Zoological Studies **46**(5): 569-582 (2007)

Zoological Studies

Phylogeography and Identification of a 187-bp-Long Duplication within the Mitochondrial Control Region of *Formosania lacustre* (Teleostei: Balitoridae)

Tzi-Yuan Wang[1,2], Chyng-Shyan Tzeng[1,*], Hui-Yu Teng[1], and Tiffany Chang[2]

[1]*Department of Life Science, National Tsing Hua University, 101 Kuang-Fu Road, Sec. 2, Hsinchu 300, Taiwan. Tel: 886-3-5742765. Fax: 886-3-5742765. E-mail: labtcs@gmail.com*

[2]*Genomics Research Center, Academia Sinica, 128 Academia Road, Sec. 2, Nankang, Taipei 115, Taiwan. Tel: 886-2-27898756. Fax: 886-2-27898757. E-mail: d868210@life.nthu.edu.tw*

(Accepted April 3, 2007)

Tzi-Yuan Wang, Chyng-Shyan Tzeng, Hui-Yu Teng, and Tiffany Chang (2007) Phylogeography and identification of a 187-bp-long duplication within the mitochondrial control region of *Formosania lacustre* (Teleostei: Balitoridae). *Zoological Studies* **46**(5): 569-582. An unusually long duplication was first discovered within the mitochondrial control region of the hillstream loach, *Formosania lacustre*. Over 68% of the 22 heteroplasmic individuals with the duplication were concentrated in the Tadu and Choshui Rivers of west-central Taiwan. This unusual duplication was located in the R1-repeated region of the mitochondrial genome, which includes a partial $tRNA^{Pro}$ and the 5'-end control region. The duplication has a tendency to form secondary hairpin structures. For the phylogeographic analysis, the mitochondrial control region of 68 individual loaches from 11 different river systems (10 in Taiwan and 1 in mainland China) were amplified and sequenced. Nested clade analysis divided these loach populations into 2 distinct groups: northern and central groups. Higher and lower F_{ST} values were respectively revealed in the northern and central groups. Unexpectedly, a newly recorded population was established in this study. This population was previously classified as *F. stigmata*, but the mtDNA lineage analysis and morphological comparison indicated a strong similarity of this population with *F. lacustre*. Furthermore, the nested clade analysis indicated a long-distance dispersal route from central Taiwan to the mainland China; the mainland population later dispersed into northern Taiwan during recent ice ages. We concluded that colonization between the mainland China and central Taiwan was a major influence leading to the low genetic divergence and recent recolonization of *F. lacustre*.
http://zoolstud.sinica.edu.tw/Journals/46.5/569.pdf

Key words: Heteroplasmy, Polymorphism, Balitoridae, D-loop, *Crossostoma lacustre*.

The higher mutation rate of mitochondrial (mt)DNA over nuclear (n)DNA is due to the inefficient replication repair which occurs during evolutionary processes (Brown et al. 1982, Clayton 1984), which leads to point mutations, variable-number tandem repeats (VNTRs), gene duplications, and gene rearrangements in animals (Quinn and Wilson 1993, Berg et al. 1995, Cecconi et al. 1995, Lunt et al. 1998, Macey et al. 1998, Mindell et al. 1998, Saitoh et al. 2000, Chen et al. 2004). These genetic variations are suitable markers for distinguishing intra- and inter-population differences in animals. The markers can provide useful information on the molecular phylogeny and phylogeography of closely related species.

Formosania lacustre (Steindachner), previously known as *Crossostoma lacustre*, was thought to be an endemic, small-sized hillstream loach (Cypriniformes: Balitoridae). Novaek et al. (2005) replaced *Crossostoma* with the next available earliest name, *Formosania*, as *Crossostoma* is preoccupied by a gastropod genus. The correct

*To whom correspondence and reprint requests should be addressed.

scientific name is thus *Formosania lacustre*. Unlike other hillstream loaches, found in southeastern China, *F. lacustre* is a derived species from crossostomoid fishes, and it inhabits mountain torrents of north-central Taiwan (Fang 1935). Previous studies showed that *F. lacustre* is the only species of crossostomoid fish in Taiwan, and it has not been found in mainland China.

A long variation within the mitochondrial DNA has only been reported in several vertebrates (Cecconi et al. 1995, Casane et al. 1997, Wilkinson et al. 1997, Wang et al. 1999, Ludwig et al. 2000, Taylor and Breden 2000, Freeman et al. 2001, Sipe and Browne 2003). Tzeng et al. (1990) also indicated an additional 200-base pair (bp)-long fragment located in the 5'-end control region of *F. lacustre*. Most *F. lacustre* individuals found in the Tachia and Tadu Rivers in west-central Taiwan exhibited heteroplasmic mtDNA within each individual according to cloning and restriction fragment length polymorphism (RFLP) of pure mtDNA. Due to the shortage of prior techniques, the investigation could only reveal polymorphism of the mitochondrial control region, and further studies were put on hold.

In this study, we amplified and sequenced the mitochondrial control region of *F. lacustre* from 11 different river systems, and the phylogeographic pattern of *F. lacustre* was revealed by genetic differences, which indicated a possible invasion event from Taiwan to the mainland China. This has never been discovered before in Taiwanese freshwater fishes. Furthermore, a long duplication was found in the mitochondrial control region which caused heteroplasmy, which seldom occurs in freshwater fishes. In conclusion, we are able to provide information on the types of polymorphism, the occurrence of genetic rearrangements, and patterns of the length fragment of mtDNA in *F. lacustre*.

MATERIALS AND METHODS

Fish samples

Specimens were collected from 10 rivers of northern and western Taiwan and 1 river of Fujian Province, China. Detailed sampling localities are described in table 1 and figure 1. Samples were preserved in 75% ethanol, and a piece of muscle or fin tissue was removed for DNA extraction. In total, 68 individuals were screened in this study.

Fifteen morphological measurements and counts were obtained from each specimen of the 53 samples. These samples consisted of 22 individuals of *F. lacustre*, 16 individuals of *F. stigmata*, and 15 individuals of *F. fascicauda*. Measurements were made with a digital caliper and rounded up to the nearest 0.10 mm (Fig. 2).

DNA extraction, amplification, electrophoresis, and sequencing

A piece of pectoral fin or pelvic fin, weighing about 50 mg, was immersed in 500 μl digestion buffer (10 mM Tris-HCl (pH 8.0), 1% SDS, 2 mM EDTA, 10 mM NaCl, 10 mg/ml DTT, 0.5 mg/ml proteinase K) (Kocher et al. 1989), and incubated for 16 h at 50°C in a dry bath. DNA was isolated and purified by the phenol/chloroform-isoamyl alcohol extraction procedure (Innis et al. 1989).

The region spanning the 3'-end of the *cytochrome b* (*cytb*) gene and the 5'-end of the *12S rRNA* gene was amplified using primers Ec7, G22, and PU originally described by Wang et al. (1999). PCR amplifications were performed in a 50 μl final volume containing 30-100 ng of total DNA, 200 μM of each dNTP, 0.3 μM of each primer, 1 unit of *SuperTaq* (Protech Technology Enterprise Co. Ltd., Taipei, Taiwan), and a buffer supplied by the manufacturer. The PCR conditions were optimized as follows: 35 cycles of hot start-up at 93°C for 3 min, denaturing at 93°C for 30 s, annealing at 50°C for 40 s, and extension at 72°C for 1 min, with a final extension at 72°C for 10 min.

