Pl Syst Evol 269: 145–157 (2007)
DOI 10.1007/s00606-007-0592-4
Printed in The Netherlands

**Plant Systematics and Evolution**

# Single nucleotide polymorphisms of *Gcyc1* (*Cycloidea*) in *Conandron ramondioides* (Gesneriaceae) from Southeast China

**L.-H. Xiao,[1,2] Y.-Z. Wang[1]**

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, Beijing, P.R. China
[2]Department of Ecological and Environmental Science, College of Life Science, Inner Mongolia University, Hohhot, Inner Mongolia, P.R. China

Received 4 August 2006; Accepted 26 June 2007; Published online 26 October 2007


**Abstract.** *Conandron ramondioides* with actinomorphic flower in Gesneriaceae is an endemic species distributed in Taiwan, Southeast of China and Japan. Populations are usually small and isolated in typically fragmented habitat. Based on SNPs of *Gcyc1* (*Cycloidea*), a TCP gene known in patterning the floral dorsoventral asymmetry, we have explored the molecular evolution and genetic differentiation of *Gcyc1* at population level, and the population history of *C. ramondioides* populations distributed in SE China. Eighteen SNPs are detected in 774-bp of the gene, of which eleven are non-synonymous. However, morphological observation of flowers shows that there is no visible differentiation in shape and size across the dorsoventral axis within each whorl. None of the eighteen SNPs is by all shared the eleven populations. Population differentiation is significant. These results reveal that evolution of *Gcyc1* at population level is well in accord with the neutral theory. Our study indicates that the SNPs of developmental genes are also useful molecular markers for exploring the genetic differentiation and population history in non-model organisms.

**Keywords:** *Conandron ramondioides* (Gesneriaceae) *Gcyc1* gene; Habitat fragmentation; Non-model organism; Population history; SNP

## Introduction

Single nucleotide polymorphisms (SNPs), i.e. single-base substitutions in the DNA sequence, are the most common type of sequence differences between alleles, and extensively exist in genomes of all organisms (Brookes 1999, Nasu et al. 2002, Rafalski 2002). Most SNPs lie in the non-coding regions of genes owing to be constrained by gene functions, but the SNPs in the coding regions of genes (cSNPs) are more important because they may alter the structure and function of the encoded proteins (Syvänen 2001). Since they encode amino acids cSNPs can be devided into synonymous and non-synonymous cSNPs, based on coding amino acids. Most recently, SNPs, as the third generation of

Correspondence: Yin-Zheng Wang, State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, The Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, P.R. China
e-mail: wangyz@ibcas.ac.cn

molecular markers following RFLP and SSR, have gradually become a welcome tool for studying linkage disequilibrium (LD), integrating genetic map and physical maps and elucidating the evolutionary history of populations for model organisms (Brumfield et al. 2003, Han and Xue 2003, Rafalski 2002, Tenaillon et al. 2001, Zhu et al. 2003). Compared with other genetic markers, SNPs are more abundant in the genome and much more stably inherited (Osman et al. 2003, Marshall 1997, Weber and Wong 1993). Furthermore, SNPs can be used for association studies because they have a low mutation rate and are amenable to automation (Halushka et al. 1999). Just as changes at the DNA level allowing populations to diverge and ultimately to form new species, SNPs may also provide fundamental novel insights into the evolution of species (Barraclough 2001, Picó et al. 2002). SNPs of known genes have recently been used to explore the population history of non-model species in plant, such as in *Ramonda* (Picó et al. 2002). Therefore, it will be interesting to use SNPs of known genes as marker to carry out an evolutionary biological and genetic research at population level in non-model species, such as *Conandron ramondioides*.

