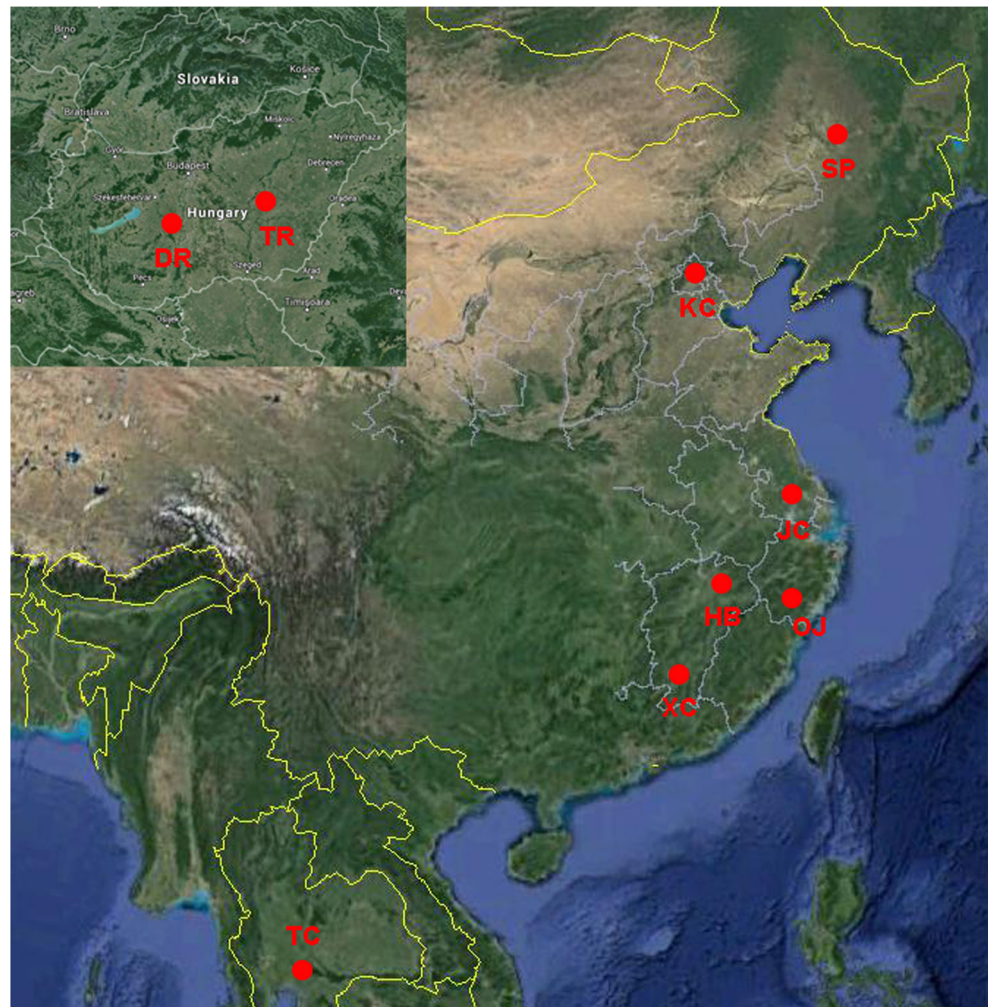
Herein, we utilized the high-throughput whole-genome resequencing data of 26 common carp individuals from nine strains and extracted the mitochondrial genome sequences using a bioinformatics approach. The CMGs were collected, assembled, and subjected to a phylogenetic analysis of these individuals. Utilizing CMG sequences will provide additional resolution to decipher difficult phylogenetic relationships and benefit germplasm conservation and genetic breeding efforts for common carp.

Materials and Methods

Sample Collection

Nine strains of common carp including 26 individuals were collected across Europe and China (Fig. 1), representing typical lineages in both geographic regions. Danube River carp

Fig. 1 Sample collection sites and their geographic distribution



(DR, $n=3$), Tisza River carp (TR, $n=3$), and Thailand carp (TC, $n=2$) were collected from the live carp gene bank of the Research Institute for Fisheries and Aquaculture, Hungary. Representatives of all other strains and wild populations were collected from China, including SP ($n=4$) from Heilongjiang Fishery Research Institute of the Chinese Academy of Fishery Sciences (CAFS), HB ($n=4$) from Wuyuan county in Jiangxi province, XC ($n=3$) from Xingguo county in Jiangxi province, OJ ($n=3$) from Longquan county in Zhejiang province, KC ($n=2$) from the Beijing Fishery Research Institute, and JC ($n=2$) from the CAFS Freshwater Fisheries Research Center.

Genome Sequencing, Extraction of the Mitochondrial Sequences, and Assembly

Blood samples or fin clips were collected from each individual. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany). Paired-end libraries were constructed following the standard Illumina protocol with an insert size of 180 bp and sequenced in one flow cell (eight lanes) on the Illumina HiSeq 2000 platform. Low-

quality and short reads were removed using SolexaQA software to obtain a set of usable reads (Cox et al. 2010). The data were used primarily to identify genome-wide single-nucleotide polymorphisms (SNPs) and to conduct the genetic diversity analysis (Xu et al. 2014). Herein, we used the data to extract the mitochondrial sequences. Paired-end reads from each individual were aligned to the mitochondrial reference genome (NC_018366) (Wang et al. 2013) using the Burrows–Wheeler Aligner (Li and Durbin 2009). All mapped reads were collected and assembled into the mitochondrial genome of each individual using Newbler (Roche Scientific, Mannheim, Germany). The assembled CMG sequences were validated by comparison to references using CLUSTALW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

Sequence Analysis

Alignment of the D-loop, Cyt *b*, and CMG sequences of the 26 individuals was performed using MEGA 6 (Tamura et al. 2013), and the sequences were subjected to base composition, conservative site, variable site, parsimony informative site,

and transition/transversion ratio analyses. DNAsp v5 (Librado and Rozas 2009) was used to calculate haplotype and DNA polymorphisms. We used Arlequin v3.5 (Excoffier and Lischer 2010) to determine the connection length between the different haplotypes and laid out the network for optimal visualization with the help of NETWORK 4.6.1.2 (<http://www.fluxus-engineering.com>) to directly show the associations between the haplotypes.