The PCR products were electrophoresed in a 2% agarose gel (FMC Bioproduct, Rockland, ME, USA) and 0.5x TAE buffer to check the yields. The heteroplasmic products were extracted using a DNA/RNA gel extraction column (Viogene Biotek Corp., Taipei, Taiwan) by standard procedures. Ambiguous DNA was cloned into the pUC19 vector and sequenced by M13-20 and M13-rev primers (Cedar Creek, TX, USA) to confirm the heteroplasmic sequences.

DNA sequencing was performed with commercial sequence kits (BigDye™ Terminator Cycle Sequencing Ready Reaction Kits of Applied Biosystems, Foster, CA, USA) and an ABI model 377 automated DNA sequencer was used to obtain the sequence data. A fragment of about 1.2 kb was sequenced for all samples in both directions. Each haplotype was submitted to GenBank with the accession numbers AY283688-AY283725 and AY283730-AY283732.

Sequence prediction

The sequences were aligned using ClustalX (Thompson et al. 1997) and adjusted manually. Putative gene regions (cytb, tRNA-*Thr*, $tRNA^{Pro}$, control, and 12S rRNA) and termination-associated sequence/conserved sequence block (TAS/CSB) structures were localized by aligning the sequences. The identified TASs and CSBs were then compared to the conserved features of other vertebrates and confirmed using a tRNAscan-SE (Lowe and Eddy 1997). The secondary structure of the repeated sequence was predicted by tRNAscan-SE and RNA mfold 3.1 (Zuker 2003).

Genetic divergence

Genetic distances were calculated by Tamura-Nei gamma distances (Tamura and Nei 1993) with 500 bootstrap replications (Felsenstein 1985) for standard error computation, as implemented in MEGA (vers. 2.1, Kumar et al. 2001). Genetic diversity was quantified at the inter- and intra-population levels using DnaSP (vers. 4.10.7, Rozas et al. 2003) to calculate the index of haplotype diversity (h) (Nei 1987), estimates of nucleotide diversity (π) (Nei 1987), and F_{ST} for gene flow (Hudson et al. 1992).

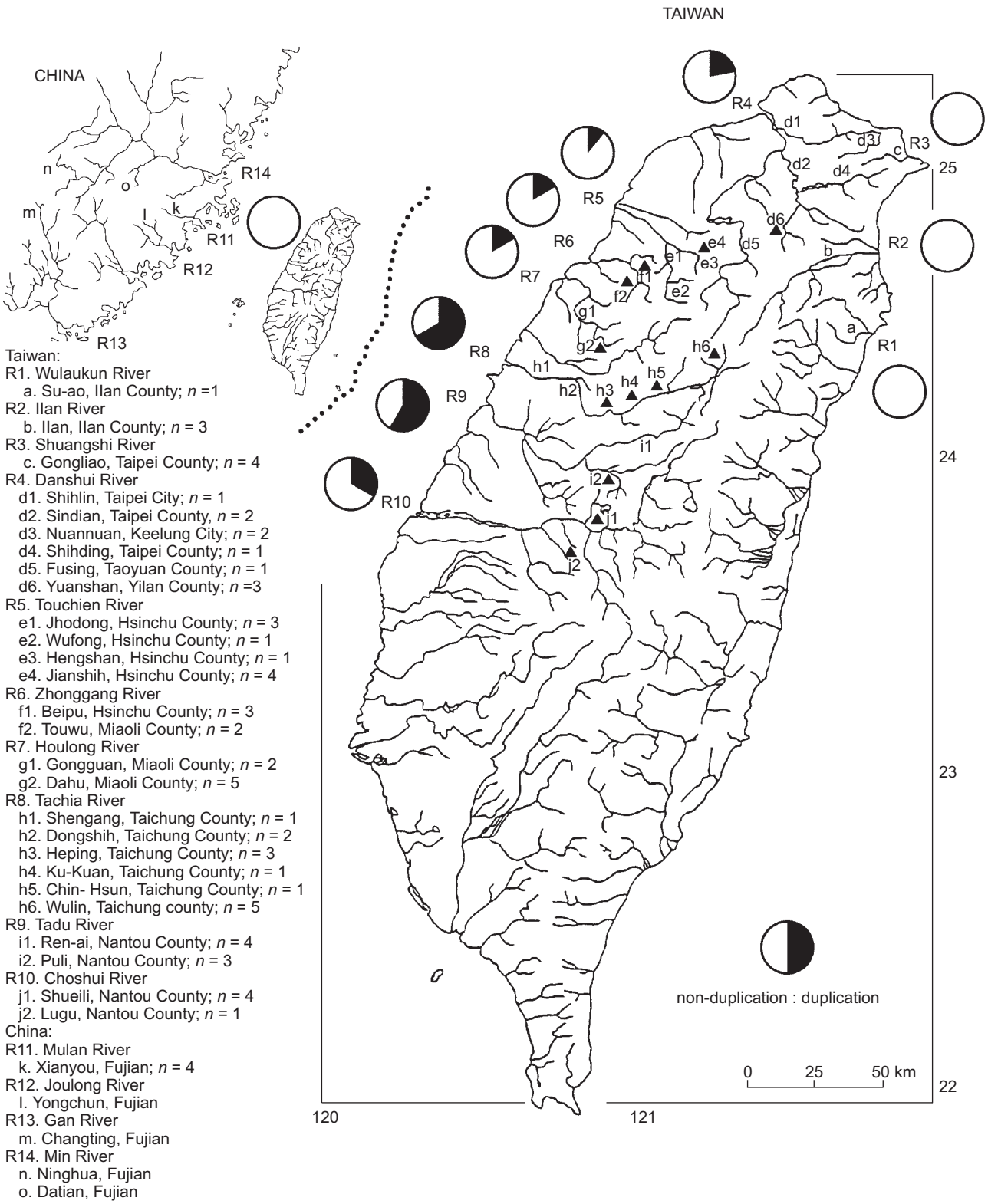
Nested clade analysis (NCA)

The data of genetic distances were employed to establish a minimum spanning network with the TCS vers. 1.13 (Clement et al. 2000). The NCA is able to incorporate expectations from population genetics into inferences about the recent geographic history of a population at the intraspecific level. There are 4 steps in the NCA: constructing a haplotype network, nesting clades on the network, testing for geographic associations, and formulating inferences about the processes (fragmentation, long-distance colonization, and isolation by distance) that have generated the pattern. The NCA is appropriate for studying geographical variations within species and was performed using GeoDis (Posada et al. 2000). The inference key was provided by Posada and Templeton (2005).