*Cycloidea* (*cyc*), a functional gene involved in establishing floral zygomorphy together with the *Dichotoma* (*dich*), was first isolated from *Antirrhinum majus* (Veronicaceae) (Luo et al. 1996, 1999, Almeida et al. 1997). This gene encodes a putative transcription factor, which controls the floral zygomorphy by regulating the expression of the *Cyclin* genes (Cubas et al. 1999b). Homologues of *cyc* have been isolated in several other families (Cubas et al. 1999a; Möller 1999; Citerne et al. 2000, 2003; Fukuda et al. 2003, Wang et al. 2004a), including Gesneriaceae that closely related to Veronicaceae *Antirrhinum* belongs to *Gcyc*, a homologue of *cyc* in Gesneriaceae, usually has one or two copies in one species (Möller 1999, Citerne et al. 2000, Wang et al. 2004a). Both the highly conserved regions (e.g. the TCP-domain and R-domain) and the highly variable regions, like *cyc* of *Antirrhinum*, were also presented in the *Gcyc* (Citerne et al. 2000). On account of rapid evolution of the highly variable regions, *Gcyc* was applied for phylogenetic analysis to investigate the possible evolutionary relationship among taxa of this family and the relationship between divergence of *Gcyc* sequences and changes in floral morphology in this family (Möller 1999, Wang et al. 2004b). The results of phylogenetic analysis showed that *Gcyc* may fall into two classes (*Gcyc1* and *Gcyc2*), as a result of an early gene duplication from an ancient polyploidy event (Wang et al. 2004a). *Gcyc1* seems to evolve more quickly than *Gcyc2* and several independent duplication events had occurred after splitting between *Gcyc1* and *Gcyc2* (Wang et al. 2004a). Since SNPs of *Gcyc2* have been successfully used as a molecular marker to explore the population history of a species in Gesneriaceae, the SNPs of the *Gcyc1* may be an ideal molecular marker to study the genetic differentiation and the population history of a species in Gesneriaceae because it evolves more quickly than *Gcyc2* in Gesneriaceare (Wang et al. 2004a).

*Conandron* with antinomorphic flowers is a monotypic genus in the Old World Gesneriaceae (subfamily Cyrtandroideae), and is geographically restricted Taiwan, Southeast of mainland in China and Japan. The only species *Conandron ramondioides* Sieb. et Zucc. is a perennial herb with only one leaf (rare two or three ones) and is usually distributed in the middle and low mountain areas (Fig. 1). Populations of *C. ramondioides* usually grow in specific fragmented habitat

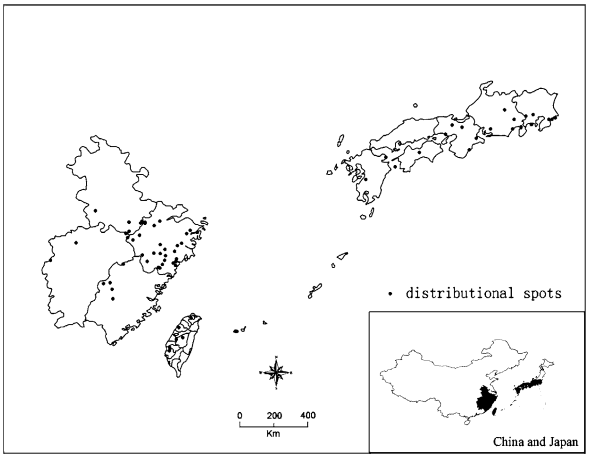


**Fig. 1.** Geographic distribution of *C. ramondioides*

along creeks, particularly on wet, dripping granite cliffs (personal observation). Populations are often small and isolated each other in long distance. There is no noticeable difference in the subtropical monsoon climatic conditions over its whole distribution area. As outlined above, the species *C. ramondioides* is an ideal candidate for the research in population biology, including genetic differentiation and population history of this species, especially using SNPs of a known gene as a molecular marker.

To understand the genetic differentiation and evolutionary history of *C. ramondioides* populations as well as the molecular evolution of *Gcyc1* within this species,we isolated the full length of ORF (open reading frame) of *Gcyc1* and investigated its variations at population level. Integrating the SNPs of *Gcyc1* and paleoclimatic and paleogeological changes of distribution area of *C. ramondioides*, we inferred the evolutionary processes of populations and re-colonization patterns of the relict *C. ramondioides* during pre-glacial and glacial episodes. In addition, we also discussed the application of SNP markers of coding regions of genes in inferring population genetic differentiation and population history among non-model species. Our results give an insight into population evolutionary history in *C. ramondioides*, and provide beneficial suggestion for the conservation of endemic and rare species, like *C. ramondioides*, in Gesneriaceae.