Divergence Time and Phylogenetic Analysis

Pairwise genetic distances were calculated in MEGA 6 based solely on the CMG sequences using the maximum composite likelihood model. Transitions and transversions were included as nucleotide substitutions. Then, Nei's formula was used with the Cyprinidae evolutionary rate to calculate divergence time. The phylogenetic relationships were estimated using the D-loop (927 bp), Cyt *b* (1141 bp), and CMG (16,581 bp) sequences and the maximum likelihood (ML) and neighbor-joining (NJ) methods, respectively. Bayesian information criterion values were calculated for each model to determine the optimum for the ML method in this study. The Tamura–Nei and maximum composite likelihood models were used in the ML and NJ analyses, respectively. A total of 1000 bootstrap replicates were conducted for each calculation. The PhyML software package was used in the phylogenetic analysis as a comparable approach.

Results

Mitochondrial Genomes and Their Features

A total of 332.3-Gb whole-genome sequences were collected from the Illumina sequencing platform, which contained reads derived from the mitochondrial and nuclear genomes of 26 individuals. The raw reads were mapped to the common carp mitochondrial reference genome to discriminate mitochondrial reads from background. A total of 2,065,531 mapped reads were collected, representing 206.5 Mb of the mitochondrial genomes from the 26 individuals. Sequencing depths were 193–1054, with a mean of $\sim 479\times$, which was comparable with the reference mitochondrial genome (Table 1). The mitochondrial genomes of the 26 individuals were assembled completely for further analysis.

All assembled CMGs were 16,581 bp in length, with GC content of 43.2–43.4 % (Table 2), which was similar to the reference mitochondrial genome. These CMGs contained 13 protein-coding genes, 22 tRNA genes, two rRNA genes, and a D-loop region. The structural organization and location of the different features in the genome conformed to the common vertebrate mitochondrial genome model.

The lengths of the D-loop, Cyt *b*, and the CMG were 927; 1141; and 16,581 bp in all 26 individuals, respectively, and contained conserved sites of 896; 1118; and 16,347 bp; variable sites of 31, 23, and 234 bp; parsimony informative sites of 31, 22, and 227 bp; transitions of 9, 7, and 69 bp; and transversions of 1, 1, and 4 bp (Table 3).

Haplotype Analysis

We identified 12 haplotypes from the 26 individuals (Table 4 and Fig. 2). The Asian and European individuals were divided into two clades, containing nine and three haplotypes, respectively. Six strains (DR, JC, KC, XC, OJ, and TC) contained only one haplotype, two strains (HB and TR) contained two haplotypes, and one strain (SP) contained three haplotypes. All haplotypes belonged exclusively to one specific strain, except haplotype H7, which was shared by the SP and XC strains. SP also contained the H6 and H8 haplotypes, which were relatively new and likely derived from H7. JC is a hybrid strain with HB as the maternal parent during multiple hybridization rounds. However, haplotype H4 belonged exclusively to JC, suggesting that either H4 is a newly evolved haplotype or our samples did not represent all HB haplotypes. Investigating additional HB and JC samples would clarify this issue.

Phylogenetic Analysis

A phylogenetic analysis was performed based on the CMGs from the 26 individuals using the ML method. The ML tree was constructed as shown in Fig. 3. We performed a NJ analysis and constructed the NJ tree in MEGA6 (Fig. 4). We also constructed a ML tree in PhyML to evaluate the reliability of the phylogenetic analysis using the ML method. The NJ tree supported the ML tree topology, and only minor differences in the bootstrap support (BS) values were detected, providing confidence for the common carp phylogenetic topology results. The PhyML topology was consistent with that from MEGA6 (Supplementary Fig.S1).

Two distinctive clades corresponding to two subspecies were found. Clade A belonged to strains inhabiting East Asia, including SP, XC, JC, HB, KC, TC, and OJ, and was clearly divided into the A1 and A2 subclades. A1 consisted of SP, XC, JC, and HB, whereas A2 consisted of KC, TC, and OJ. Clade B belonged to the subspecies *C. carpio carpio* inhabiting Europe, including DR and TR. All strains and populations in clade A belonged to the East Asian subspecies *C. carpio haematopterus*, except SP, which originated from a European mirror carp (*C. carpio carpio*) strain. SP was introduced into China in the 1970s and was bred to adapt to the local aquaculture environment. SP showed high similarity with XC in subclade A1. Further discussion will follow below regarding the exception.

Table 1 Total paired-end reads, data size, and sequencing coverage of mitochondrial genomes in each individual

Strains	No. of left fragments	No. of right fragments	Data size (bp)	Sequencing coverage
DR1	31,848	8588	4,043,600	244
DR2	34,954	34,334	6,928,800	418
DR3	34,954	34,334	6,928,800	418
TR1	44,296	44,072	8,836,800	533
TR2	26,486	26,486	5,297,200	319
TR3	36,916	36,626	7,354,200	444
JC1	17,314	17,160	3,447,400	208
JC2	16,070	15,966	3,203,600	193
KC1	16,060	15,988	3,204,800	193
KC2	44,272	43,448	8,772,000	529
SP1	25,158	35,080	6,023,800	363
SP2	31,296	25,066	5,636,200	340
SP3	51,802	51,620	10,342,200	624
SP4	82,880	82,508	16,538,800	997
OJ1	80,508	80,008	16,051,600	968
OJ2	80,852	72,776	15,362,800	927
OJ3	87,960	86,830	17,479,000	1054
TC1	44,296	44,072	8,836,800	533
TC2	36,916	36,626	7,354,200	444
HB1	46,389	47,869	9,425,800	568
HB2	36,789	34,156	7,094,500	428
HB3	35,681	29,623	6,530,400	394
HB4	46,258	48,452	9,471,000	571
XC1	26,052	25,796	5,184,800	313
XC2	17,894	17,684	3,557,800	215
XC3	18,288	18,174	3,646,200	220