Table 1. Sample information of selected mitochondrial (mt)DNA for sequencing the D-loop

River basin / Locality	Population code	No. of individuals sampled[a]	mtDNA lineage	Haplotypes	Proportion of heteroplasmic individuals[b]	Tandem repeat $(AT)_n$	Haplotype diversity (h)	Nucleotide diversity (π)
Wulaukun R., Northeastern Taiwan	R1	1	2-1 (WN)	Cl1	-	$(AT)_6$	-	-
Ilan R., Northeastern Taiwan	R2	3	2-1 (WN)	Cl1, Cl2	-	$(AT)_6$	0.66667	0.0015 ± 0.0010
Shuangshi R., Northeastern Taiwan	R3	4	2-1 (WN)	Cl3	-	$(AT)_8$	0	0.0000 ± 0.0000
Danshui R., Northern Taiwan	R4	10	2-1 (WN)	Cl1, Cl3-Cl7	9.09%	$(AT)_6$	0.86667	0.0040 ± 0.0013
Touchien R., Northwestern Taiwan	R5	9	3-1 (WC)	Cl8-Cl12	4.55%	$(AT)_6$	0.72222	0.0030 ± 0.0011
Zhonggang R., Northwestern Taiwan	R6	5	3-1 (WC)	Cl13-Cl16	4.55%	$(AT)_6$	1	0.0022 ± 0.0009
Houlong R., Northwestern Taiwan	R7	7	1-6 (WC)	Cl17-Cl20, Cl31	4.55%	$(AT)_6$	0.71429	0.0034 ± 0.0011
Tachia R., West-central Taiwan	R8	13	3-1 (WC)	Cl21-Cl30	36.36%	$(AT)_6$	0.96154	0.0050 ± 0.0016
Tadu R., West-central Taiwan	R9	7	3-1 (WC)	Cl21, Cl23, Cl30-Cl32	31.82%	$(AT)_6$	0.85714	0.0044 ± 0.0015
Choshui R., West-central Taiwan	R10	5	1-5 (WC)	Cl23, Cl33, Cl34	9.09%	$(AT)_6$, $(AT)_{12}$	0.9	0.0011 ± 0.0008
Mulan R., East-central China	R11	4	2-1 (CN)	Cla, Clb, Clc	-	$(AT)_6$	0.83333	0.0030 ± 0.0014

[a]Sixty-eight individuals were used to sequence the control region. [b]Twenty-two heteroplasmic individuals were found in this study.



Taiwan:
R1. Wulaukun River
a. Su-ao, Ilan County; *n* =1
R2. Ilan River
b. Ilan, Ilan County; *n* = 3
R3. Shuangshi River
c. Gongliao, Taipei County; *n* = 4
R4. Danshui River
d1. Shihlin, Taipei City; *n* = 1
d2. Sindian, Taipei County, *n* = 2
d3. Nuannuan, Keelung City; *n* = 2
d4. Shihding, Taipei County; *n* = 1
d5. Fusing, Taoyuan County; *n* = 1
d6. Yuanshan, Yilan County; *n* =3
R5. Touchien River
e1. Jhodong, Hsinchu County; *n* = 3
e2. Wufong, Hsinchu County; *n* = 1
e3. Hengshan, Hsinchu County; *n* = 1
e4. Jianshih, Hsinchu County; *n* = 4
R6. Zhonggang River
f1. Beipu, Hsinchu County; *n* = 3
f2. Touwu, Miaoli County; *n* = 2
R7. Houlong River
g1. Gongguan, Miaoli County; *n* = 2
g2. Dahu, Miaoli County; *n* = 5
R8. Tachia River
h1. Shengang, Taichung County; *n* = 1
h2. Dongshih, Taichung County; *n* = 2
h3. Heping, Taichung County; *n* = 3
h4. Ku-Kuan, Taichung County; *n* = 1
h5. Chin- Hsun, Taichung County; *n* = 1
h6. Wulin, Taichung county; *n* = 5
R9. Tadu River
i1. Ren-ai, Nantou County; *n* = 4
i2. Puli, Nantou County; *n* = 3
R10. Choshui River
j1. Shueili, Nantou County; *n* = 4
j2. Lugu, Nantou County; *n* = 1
China:
R11. Mulan River
k. Xianyou, Fujian; *n* = 4
R12. Joulong River
l. Yongchun, Fujian
R13. Gan River
m. Changting, Fujian
R14. Min River
n. Ninghua, Fujian
o. Datian, Fujian

Fig. 1. Sample localities and the heteroplasmic proportions in each river. ▲, Indicates the localities where heteroplasmic individuals appeared. Pie charts give the apparent proportion of individuals with the 187-bp duplication within each river. Details are given in table 1.

RESULTS

Gene structural analysis

Sixty-eight individuals of *F. lacustre* were sampled from 10 different river basins in Taiwan and 1 in mainland China, and 22 individuals exhibited length variations (heteroplasmic fragments) within their mitochondrial control regions (Table 1). Individuals from northeastern Taiwan rivers were homoplasmic while the rest showed various extents of heteroplasmy (Fig. 1). Over 68% of the heteroplasmic individuals were concentrated in the Tadu and Choshui Rivers of west-central Taiwan.

The gene structure of the mitochondrial control region showed a unique direct repeat located in the left domain (Fig. 3). This repeat included a 42-bp origin from the 3'-end of the mitochondrial $tRNA^{Pro}$ and 145 bp from the 5'-end of the mitochondrial control region. We observed that CCTGG repeat sequences were located at both the 5'- and 3'-ends, so these repeat segments possibly participate in duplication of the unique repeat in the left domain, which in turn could have resulted in the heteroplasmy within the mtDNA of *F. lacustre*. Some functional motifs related to duplication, including TAS, CSB-D, CSB-II, and CSB-III, were discovered in our study as well (Fig. 4). The classification results indicated that the unique direct repeat within the left domain of the mitochondrial control region belonged to the R1 repeat. In addition, AT repeats observed in the right domain were classified as the R2 repeat. This dinucleotide repeat occurred 6 times in most individuals, but in individuals from the Shuangshi and Choshui Rivers, the AT repeats were respectively duplicated 8 and 12 times (Table 1).

Secondary structure predictions

Several other predictions were employed to further assist in understanding the stabilities of the heteroplasmic fragments within the mitochondrial control region. Two different computer programs were utilized to predict the free energy of all secondary structures of the heteroplasmic fragments, and the results of the highly stable structures (with a minimum amount of free energy) are illustrated in figure 5. Although 2 different computer programs predicted 2 different results, the highly stable structures revealed similar patterns (data not shown). The predicted secondary structure of each direct repeat of the left domain indicated an extra hairpin structure in the mitochondrial control region (Fig. 5A). Interestingly, this hairpin structure could form by itself even after separation from the control region (Fig. 5B).

Genetic divergence

Thirty-seven haplotypes were found among all samples (Table 2). In addition, there were 42 variable sites and 18 parsimoniously informative sites among the haplotypes; only 3 parsimoniously informative sites were in the direct repeat on the left domain of the mitochondrial control region. The DnaSP data revealed that the haplotype diversity (*h*) ranged from 0 to 1 (the value of the pooled specimens was 0.97558) within populations (Table 1). The nucleotide diversity (π) ranged from 0 to 0.0058 (the value of pooled specimens was 0.0071 ± 0.0017) within populations. Table 3 shows that the inter-population genetic distances ranged 0.15%-1.28%. The highest genetic distance (1.28%) was between the Tadu River population (R9) and populations of northern Taiwan (R1 and R2). Furthermore, genetic distances between the northern Taiwanese populations and the Chinese population (R11) were smaller than those compared to populations in central Taiwan.

The F_{ST} values between the rivers of northern Taiwan (R2, R3, and R4) and others were higher than the mean value (0.65). The F_{ST} value between the Mulan River in mainland China (R11) and others were also higher than the mean value. In contrast, low F_{ST} values were exhibited in central Taiwan (R5, R6, R7, R8, R9, and R10), indicating a greater occurrence of gene flow in the region.