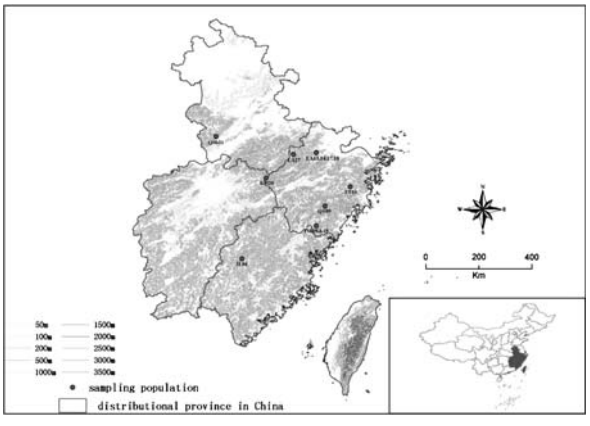


**Fig. 2.** Location and topography of investigated populations

## Materials and methods

**Sampling, material collection and morphological observation at population level.** During the summer and autumn of 2003 and 2004, eleven populations were investigated in *C. ramondioides* from SE China. DNA materials were collected during the field survey (Fig. 2). In addition to the Qianshan population (QSh21) where only five individuals were sampled, leaf samples from at least sixteen randomly chosen individuals per population were collected, reflecting both the size of these populations and the extent of clonal growth. In fact, we have avoided clonal sampling as far as possible in samples gathering process. Moreover, some living plants, rhizomes and seeds were brought back to IBCAS (Institute of Botany, the Chinese Academy of Sciences) for future studies. Voucher specimens were deposited in the China National Herbarium (PE). Living plants are cultured in glasshouse of IBCAS. The collection details materials are shown in Table 1. Morphological observation was carried out using dissecting microscope.

**DNA extraction, primers, amplification and sequencing.** Genomic total DNA from a total 219 individuals was extracted using a modified CTAB method (Doyle and Doyle 1987) from silica gel dried leaves, and purified using the Genomic DNA Rapid Purification Kit (BioDev Tech. Inc.). In this study, we chose *Gcyc1*, due to its rapid evolution, as the target fragment. According to the sequence of *cyc* in *Antirrhinum majus* (accession no.Y16313), *Lcyc* in *Linaria vulgaris* (accession no. AF161252) from GenBank and *Saintpaulia velutina cyc1A* (Wang et al. 2006), we designed the forward degenerate primer, JC (5′-GTT TGG SAA GAA CWC RTA CC-3′). According to the 70% complete sequence of *Gcyc1* in *C. ramondioides* from GenBank (accession no. AF20839 and AF208321), the specific reverse primer GcycR1 (5′-CAT TGA CAT TAA GAG ATG GGA G-3′) were designed. Then JC/GcycR1 were used for amplifying and sequencing to obtain the 5′-teminal sequence of encoding region of *Gcyc1* gene in *C. ramondioides*. After that, we also designed the forward primer GC1 (5′-ATG TTT GGC AAG AAC TCG TAC C-3′) as well as the sequencing primers SPF (5′-AGC AAG ACA TGC TTT CTG G-3′) and SPR (5′-GAC ATT GGA TTC AAT ATG GTG-3′) (Fig. 3). Then we amplified and sequenced the total 219 samples from eleven populations in *C. ramondioides*, using the PCR primers GC1/GcycR (5′-ATg AAT TTG TGC TGA TCC AAA ATG-3′)

**Table 1.** Detailed sources of materials used in this study

| Pop. | Source | Elevation (m) | Pop. size | No. of samples | Collected time |
|---|---|---|---|---|---|
| QSh21 | Qianshan, Anhui | 940 | 6 in 2 $m^2$ | 5 | July, 2004 |
| LAL13 | Longtangshan, Lin'an, Zhejiang | 730 | 42 in 4 $m^2$ | 17 | July, 2004 |
| LAL14 | Longtangshan, Lin'an, Zhejiang | 720 | 24 in 3 $m^2$ | 19 | July, 2004 |
| LAN17 | Niulangping, Lin'an, Zhejiang | 1140 | 120–160 in 8 $m^2$ | 32 | July, 2004 |
| LAD18 | Dong'aotou, Lin'an, Zhejiang | 960 | 170 in 10 $m^2$ | 23 | July, 2004 |
| LASh27 | Shunxizhiyuan, Lin'an, Zhejiang | 510 | 40 in 5 $m^2$ | 16 | October, 2003 |
| KH20 | Kaihua, Zhejiang | 730 | 80–100 in 10 $m^2$ | 22 | July, 2004 |
| TT10 | Tiantai, Zhejiang | 760 | 40 in 4 $m^2$ | 17 | July, 2004 |
| QT09 | Qingtian, Zhejiang | 780 | 140 in 4 $m^2$ | 19 | July, 2004 |
| TSh08 & 39 | Taishun, Zhejiang | 740 | 80 in 8 $m^2$ | 27 | June, 2004 & October, 2003 |
| JL04 | Jiangle, Fujian | 940 | 253 in 11 $m^2$ | 22 | June, 2004 |

and sequencing primers SPF/SPR (Fig. 3). For each sample, the *Gcyc1* gene sequence was sequenced from both forward and reverse directions. We identified the heterozygotes by double peaks of the same intensity from the electropherograms. For those samples with double peak, resequencing was carried out in order to confirm heterozygous nucleotides.