We further constructed the ML and NJ trees based on the partial mitochondrial genomes, including *Cyt b* (Figs. 5 and 6) and the D-loop (Figs. 7 and 8), which are commonly used in phylogenetic analyses. The topologies were also consistent in the major portion of the trees based on the CMG. However, we detected some differences between the CMG and partial

mitochondrial genome sequence trees. For example, the ML and NJ trees based on *Cyt b* hardly discriminated HB from SP and XC, whereas HB was clearly separated from SP and XC and merged with JC in the CMG trees. The ML and NJ trees based on the D-loop region and CMG were significantly different for the OJ, KC, and TC topology. OJ and TC were

Table 2 Nucleotide composition of D-loop, *Cyt b*, and CMG

Strains	D-loop (%)				<i>Cyt b</i> (%)				CMG (%)			
	T	C	A	G	T	C	A	G	T	C	A	G
SP	33.0	19.7	33.4	13.8	26.2	30.0	29.8	14.0	24.8	27.5	31.9	15.7
OJ	32.8	19.8	33.0	14.3	26.1	30.1	30.1	13.7	24.8	27.5	31.9	15.8
TC	33.0	19.6	33.1	14.2	26.1	30.1	30.1	13.7	24.8	27.5	31.9	15.7
HB	33.0	19.7	33.2	14.0	26.1	30.0	29.8	14.1	24.8	27.6	31.9	15.8
XC	32.9	19.7	33.5	13.8	26.2	30.0	29.8	14.0	24.8	27.5	31.9	15.7
DR	33.1	19.6	33.1	14.1	26.3	29.9	30.1	13.7	24.9	27.5	31.9	15.7
TR	33.1	19.6	33.1	14.1	26.3	29.9	30.1	13.7	24.9	27.5	31.9	15.7
JC	32.9	19.7	33.2	14.1	26.2	30.0	29.9	13.9	24.8	27.5	31.9	15.8
KC	33.0	19.6	33.1	14.2	26.0	30.2	30.1	13.7	24.8	27.5	31.9	15.7
Avg.	33.0	19.7	33.2	14.1	26.2	30.0	30.0	13.8	24.8	27.5	31.9	15.7

Table 3 Statistics of the D-loop, Cyt *b*, and CMG in 26 individuals

	bp	vs	pi	cs	ti/tv
D-loop	927	31	31	896	9
Cyt <i>b</i>	1141	23	22	1118	7
CMG	16,581	234	227	16,347	17.3

bp number of base pairs, *vs* number of variable sites, *pi* number of parsimony informative sites, *cs* number of conserved sites, *ti/tv* the transition-transversion ratio given for in-groups only

clustered first in the CMG tree and then clustered with KC. However, OJ and KC were clustered first in the D-loop tree and then with TC.

Time of Divergence Estimate

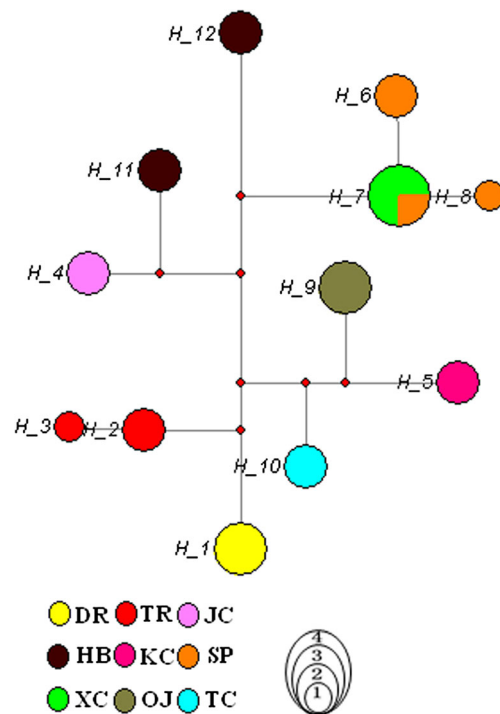
We calculated Nei's genetic distance (*D*) of the nine strains based on the CMG sequences (Table 5). The *D* values ranged from 0.00006 to 0.00699, with a mean of 0.0048. XC and SP had the smallest *D* value (0.00006), whereas OJ and DR had the largest *D* value (0.00699). The *D* values of SP and the strains belonging to the East Asian subspecies *C. carpio haematopterus* were much smaller than those between SP and the strains of the European subspecies *C. carpio carpio*. These results conflict with the breeding history of SP, which was bred based on mirror carp introduced from Europe.

Mitochondrial genomes usually have a faster evolutionary rate than that of nuclear genomes. Generally, a rate of 2 % per million years has been calculated for Cyt *b* genes in several bony fish species (Wu et al. 2013). A rate of 3.3 % per million years has been calculated for Cyt *b* and 16S rRNA genes to estimate the divergence time of cyprinids. Different genes and regions in the mitochondrial genome may have different

Table 4 Haplotypes in nine common carp stains and their frequencies

Hap.	Stains	Freq.	SD
1	DR	0.115385	0.063897
2	TR	0.076923	0.053294
3	TR	0.038462	0.038462
4	JC	0.076923	0.053294
5	KC	0.076923	0.053294
6	SP	0.076923	0.053294
7	XC	0.153846	0.07216
8	SP	0.038462	0.038462
9	SP	0.115385	0.063897
10	TC	0.076923	0.053294
11	HB	0.076923	0.053294
12	HB	0.076923	0.053294