Nested clade analysis (NCA)

Two separate groups of *F. lacustre* were classified according to their genetic structures (Fig. 6). Surprisingly, the first group included 1 population from mainland China (CN) and populations from the Wulaukun, Ilan, Shuangshi, and Danshui Rivers (the northern group, WN). The second group was the central group (WC), which contained the rest of the populations from 6 rivers of western Taiwan. The NCA revealed that the genetic distance was about 0.99% ± 0.28% between the northern and central groups. Moreover, the analysis also indicated that the genetic distance of the population from mainland China was 0.63% ± 0.21% with the northern Taiwanese group and 0.86% ± 0.24% with the central Taiwanese group.

The NCA of the total cladogram provided 2

possible explanations for the phylogeography of *F. lacustre* (Table 4). The first explanation predicted that the gene flow in the past was followed by the extinction of some populations in the intermediate area. Another explanation indicates the long-distance dispersal over intermediate areas not occupied by this species.

In addition, the same analysis also showed that colonization and recent fragmentation were 2 major mechanisms inferred for the northern and central subclades. Most of the duplication sequences were found in the interior nodes of the central group and also found in several individuals of the northern group (Fig. 6).

Table 2. Haplotypes found in each population of *Formosania lacustre*

Haplotype	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Cl1	1	1		3							
Cl2		2									
Cl3			4	1							
Cl4				3							
Cl5				1							
Cl6				1							
Cl7				1							
Cl8					1						
Cl9					1						
Cl10					5						
Cl11					1						
Cl12					1						
Cl13						2					
Cl14						1					
Cl15						1					
Cl16						1					
Cl17							1				
Cl18							1				
Cl19							1				
Cl20							1				
Cl21								1	1		
Cl22								1			
Cl23								3	1	2	
Cl24								1			
Cl25								1			
Cl26								1			
Cl27								1			
Cl28								1			
Cl29								1			
Cl30								2	3		
Cl31							3		1		
Cl32									1		
Cl33										2	
Cl34										1	
Cla											1
Clb											1
Clc											2

Morphological comparisons

Due to the shallow genetic divergence between the Chinese population (CN) and Taiwanese populations (WN and WC), clarifying whether the Chinese population belongs to *F. lacustre* can assist in explaining the genetic diversity. Therefore, 22 specimens of *F. lacustre* and 31 specimens of the closely related species, *F. stigmata* and *F. fascicauda*, were morphometrically analyzed. In total, 15 measurements and counts were made (Table 5). Most measurements and counts showed that the CN population was almost the same as *F. lacustre*, but differed from *F. stigmata* and/or *F. fascicauda*. For example, the ratios of the head length to mouth width showed that the CN population was similar to *F. lacustre* (*t*-test; $p > 0.05$) and significantly differed from the other 2 species ($p < 0.001$). Thus, the morphological comparison indicated that the CN population is a new record of *F. lacustre* in mainland China.

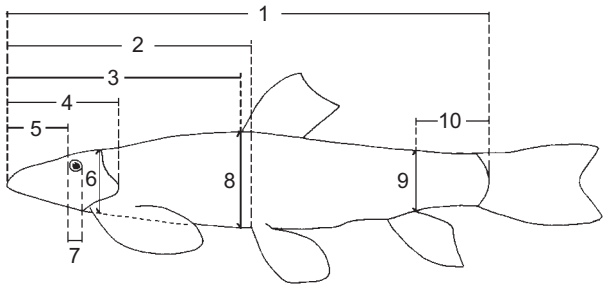


Fig. 2. Measurements of *Formosania species*. 1, Standard length, measured from the tip of the head to the posterior edge of the hypural plate; 2, prepelvic length, measured from the tip of the upper jaw to the anterior edge of the pelvic insertion; 3, predorsal length, measured from the tip of the upper jaw to the anterior edge of the dorsal insertion; 4, head length, measured from the tip of the head to the posterior edge of the operculum; 5, snout length, measured from the tip of the head to the anterior margin of the osseous orbit; 6, head depth, at the level of the occiput; 7, orbital diameter, the distance between the horizontal margins of the osseous orbit; 8, body depth, measured at the level of the dorsal fin origin; 9, caudal peduncle depth, measured at the level of the end of the anal fin base; 10, caudal peduncle length, measured from the end of the anal fin base to the end of the hypural plate; interorbital width, the cross distance between the upper margins of each osseous orbit; head width, the cross distance between the anterior pectoral fin bases; body width, the cross distance between the anterior pelvic fin bases; mouth width, the cross distance between the interior corners of the mouth.

DISCUSSION

Our present study revealed an unusual duplication of the mitochondrial control region in various *F. lacustre* populations, and this duplication was absent from populations in northeastern Taiwan (R1, R2, and R3) and mainland China (R11). However, the duplication was observed in 3/4 of individuals from the Tachia, Tadu, and Choshui Rivers of west-central Taiwan. In addition, most of the heteroplasmic individuals were found in the upper reaches of these rivers (Fig. 1).

Formation of the direct duplication

The 187-bp duplication within the mitochondrial control region of *F. lacustre* has been seldom reported in vertebrates. This unusual duplication forms a hairpin structure consisting of a partial 5'-control region of the TAS motif, and 45 bp of the partial tRNA^{Pro} gene. Ludwig et al. (2000) indicated a similar duplication, which was located near the 5' end of the D-loop and was separated by a few nucleotides from the tRNA^{Pro} gene. This duplication was found to be responsible for length variations and heteroplasmy in sturgeon. Similar repeat units as mentioned above were also discovered in seabass (Cecconi et al. 1995) and 3 percid fish (Nesbo et al. 1998). The TAS motifs discovered in this study resemble motifs in previous studies (Fig. 2).

Several models were proposed to explain the formation of the direct duplication found in our study. The illegitimate elongation model was designed to explain the frequent misalignment in the mitochondrial repeat region prior to elongation. This misalignment was facilitated by a stable secondary structure in the displaced strand, which led to a direct repeat within the D-loop (Buroker et al. 1990). Taylor and Breden (2000) proposed that slipped-strand mispairings at noncontiguous repeats could give rise to a minisatellite. Freeman et al. (2001) discovered an 80-bp repetitive sequence motif, which is believed to form the secondary structure, and it may function in regulating control region replication that could in turn cause

Table 3. Genetic distance and gene flow between different rivers. The upper right matrix and lower left matrix respectively contain F_{ST} values and genetic distances. A high F_{ST} value indicates low gene flow between rivers

	R1	R2	R3	R4	R5
R1		-	-	-	-
R2	0.0015 ± 0.0011		0.9	0.3	0.74
R3	0.0056 ± 0.0024	0.0071 ± 0.0025		0.7	0.89
R4	0.0025 ± 0.0008	0.0040 ± 0.0013	0.0047 ± 0.0017		0.65
R5	0.0099 ± 0.0030	0.0099 ± 0.0029	0.0095 ± 0.0031	0.0105 ± 0.0029	
R6	0.0109 ± 0.0034	0.0109 ± 0.0033	0.0093 ± 0.0032	0.0112 ± 0.0032	0.0044 ± 0.0017
R7	0.0120 ± 0.0033	0.0120 ± 0.0032	0.0090 ± 0.0029	0.0118 ± 0.0031	0.0074 ± 0.0023
R8	0.0123 ± 0.0034	0.0123 ± 0.0033	0.0100 ± 0.0030	0.0123 ± 0.0032	0.0057 ± 0.0017
R9	0.0128 ± 0.0035	0.0128 ± 0.0034	0.0095 ± 0.0029	0.0125 ± 0.0032	0.0060 ± 0.0020
R10	0.0111 ± 0.0034	0.0111 ± 0.0034	0.0122 ± 0.0037	0.0122 ± 0.0034	0.0053 ± 0.0019
R11	0.0074 ± 0.0027	0.0074 ± 0.0026	0.0062 ± 0.0025	0.0076 ± 0.0024	0.0076 ± 0.0023