Hap. number of haplotypes, *Freq.* haplotype frequency, *SD* haplotype standard deviations

**Fig. 2** Haplotype network of the nine common carp strains studied here. Each strain is represented by a different colored circle, and the area of the circle represents the number of individual samples ($n=1-4$) sequenced in that strain. The 12 identified haplotypes are labeled H_1-H_{12}

evolutionary rates; therefore, we used the evolutionary rate of 3.3×10^{-8} to estimate the time of divergence of the common carp lineages (Tong and Wu 2001). The divergence time between two different lineages was calculated using Nei's formula of $T=D/2\alpha$ (where *T* is the number of years since divergence and α is the evolutionary rate). The results showed that the time of divergence was $0.91 \times 10^3-1.06 \times 10^5$ years ago, with a mean of 7.27×10^4 years ago. OJ and DR showed the longest divergence time, suggesting that the European subspecies (*C. carpio carpio*) (clade B) and the Asian subspecies (*C. carpio haematopterus*) (clade A) diverged about 100 KYA. The average *D* value between clade A1 and A2 was 0.0058, suggesting a divergence time of about 8.79×10^4 years ago for the two Asian subspecies.

Discussion

As low-cost and high-throughput, next-generation sequencing technologies have been employed for genome sequencing in the past several years, CMG sequences can be easily extracted from whole-genome sequencing data based on sequence similarities and assembled into the CMG of the target species. The CMG can then be sequenced on a next-generation sequencing platform with discriminating barcodes to separate different samples. A number of phylogenetic and evolutionary studies have reported CMGs sequenced using next-generation

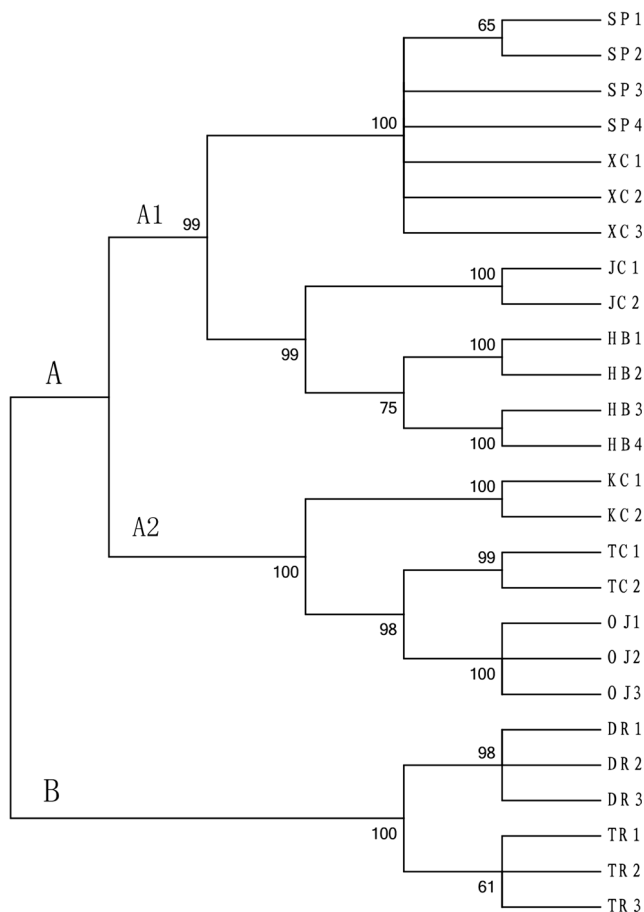


Fig. 3 Maximum likelihood-based phylogenetic tree of the full-length mitochondrial DNA sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (*A1* and *A2*) of clade A

sequencing platforms for various applications (Davila et al. 2011; Fujii et al. 2010; Rodriguez-Moreno et al. 2011). We extracted the CMG sequences from genome resequenced data of 26 common carp individuals from nine lineages based on the Illumina platform, and a phylogenetic analysis was performed based on the CMG sequences in cyprinid species for the first time. This approach will provide a rapid and effective alternative method to collect CMG sequences and conduct phylogenetic analyses on other aquaculture species. However, the Illumina platform usually has ultra-high throughput and sequencing capacity compared with the size of the mitochondrial genome. A large number of barcoding and sequencing libraries are required to reduce the cost of each mitochondrial genome and increase efficiency. However, it is not feasible when only a small number of mitochondrial genomes are sequenced. Ion Torrent (Life Technologies, Carlsbad, CA, USA) is a better alternative for this case. Ion Torrent sequences 100 Mb to 1 Gb within several hours using various chips, which is effective for rapid and economic sequencing of a mitochondrial genome for various applications.

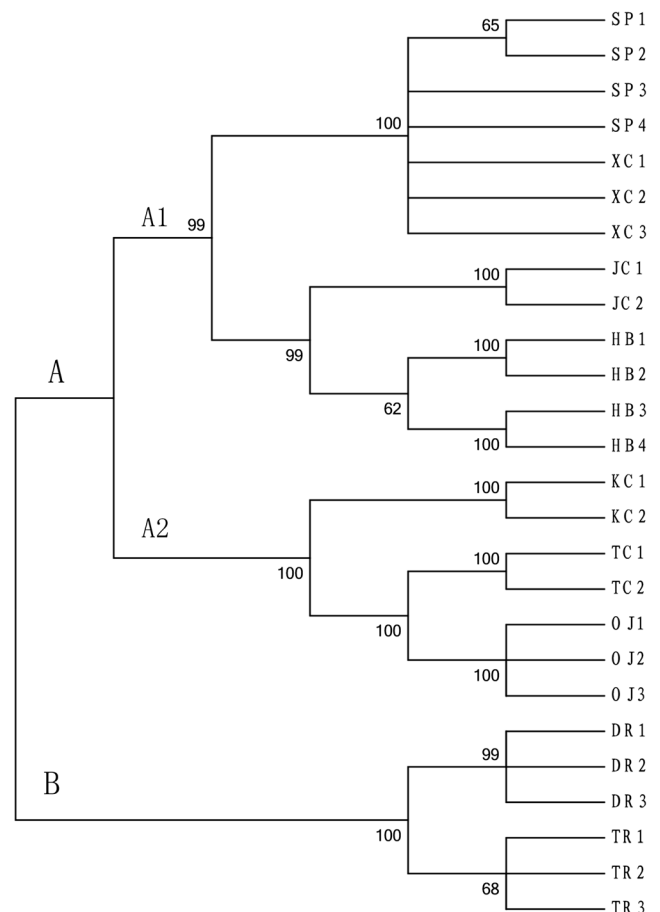


Fig. 4 Neighbor-joining-based phylogenetic tree of the full-length mitochondrial DNA sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (*A1* and *A2*) of clade A

Three red common carp strains are commonly cultured in China, such as HB, OJ, and XC, with 800, 1200, and 1300 years of documented culture history, respectively. OJ is mainly distributed in Zhejiang province, and HB and XC are cultured in Jiangxi province, as shown in Fig. 1 (Wang and Li 2004a). A phylogenetic analysis of these three strains based on cytochrome oxidase II mtDNA suggested that XC and OJ may have originated from one monophyletic group, whereas HB may have originated independently (Li and Wang 2001). However, these data were inconsistent with the biogeographic evidence that HB and XC both inhabit the Yangtze River watershed, whereas OJ is found in the Oujiang River watershed. We found here that XC and HB were clustered in clade A1, and OJ was clustered in clade A2 with TC and KC. Our data suggest that XC is more closely related to HB than OJ, which is consistent with the biogeographic evidence.

JC is bred based on the HB (maternal parent) and Yuanjiang carp (paternal parent) hybrid, which has been commonly cultured in China in the past 20 years. According to our phylogenetic analysis, JC and HB were closely clustered into

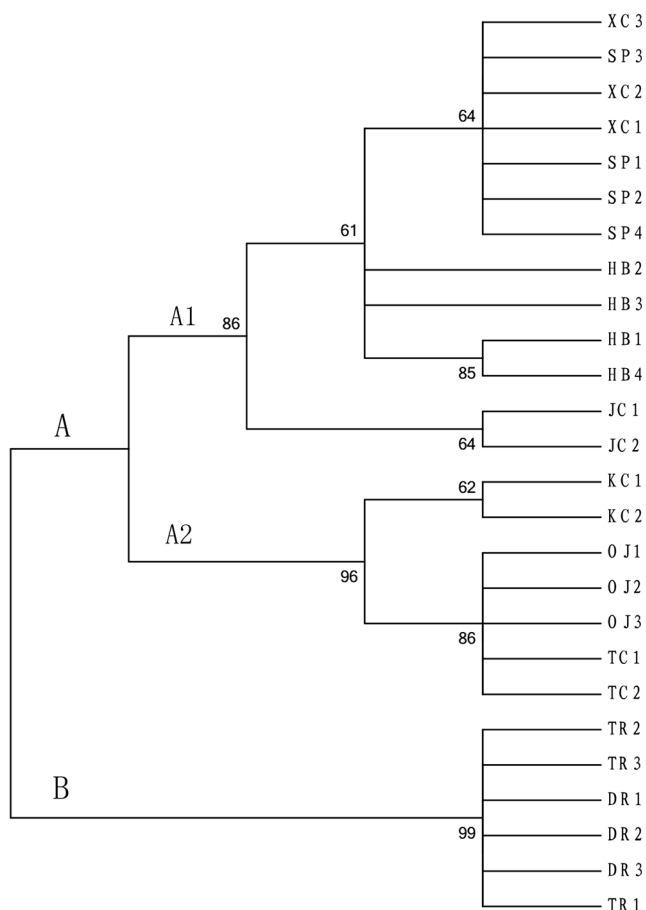


Fig. 5 Maximum-likelihood-based phylogenetic tree of cytochrome *b* (*Cyt b*) sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (*A1* and *A2*) of clade A

subclade *A1*, which is consistent with the JC breeding history. However, we also determined that the genetic distance between JC and HB was about 0.00224, suggesting that slight genetic divergence occurs in JC during intensive selection and breeding after the initial hybridization between HB and Yuanjiang carp.

KC is the most renowned ornamental fish, and more than 100 color patterns have been developed through artificial selection in the past century. However, some controversy exists in the origin of KC and its relationships with other common carp lineages. According to our phylogenetic topology, OJ, TC, and KC were clustered into subclade *A2* and separated distinctively from subclade *A1* (XC, HB, JC, and SP). Therefore, we suggest that ancestral populations of KC or TC may have originated from OJ, or they could have the same origin.

SP is a Chinese domesticated mirror carp strain based on German mirror carp and was introduced to China in the 1970s. Breeding records show that no hybrids with Asian strains have occurred. Thus, SP genetically belongs to the European