	R6	R7	R8	R9	R10	R11
R1	-	-	-	-	-	-
R2	0.77	0.74	0.67	0.76	0.68	0.72
R3	0.89	0.85	0.8	0.87	0.74	0.84
R4	0.68	0.65	0.6	0.67	0.64	0.62
R5	0.4	0.54	0.26	0.4	0.51	0.68
R6		0.39	0.29	0.44	0.56	0.69
R7	0.0048 ± 0.0017		0.39	0.49	0.6	0.7
R8	0.0053 ± 0.0018	0.0076 ± 0.0023		-0.03	0.38	0.62
R9	0.0059 ± 0.0021	0.0078 ± 0.0024	0.0044 ± 0.0014		0.43	0.69
R10	0.0061 ± 0.0023	0.0095 ± 0.0030	0.0044 ± 0.0015	0.0044 ± 0.0016		0.65
R11	0.0071 ± 0.0024	0.0102 ± 0.0030	0.0092 ± 0.0025	0.0095 ± 0.0027	0.0094 ± 0.0029	

convergent evolution. All of these models were able to explain the presence of the duplication found in this study.

In this study, the stable hairpin structure and the slipped-strand mispairing led to a partial repeat near the tRNA^{Pro} gene participating in gene duplication. When the proto-repeat is folded into a stem-loop structure during gene replication, then this repeat sequence can be the anchor for the new heavy strand to the light strand, thereby producing gene duplication. In addition, the duplication process would leave partial repeats at both ends of the duplicated region. Thus, the functional constraint in the replication control region can lead to R1 repeats with similar lengths and cause convergent evolution.

Mitochondria are cellular organelles important in cellular respiration, and mutagenic reactive oxygen species are generated in the process of energy production (Croteau and Bohr 1997). Since mtDNA is subjected to relatively high amounts of oxidative damage, it seems that mitochondria would need efficient DNA repair mechanisms to repair oxidative damage from its DNA (Croteau and Bohr 1997). In response to low oxygen levels, mitochondria produce chemicals such as hydrogen peroxide to induce apoptosis (Simon et al. 2000). An increase in hydrogen peroxide in mitochondria

Fig. 3. The repeated sequence and control region of *Formosania lacustre*. The shaded nucleotides ( ) are tRNA^{Pro}; underlined letters are the direct repeat; "CCTGG", and the putative termination-associated sequence elements appear between repeats.

might cause mutations and damage to mtDNA. Duplications within mtDNA can counterbalance the damage and repair mtDNA replication, so the multiple repeats are able to increase the longevity of mitochondria. The multiple repeats within mtDNA can compensate for deleterious mutations during the lifetime of an individual (Wilkinson et al. 1997). This might explain the accumulation of a large quantity of duplications in mitochondria and the functional constraints causing heteroplasmy in *F. lacustre*.

Phylogeography

Formosania lacustre was previously thought to be a species endemic to Taiwan, and the Chinese population in central Fujian Province was previously reported to be *F. stigmata* Nichols (Fang 1935, Tang and Chen 2000). However, the genetic relationships in our study showed that the Chinese population of *F. stigmata* is in fact a new distribution record for *F. lacustre*. Figure 6 reveals that the mtDNA lineages of populations from northern Taiwan (clades 1-1 and 1-2) are closely related to the population from mainland China (clade 1-3). Thus, this population should become a new lineage of *F. lacustre* instead of *F. stigmata* according to the genetic relationship established and morphological comparisons in this study (Tables 3, 5). The phylogenetic relationships within *Formosania* also support this conclusion (Wang 2004).

The NCA predicted that the new lineage was caused by gene flow in the past, which was then followed by extinction of populations in intermediate areas (Table 4). The principle concept of the NCA states that an ancestral center should be defined by 2 characteristics: first, it must be the interior node of a haplotype network, and second, it must contain the most haplotypes among all clades. Thus, clade 2-2 (WC) is classified as the ancestral center according to the analysis because

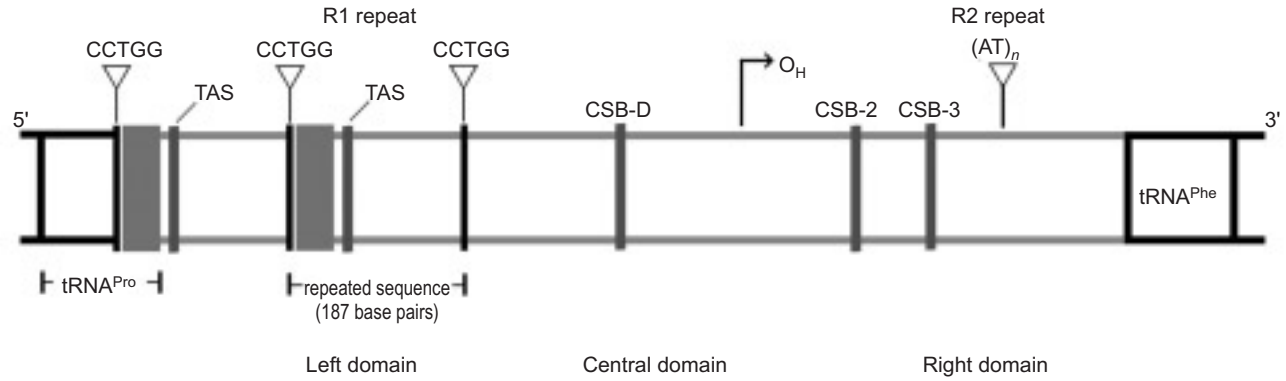


Fig. 4. Genetic structure of the mitochondrial control region. The 187-bp duplication of *Formosania lacustre* includes 42 bp of partial $tRNA^{Pro}$ and 145 bp of the 5'-end control region. The gray region is the control region, which is located between $tRNA^{Pro}$ and $tRNA^{Phe}$. The central domain consists of a conserved sequence block-D (CSB-D) that is involved in heavy-strand replication; furthermore, this replication is initiated by strand displacement at the heavy-strand origin (O_H), which is located in the right domain along with 3 CSBs (CSB-1, -2, and -3). One or more copies of the termination-associated sequences (TASs) can usually be identified in the left domain, and these TASs can signal termination of the displacement strands (D-loop; Lee et al. 1995).