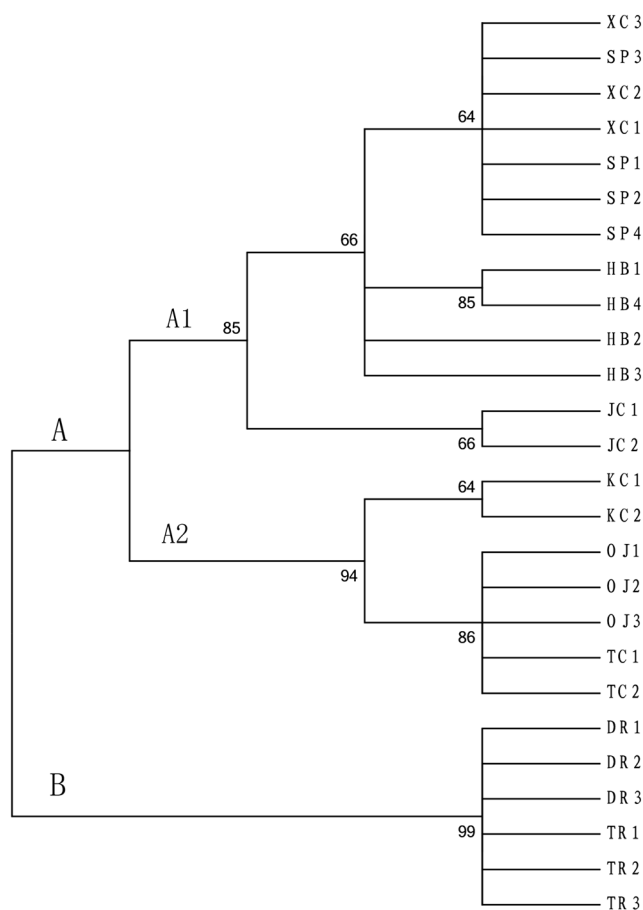


Fig. 6 Neighbor-joining-based phylogenetic tree of cytochrome *b* (*Cyt b*) sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (*A1* and *A2*) of clade A

subspecies. A phylogenetic study based on genome-wide SNP loci confirmed consanguinity (Xu et al. 2014). Surprisingly, our results show that SP was closely clustered with XC based on the very small genetic distance. The only reasonable explanation is that XC was the maternal parent when SP was domesticated in the 1970s. Notably, the divergence time between SP and HB was estimated to be about 0.91×10^3 years ago based on molecular evidence, which is a significant overestimate compared with the actual divergence time if SP hybridized with XC in China. Divergence time estimates based on commonly used mtDNA sequences likely suffer from a high substitution rate. For example, divergence time derived from mitochondrial data could be considerably overestimated and systematically biased toward the calibration point because of substitution saturation in the absence of younger more effective calibration points. The overestimate may be ten times older than the actual divergence date for recent splitting events (Zheng et al. 2011). The results of a population structure analysis based on genome-wide SNP loci showed that SP has a genetic introgression signature from Asian populations (refer

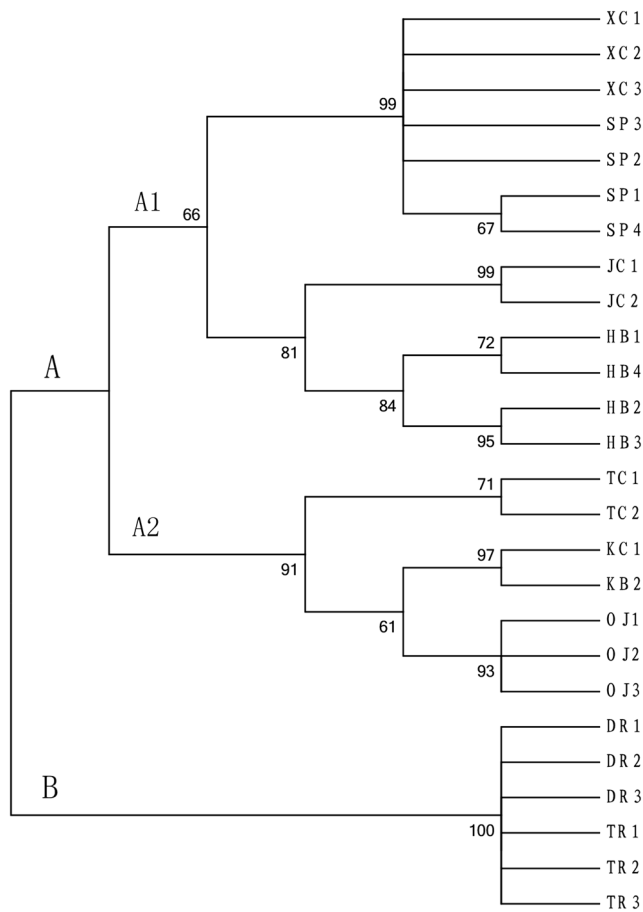


Fig. 7 Maximum-likelihood-based phylogenetic tree of the D-loop sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (A1 and A2) of clade A

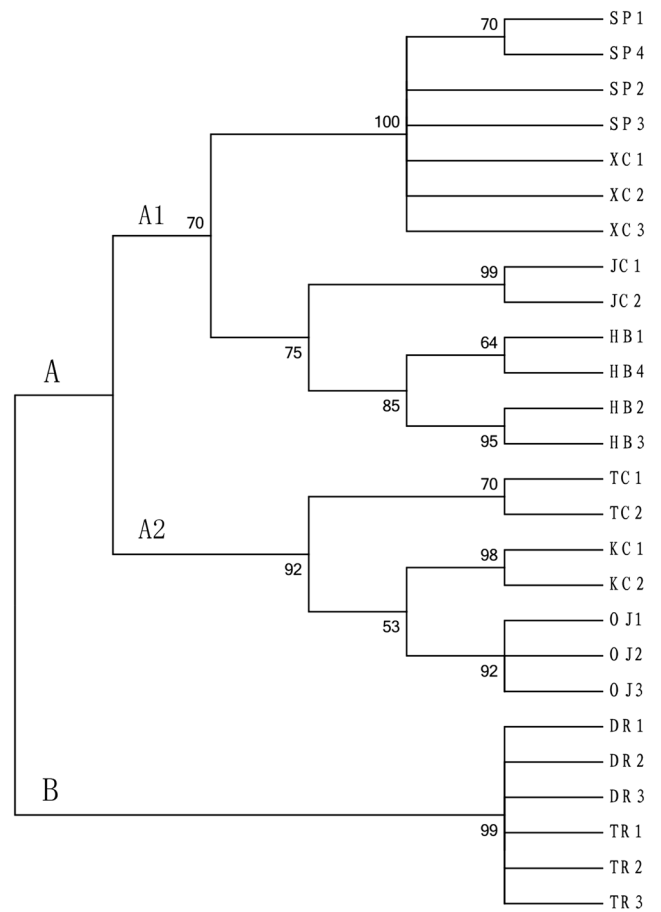


Fig. 8 Neighbor-joining-based phylogenetic tree of the D-loop sequences in 26 samples from nine strains of common carp. Bootstrap support values for each branch are included in the tree. Two distinct clades, A and B, are identified along with two subclades (A1 and A2) of clade A

to Fig. 3c in Xu et al. [2014]). Therefore, we suggest that European mirror carp was first hybridized with XC to enhance adaptability after its introduction into China and was then backcrossed continuously with European mirror carp to achieve desired traits, such as high growth rate and stress tolerance. SP was selected with stable traits and performance after several rounds of backcrossing and introgression.

Conclusion

We extracted mitochondrial genome sequences from high-throughput genome resequencing data of 26 individuals representing nine common carp lineages from Europe and East Asia. The phylogenetic analysis was performed based on these CMGs and partial mitochondrial sequences. The

Table 5 Nei’s standard genetic distance of nine common carp stains based on the CMG

	DR	TR	JC	KC	SP	OJ	TC	HB
DR								
TR	0.00030							
JC	0.00589	0.00570						
KC	0.00680	0.00662	0.00528					
SP	0.00576	0.00558	0.00315	0.00540				
OJ	0.00699	0.00680	0.00570	0.00163	0.00583			
TC	0.00686	0.00668	0.00546	0.00151	0.00558	0.00097		
HB	0.00631	0.00613	0.00224	0.00558	0.00333	0.00601	0.00577	
XC	0.00570	0.00552	0.00309	0.00534	0.00006	0.00576	0.00552	0.00327

results clarified several undetermined and controversial common carp phylogenetic relationships and breeding history and provide additional resources for germplasm conservation and genetic breeding of common carp.

Acknowledgments We acknowledge grant support from the National High-Technology Research and Development Program of China (863 Program; 2011AA100401), the National Natural Science Foundation of China (No. 31422057), and the Special Scientific Research Funds for Central Non-profit Institutes of Chinese Academy of Fishery Sciences (2013C011).

Conflict of Interest The authors declare no competing interests.

Author's Contributions PX conceived the study. CD and PX wrote the manuscript. CD, JX, and BW performed the bioinformatics analysis to obtain the complete mitochondrial genome sequences. CD conducted the phylogenetic analysis. JF and ZJ collected samples. XS supervised the genome study. All authors have read and approved the final version of the manuscript.

References

- Balon EK (1995) Origin and domestication of the wild carp. *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture* 129:3–48
- Bartfai R et al (2003) Genetic analysis of two common carp broodstocks by RAPD and microsatellite markers. *Aquaculture* 219:157–167. doi:10.1016/S0044-8486(02)00571-9
- Botero-Castro F, Tilak MK, Justy F, Catzeflis F, Delsuc F, Douzery EJP (2013) Next-generation sequencing and phylogenetic signal of complete mitochondrial genomes for resolving the evolutionary history of leaf-nosed bats (Phyllostomidae). *Mol Phylogenet Evol* 69:728–739. doi:10.1016/j.ympev.2013.07.003
- Cheng L, Liu L, Yu X, Wang D, Tong J (2010) A linkage map of common carp (*Cyprinus carpio*) based on AFLP and microsatellite markers. *Anim Genet* 41:191–198. doi:10.1111/j.1365-2052.2009.01985.x
- Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* 11. doi:10.1186/1471-2105-11-485
- Cunningham CW (1997) Can three incongruence tests predict when data should be combined? *Mol Biol Evol* 14:733–740
- Davila JI et al. (2011) Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. *BMC Biol* 9. doi:10.1186/1741-7007-9-64
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10:564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Froufe E, Magyary I, Lehoczy I, Weiss S (2002) mtDNA sequence data supports an Asian ancestry and single introduction of the common carp into the Danube Basin. *J Fish Biol* 61:301–304. doi:10.1006/jfbi.2002.2046
- Fujii S, Kazama T, Yamada M, Toriyama K (2010) Discovery of global genomic re-organization based on comparison of two newly sequenced rice mitochondrial genomes with cytoplasmic male sterility-related genes. *BMC Genomics* 11. doi:10.1186/1471-2164-11-209
- Ji PF et al (2012) High throughput mining and characterization of microsatellites from common carp genome. *Int J Mol Sci* 13: 9798–9807. doi:10.3390/Ijms13089798
- Kohlmann K, Kersten P (2013) Deeper insight into the origin and spread of European common carp (*Cyprinus carpio carpio*) based on mitochondrial D-loop sequence polymorphisms. *Aquaculture* 376:97–104. doi:10.1016/j.aquaculture.2012.11.006
- Kongchum P, Palti Y, Hallerman EM, Hulata G, David L (2010) SNP discovery and development of genetic markers for mapping innate immune response genes in common carp (*Cyprinus carpio*). *Fish Shellfish Immun* 29:356–361. doi:10.1016/j.fsi.2010.04.013
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760
- Li S, Wang C (2001) Genetic diversity and selective breeding of red common carps in China. *Naga ICLARM Quarterly* 24:56–61
- Li Y, Xu P, Zhao ZX, Wang J, Zhang Y, Sun XW (2011) Construction and characterization of the BAC library for common carp *cyprinus carpio* l. and establishment of microsynteny with zebrafish *Danio Rerio*. *Mar Biotechnol* 13:706–712. doi:10.1007/s10126-010-9332-9
- Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452. doi:10.1093/bioinformatics/btp187
- Mabuchi K, Senou H, Nishida M (2008) Mitochondrial DNA analysis reveals cryptic large-scale invasion of non-native genotypes of common carp (*Cyprinus carpio*) in Japan. *Mol Ecol* 17:796–809
- Moens LN, van der Ven K, Van Remortel R, Del-Favero J, De Coen WM (2007) Gene expression analysis of estrogenic compounds in the liver of common carp (*Cyprinus carpio*) using a custom cDNA Microarray. *J Biochem Mol Toxic* 21:299–311. doi:10.1002/Jbt.20190
- Rodriguez-Moreno L, Gonzalez VM, Benjak A, Marti MC, Puigdomenech P, Aranda MA, Garcia-Mas J (2011) Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. *BMC Genomics* 12. doi:10.1186/1471-2164-12-424
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30:2725–2729. doi:10.1093/molbev/mst197
- Tong J-g, Wu Q-J (2001) Sequence conservation on segments of mitochondrial 16S rRNA and Cytochrome b in strains of common carp (*Cyprinus carpio* L. var.). *Acta Hydrobiologica Sinica* 25:60–65
- Wang CH, Li SF (2004a) Genetic variability and relationships in mitochondrial DNA CO II gene sequence of red common carps in China. *Yi Chuan Xue Bao. Acta Genet Sinica* 31:1226–1231
- Wang CH, Li SF (2004b) Phylogenetic relationships of ornamental (koi) carp, Oujiang color carp and long-fin carp revealed by mitochondrial DNA COII gene sequences and RAPD analysis. *Aquaculture* 231: 83–91. doi:10.1016/j.aquaculture.2003.10.040
- Wang BS, Ji PF, Wang J, Sun JS, Wang CH, Xu P, Sun XW (2013) The complete mitochondrial genome of the Oujiang color carp, *Cyprinus carpio* var. color (*Cypriniformes*, *Cyprinidae*). *Mitochondrial DNA* 24:19–21. doi:10.3109/19401736.2012.710230
- White TR, Conrad MM, Tseng R, Balayan S, Golding R, Martins AMD, Dayrat BA (2011) Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships. *BMC Evol Biol* 11. doi:10.1186/1471-2148-11-295
- Wiens JJ (1998a) Combining data sets with different phylogenetic histories. *Syst Biol* 47:568–581
- Wiens JJ (1998b) Does adding characters with missing data increase or decrease phylogenetic accuracy? *Syst Biol* 47:625–640
- Williams DR et al (2008) Genomic resources and microarrays for the common carp *Cyprinus carpio* L. *J Fish Biol* 72:2095–2117. doi:10.1111/j.1095-8649.2008.01875.x
- Wu XY, Luo J, Huang S, Chen ZM, Xiao H, Zhang YP (2013) Molecular phylogeography and evolutionary history of *Poropuntius*