Table 4. Chain of inference from the nested clade analysis. Haplotype and clade designations are given in figure 6

Clade	Inference chain	Inferred pattern
1-1	1-2-11-12 NO	Colonization event is inferred, perhaps associated with recent fragmentation (CRF)
2-1	1-2-11-12-13 YES	Long-distance colonization possibly coupled with subsequent fragmentation (LDC-SF) or past fragmentation followed by range expansion (PF-RE)
1-4	1-2-11-12 NO	Colonization event is inferred, perhaps associated with recent fragmentation (CRF)
2-2	1-2-11-12-13-14 YES	Past fragmentation (PF) or long distance colonization (LDC)
2-3	1-2-11-12 NO	Colonization event is inferred, perhaps associated with recent fragmentation (CRF)
Total	1-2-3-5-6-7-8 NO	Past gene flow followed by extinction of intermediate populations (GF-EX) or restricted gene flow/dispersal but with some long-distance dispersal over intermediate areas not occupied by the species (GF-LF)

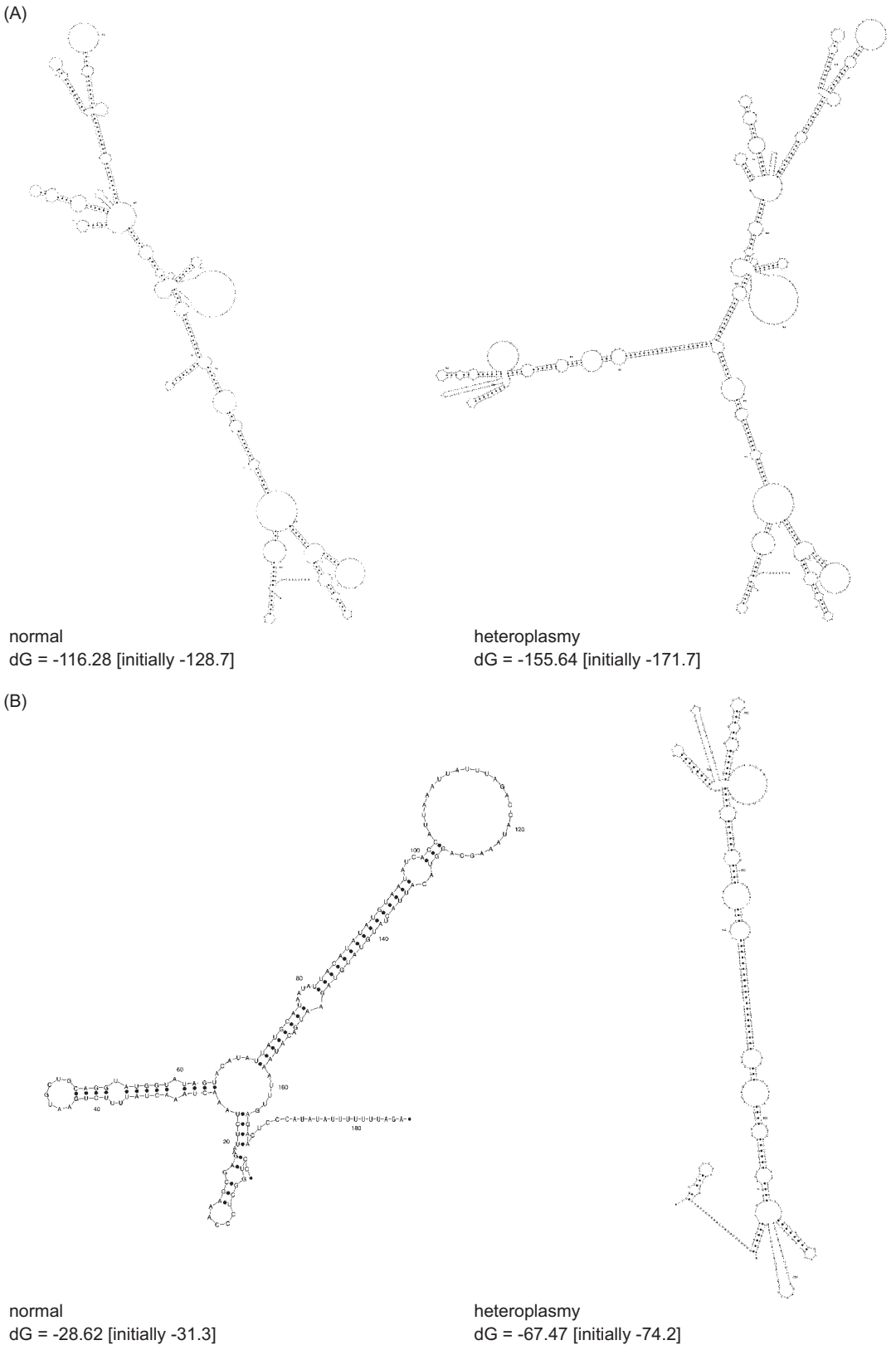


Fig. 5. Secondary structure prediction of the duplicated region. (A) Left domain of the control region: an extra-hairpin loop was formed in the heteroplasmic sequence; (B) only repeated sequence: a long hairpin was also formed in the heteroplasmic sequence. dG, free energy (kcal/mole).

it is the interior node of the haplotype network; furthermore, 12 of 34 haplotypes are clustered in this clade. Therefore, a sequential dispersal route may have occurred from clade 2-2 (WC) to clade 2-1 (CN+WN). In addition, the slightly larger genetic distance also indicates that the central populations (WC) are more distinct from the northern populations (CN+WN). Low genetic divergence among WC, WN, and CN implies a recent separation among these populations. The F_{ST} values also showed that lower gene flows have occurred among these distinct populations, while higher gene flows have occurred in the central populations. Therefore, the populations from central

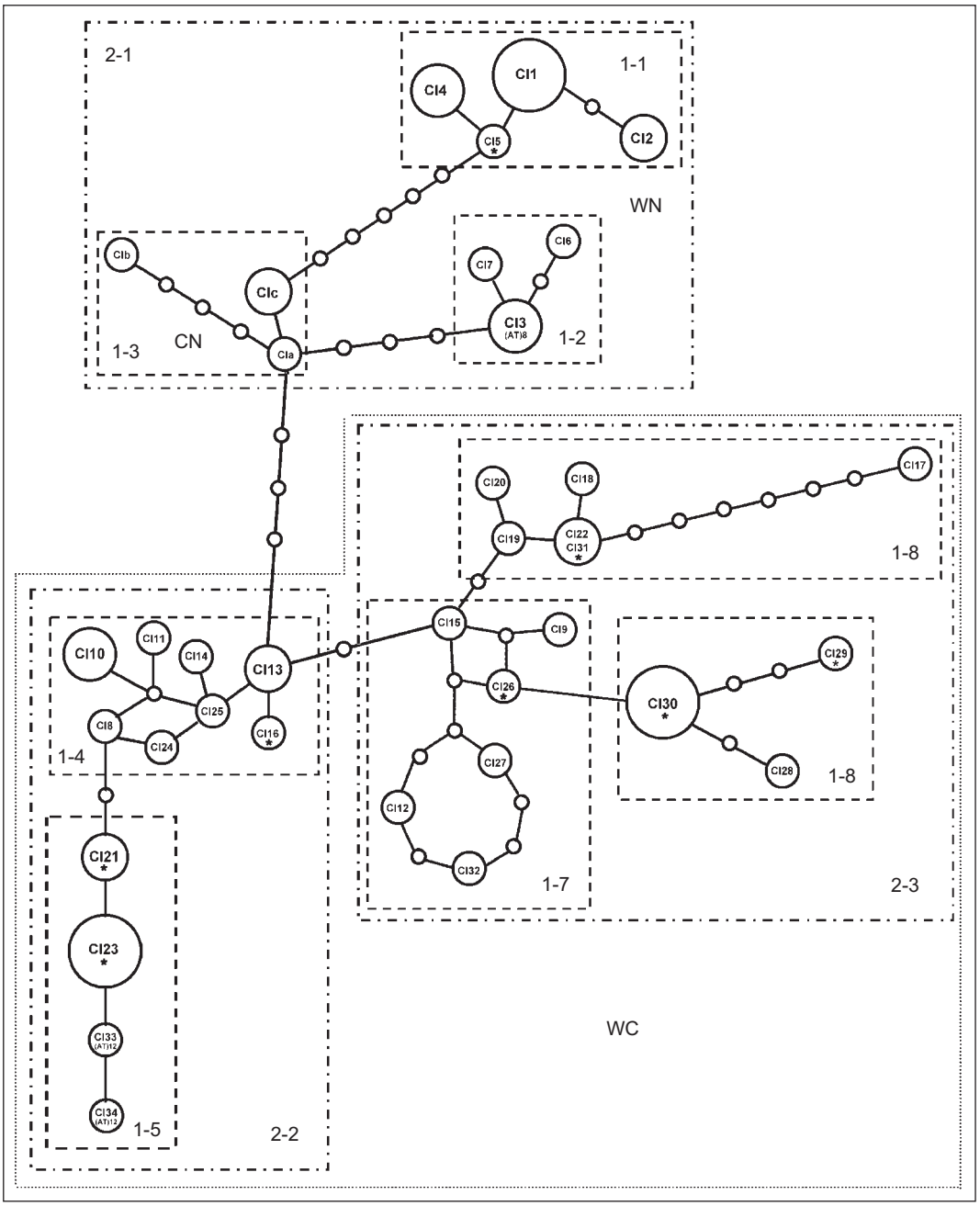


Fig. 6. Haplotype network based on the complete deletion sequences. Cl30 and Cl26 were identical when 1 indel was excluded. Evolution of the direct repeat is ill-defined as a single event or multiple events; therefore, the repeat was excluded when constructing the haplotype network and then was readdressed in the original haplotypes. An asterisk (*) represents the appearance of the R1 repeat, while (AT) *n* indicates tandem repeats. *n*, number of times repeated.

Taiwan could have dispersed to mainland China, and then later made it to northern Taiwan, which was followed by geographical isolation. According to the literature, many freshwater fishes dispersed from mainland China to Taiwan and then subdivided in Taiwan (Tzeng et al. 2006, Wang et al. 2007). However, no record has indicated a colonization route from Taiwan to mainland China. The unusual colonization route in our study implies that long-distance migrations might have taken place when the sea level was more than 100 m lower during the last Pleistocene glaciation, and were facilitated by the land bridge connection between mainland China and north-central Taiwan. Recently, the genetic structure of freshwater shrimp also revealed a similar colonization pattern, which implies that some animals might have also dispersed from Taiwan to the mainland China during the ice ages (Liu 2006, Liu et al. 2007). Further comparative phylogeography may provide better and more-detailed scenarios for the colonization routes of freshwater animals.

In this study, a new phylogeographical pattern and an unusual gene duplication were observed in *F. lacustre*. A newly recorded population from the Mulan River of Fujian Province, China may be a surviving intermediate population or recolonization during the ice ages. Further studies will focus on detailed analyses of anatomical comparisons among the closely related *Formosania* species in East Asia. The recent isolation due to lower sea levels during the last ice age may have led to the lower genetic divergence of *F. lacustre* between mainland China and Taiwan.

Table 5. Morphological comparisons of *Formosania lacustre* with its sister species. Unbranched fin-rays are shown as Roman numerals, and branched rays are shown as Arabic numerals. *n*, number of individuals sampled. WN, WC, and CN represent the subdivision within the haplotype network in figure 6

Species	*F. lacustre*			*F. stigmata*	*F. fascicauda*
Locality (*n*)	WN, Northern Taiwan (11)	WC, West-central Taiwan (8)	CN, Mulan R., China (3)	Southern Fujian[a], China (16)	Southern Fujian[b], China (15)
Dorsal fin rays	III/7-8 (7.9)	III/8-9 (8.3)	III/8-9 (8.7)	III/7-9 (8.6)	III/7-9 (8.3)
Anal fin rays	II/5-6 (5.5)	II/5-6 (5.3)	II/7-8 (7.3)	II/5-6 (5.8)	II/5-7 (5.9)
Pectoral fin rays	I/13-15 (14.2)	I/13-15 (13.9)	I/14-15 (14.7)	I/13-15 (14.1)	I/10-15 (13.1)
Ventral fin rays	I/8-10 (8.4)	I/8 (8)	I/8 (8)	I/7-8 (7.9)	I/7-9 (7.9)
Lateral line scales	82-107 (97)	91-125 (110)	88-89 (88)	83-105 (94)	84-109 (92)
Standard length (mm)	42.7-95.1 (54.34)	42-83 (61)	43.6-60.1 (49.80)	39.0-84.2 (63.28)	43.4-77.0 (61.60)
Standard length / body depth[c]	4.85-6.08 (5.47)	5.41-7.54 (6.12)	5.25-5.56 (5.39)	4.5-6.8(5.86)	3.7-6.2 (5.11)
Standard length / head length	4.47-5.05 (4.7)	4.34-5.45 (4.89)	4.32-5.07 (4.62)	4.1-5.0(4.54)	3.9-5.7 (4.44)
Standard length / caudal peduncle length[c]	7.12-12.2 (8.76)	8.6-10.5 (9.25)	7.91-8.96 (8.53)	5.0-9.2 (7.09)	6.9-9.6 (7.81)
Standard length / caudal peduncle depth[c]	7.62-9.96 (8.5)	6.88-9.77 (8.83)	7.88-8.97 (8.31)	6.7-11.6 (10.19)	7.8-11.7 (9.25)
Head length / snout length[c]	1.74-2.06 (1.89)	1.74-2.11 (1.94)	1.8-1.88 (1.85)	1.7-2.1 (1.89)	1.7-2.2 (1.97)
Head length / orbital diameter[c]	4.03-5.76 (4.81)	3.62-5.73 (4.78)	5-5.64 (5.23)	4.1-7.0 (5.79)	4.8-7.8 (6.05)
Head length / interorbital width	1.03-2.51 (2.24)	2.04-3.31 (2.54)	1.74-2.49 (2.18)	1.8-3.1 (2.55)	2.2-3.1 (2.53)
Head length / mouth width[c]	3.24-5.17 (4.07)	2.39-4.3 (3.05)	2.86-3.26 (3.12)	4.0-5.6 (4.68)	3.8-5.4 (4.80)
Caudal peduncle length / caudal peduncle depth[c]	0.64-1.43 (1.06)	0.77-1.29 (1.01)	0.88-1.13 (0.98)	1.0-1.8 (1.46)	0.8-1.5 (1.19)

[a]Samples were collected from the Gan (R13) and Min (R14) Rivers. [b]Samples were collected from the Chin (R12), Gan (R13), and Min (R14) Rivers. [c]The ratio from Mulan R. (R11) was similar to that of *F. lacustre* but differed from those of *F. stigmata* and/or *F. fascicauda* by *t*-test.

Acknowledgments: This research was funded by the National Science Council, Taiwan (NSC86-2311-B-002-028-B17 and NSC88-2311-B-002-039). We would like to thank Ms. Shu-Hua Hsu, Ms. Wan-Hui Lin, Ms. Yi-Ting Hsu, Mr. Te-Yu Liao, and Dr. Huei-Luen Huang for their assistance.

REFERENCES

Berg T, T Moum, S Johansen. 1995. Variable numbers of simple tandem repeats make birds of the order Ciconiiformes heteroplasmic in their mitochondrial genomes. Curr. Genet. **27:** 257-262.