- huangchuchieni (Cyprinidae) in Southwest China. *PLoS One* 8. doi:10.1371/journal.pone.0079975
- Xu J, Huang W, Zhong CR, Luo DJ, Li SF, Zhu ZY, Hu W (2011a) Defining global gene expression changes of the hypothalamic-pituitary-gonadal axis in female sGnRH-antisense transgenic common carp (*Cyprinus carpio*). *PLoS One* 6. doi:10.1371/journal.pone.0021057
- Xu P et al (2011b) Genomic insight into the common carp (*Cyprinus carpio*) genome by sequencing analysis of BAC-end sequences. *BMC Genomics* 12. doi:10.1186/1471-2164-12-188
- Xu P et al. (2011c) Generation of the first BAC-based physical map of the common carp genome *BMC Genomics* 12. doi:10.1186/1471-2164-12-537
- Xu P et al (2014) Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet* 46:1212–1219. doi:10.1038/Ng.3098
- Yan XC, Ding L, Li YC, Zhang XF, Liang Y, Sun XW, Teng CB (2012) Identification and profiling of microRNAs from skeletal muscle of the common carp. *PLoS One* 7. doi:10.1371/journal.pone.0030925
- Zhang XF (2013) A consensus linkage map provides insights on genome character and evolution in common carp (*Cyprinus carpio* L.). *Mar Biotechnol* 15:275–312. doi:10.1007/s10126-012-9485-9
- Zhang Y, Liang LQ, Jiang P, Li DY, Lu CY, Sun XW (2008) Genome evolution trend of common carp (*Cyprinus carpio* L.) as revealed by the analysis of microsatellite loci in a gynogenetic family. *J Genet Genomics* 35:97–103. doi:10.1016/S1673-8527(08)60015-6
- Zheng YC, Peng R, Kuro-O M, Zeng XM (2011) Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA Sequence Data in a Particular Lineage: a case study of salamanders (order Caudata). *Mol Biol Evol* 28:2521–2535. doi:10.1093/molbev/msr072
- Zhou JF, Wang ZW, Ye YZ, Wu QJ (2003a) PCR-RFLP analysis of mitochondrial DNA ND5/6 region among 3 subspecies of common carp (*Cyprinus carpio* L.) and its application to genetic discrimination of subspecies Chinese. *Sci Bull* 48:465–468. doi:10.1360/03tb9098
- Zhou JF, Wu QJ, Ye YZ, Tong JG (2003b) Genetic divergence between *Cyprinus carpio carpio* and *Cyprinus carpio haematopterus* as assessed by mitochondrial DNA analysis, with emphasis on origin of European domestic carp. *Genetica* 119:93–97. doi:10.1023/A:1024421001015
- Zhou J, Wu Q, Wang Z, Ye Y (2004a) Genetic variation analysis within and among six varieties of common carp (*Cyprinus carpio* L.) in china using microsatellite markers *Russ. J Genet* 40:1144–1148. doi:10.1023/B:Ruge.0000044758.51875.25
- Zhou J, Wu Q, Wang Z, Ye Y (2004b) Molecular phylogeny of three subspecies of common carp *Cyprinus carpio*, based on sequence analysis of cytochrome b and control region of mtDNA. *J Zool Syst Evol Res* 42:266–269. doi:10.1111/j.1439-0469.2004.00266.x