Brown WM, EM Prager, A Wang, AC Wilson. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18:** 225-239.

Buroker NE, JR Brown, TA Gilbert, PJ O'Hara, AT Beckenbach, WK Thomas, MJ Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124:** 157-163.

Casane D, N Dennebouy, H de Rochambeau, JC Mounolou, M Monnerot. 1997. Nonneutral evolution of tandem repeats in the mitochondrial DNA control region of lagomorphs. Mol. Biol. Evol. **14:** 779-789.

Cecconi F, M Giorgi, P Mariottini. 1995. Unique features in the mitochondrial D-loop region of the European seabass *Dicentrarchus labrax*. Gene **160:** 149-155.

Chen AC, MC AnonuevoAblan, JW McManus, J DiepernkBell, VS Tuan, AS Cabanban, KT Shao. 2004. Variable numbers of tandem repeats (VNTRs), heteroplasmy, and sequence variation of the mitochondrial control region in the three-spot Dascyllus, *Dascyllus trimaculatus* (Perciformes: Pomacentridae). Zool. Stud. **43:** 803-812.

Clayton DA. 1984. Transcription of the mammalian mitochondrial genome. Annu. Rev. Biochem. **53:** 573-594.

Clement M, D Posada, KA Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:** 1657-1659.

Croteau DL, VA Bohr. 1997. Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. J. Biol. Chem. **272:** 25409-25412.

Fang PW. 1935. Study on the crossostomoid fishes of China. Sinensia **6:** 44-97.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:** 783-791.

Freeman AR, DE MacHugh, S McKeown, C Walzer, DJ McConnell, DG Bradley. 2001. Sequence variation in the mitochondrial DNA control region of wild African cheetahs (*Acinonyx jubatus*). Heredity **86:** 355-362.

Hudson RR, M Slatkin, WP Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics **132:** 583-589.

Innis MA, DH Gelfand, JJ Sninsky, TJ White. 1989. PCR protocols: a guide to methods and application. San Diego, CA: Academic Press.

Kocher TD, WK Thomas, A Meyer, SV Edwards, S Paabo, FX Villablanca, AC Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86:** 6196-6200.

Kumar S, K Tamura, IB Jakobsen, M Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Bioinformatics **17:** 1244-1245.

Lee WJ, J Conroy, WH Howell, TD Kocher. 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. **41:** 54-66.

Liu MY. 2006. Molecular systematics of the Genus *Macrobrachium* with notes on the phylogeography and population genetics of *M. asperulum* in Taiwan. PhD dissertation, Department of Life Science, National Tsing Hua Univ., Hsinchu, Taiwan.

Liu MY, YX Cai, CS Tzeng. 2007. Molecular systematics of the freshwater prawn genus *Macrobrachium* Bate, 1868 (Crustacea: Decapoda: Palaemonidae) from mtDNA sequences, with emphasis on East Asian species. Zool. Stud. **46:** 272-289

Lowe TM, SR Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic. Acids Res. **25:** 955-964.

Ludwig A, B May, L Debus, I Jenneckens. 2000. Heteroplasmy in the mtDNA control region of sturgeon (*Acipenser*, *Huso* and *Scaphirhynchus*). Genetics **156:** 1933-1947.

Lunt DH, LE Whipple, BC Hyman. 1998. Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. Mol. Ecol. **7:** 1441-1455.

Macey JR, JA Schulte, A Larson, TJ Papenfuss. 1998. Tandem duplication via light-strand synthesis may provide a precursor for mitochondrial genomic rearrangement. Mol. Biol. Evol. **15:** 71-75.

Mindell DP, MD Sorenson, DE Dimcheff. 1998. Multiple independent origins of mitochondrial gene order in birds. Proc. Natl. Acad. Sci. USA **95:** 10693-10697.

Nei M. 1987. Molecular Evolutionary Genetics. New York: Columbia Univ. Press.

Nesbo CL, MO Arab, KS Jakobsen. 1998. Heteroplasmy, length and sequence variation in the mtDNA control regions of three percid fish species (*Perca fluviatilis*, *Acerina cernua*, *Stizostedion lucioperca*). Genetics **148:** 1907-1919.

Novaek J, L Hanel, O Rican. 2005. *Formosania*: a replacement name for *Crossostoma* Sauvage, 1878 (Teleostei), a junior homonym of *Crossostoma* Morris and Lycett, 1851 (Gastropoda). Cybium **30:** 92.

Posada D, KA Crandall, AR Templeton. 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol. Ecol. **9:** 487-488.

Posada D, AR Templeton. 2005. Inference key for the nested haplotype tree analysis of geographical distances. Available at http://darwin.uvigo.es/download/geodisKey_11Nov05.pdf. 11, November, 2005.

Quinn TW, AC Wilson. 1993. Sequence evolution in and around the mitochondrial control region in birds. J. Mol. Evol. **37:** 417-425.

Rozas J, JC Sánchez-DelBarrio, X Messegyer, R Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

Saitoh K, K Hayashizaki, Y Yokoyama, T Asahida, H Toyohara, Y Yamashita. 2000. Complete nucleotide sequence of Japanese flounder (*Paralichthys olivaceus*) mitochondrial genome: structural properties and cue for resolving teleostean relationships. J. Hered. **91:** 271-278.

Simon HU, A Haj-Yehia, F Levi-Schaffer. 2000. Role of reactive oxygen species (ROS) in apoptosis induction. Apoptosis **5:** 415-418.

Sipe TW, RA Browne. 2003. Intra-specific phylogeography of the masked shrew (*Sorex cinereus*) and smoky shrew (*S. fumeus*) in the southern Appalachians. J. Mammal. **84:**

161-175.
Tamura K, M Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. **10:** 512-526.
Tang WQ, YY Chen. 2000. Study on taxonomy of Homalopteridae. J. Shanghai Fishery Univ. **9:** 1-10.
Taylor JS, F Breden. 2000. Slipped-strand mispairing at noncontiguous repeats in *Poecilia reticulata*: a model for minisatellite birth. Genetics **155:** 1313-1320.
Thompson JD, TJ Gibson, F Plewniak, F Jeanmougin, DG Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic. Acids Res. **25:** 4876-4882.
Tzeng CS, YS Lin, SM Lin, TY Wang, FY Wang. 2006. The phylogeography and population demographics of freshwater fishes in Taiwan. Zool. Stud. **45:** 285-297.
Tzeng CS, SC Shen, PC Huang. 1990. Mitochondrial DNA identity of *Crossostoma* (Homalopteridae, Pisces) from two river systems of the same geographical origin. Bull. Inst. Zool. Acad. Sinica **29:** 11-19.
Wang TY. 2004. Taxonomy, evolution and phylogeography of Balitoridae in East Asia. PhD dissertation, Department of Life Science, National Tsing Hua Univ., Hsinchu, Taiwan.
Wang TY, TY Liao, CS Tzeng. 2007. Phylogeography of the Taiwanese endemic hillstream loaches, *Hemimyzon formosanus* and *H. taitungensis* (Cypriniformes: Balitoridae). Zool. Stud. **46:** 547-560.
Wang TY, CS Tzeng, SC Shen. 1999. Conservation and phylogeography of Taiwan paradise fishes, *Macropodus operculari*s Linnaeus. Acta Zool. Taiwanica **10:** 121-134.
Wilkinson GS, F Mayer, G Kerth, B Petri. 1997. Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics **146:** 1035-1048.
Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic. Acids Res. **31:** 3406-3415.