The reaction system of polymerase chain reaction (PCR) was as following and made up to 25 μl with sterile distilled water: 0.15 μl (5 U/μl) *Taq* DNA polymerase (TaKaRa), 2.5 μl 10 × PCR buffer (TaKaRa), 2.5 μl dNTP at 2 mM (final concentration 0.2 mM) (TaKaRa), 1.0 μl $MgCl_2$ at 25 mM (final concentration 2.5 mM) (TaKaRa), 2.5 μl each primers at 2 μM (0.2 μM) (TaKaRa) and about less 50 ng template DNA. And PCR cycling parameters were as follows: 1 cycle of 5 min at 94°C; 32 cycle of 1 min at 94°C, 1 min at 53°C, and at 1 min 30 s 72°C, with a final extension step for 10 min at 72°C. PCR products were separated by 1.5% (w/v) agarose gel electrophoresis. The band with the right size was cut out and purified using the GFX PCR DNA and Gel Band Purification Kit (Pharmacia) and then the sequencing reactions carried out using the sequencing primers. The products were applied to an ABI $3730_{XL}$ automatic sequencer. All sequences of *Gcyc1* in the eleven populations of *C. ramondioides* are deposited in GenBank (accession nos. EU036762-036980).

**Methods in SNP identification and genetic analysis.** For each individual, the sequences were corrected with Contig Express software. Double peaks were identified as SNPs when it was half of the single peak on the electropherograms. Then sequence alignments were made with Clustal X and refined manually. The different bases between sequences were all identified as SNPs. Each identified SNP we considered as one information site. They were categorized as homozygous (minor and major) and

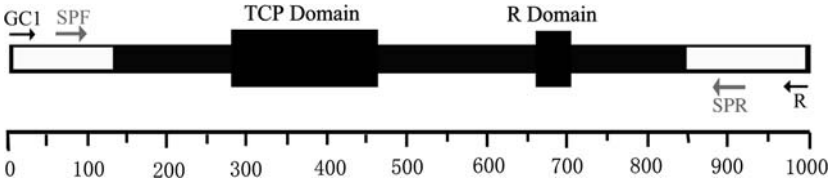


**Fig. 3.** Schematic strcture and map of the *Gcyc1* locus

heterozygous sites. The results were entered in a Microsoft Excel worksheet. Disregarding the linkage of sites, the genetic diversity and cluster analysis was constituted with the program POPGENE32 (v.1.31). Percentage of polymorphic sites and mean heterozygosity between populations were estimated using POPGENE32. The Ht at the polymorphic sites and the Gst were determined followed the G-statistic of Nei (1978) with the assistance of the FSTAT computer program (v.2.9.3.2). Nei's unbiased distance (D) was estimated between populations to generate average linkage clustering using the UPGMA method. A phenogram of Nei's genetic distance (1978) were prepared using MEGA (version3.0 (According to the method of Almeida (2003)).

## Results

**Field observation.** In this study, eleven populations of *C. ramondioides* were investigated (Table 1). All the populations are distributed in valleys of shady mountain slopes and are usually found on granite rock at the headstream, rooted in moss-covered rock clefts with shallow soil under subtropical evergreen broadleaved forest. Eleven investigated and sampled populations are all distributed in such similar fragmented primary habitats (Table 1). The size of each population is very small and isolated from each other by long distance. The individuals in each population are usually clumped together. The number of individuals is correlative with the moisture degree of the habitat, that is, more individuals in moister habitats. Two reproductive patterns, rhizomes and seed reproduction, were observed both in the field and glasshouse. However, rhizomes grow very slowly and seeds are the major source of reproduction. Seeds can be diffused by the force of dehiscence of the capsular fruit. There are many tiny seeds in a capsule. The seeds escape from the fruit usually spread by water. No visible differentiation in floral symmetry has been observed among and within populations in *C. ramondioides*.

**Sequence analysis and SNP identification.** A total of eighteen SNPs were detected in the 774 sequenced bases of *Gcyc1* (Table 2). The SNP frequency is one per 43 bp. Bases near primers at both ends of the sequence were not included in the analysis because of uncertainty. The sequences of *Gcyc1* obtained from eleven *C. ramondioides* populations were identical in length. Each SNP included two bases and the numbers of SNPs were different among populations, the results were summarized in Table 2. There were eleven non-synonymous substitutions (not correlated with the population number) and seven synonymous substitutions in all eighteen SNPs. SNP-1 to SNP-7 were detected in the variable region from 5′-terminal to TCP-domain. SNP-8 (AAA/AAg), SNP-9 (TCg/TCT) and SNP-10 (GAT/GAA) present in basic region of TCP domain, in which the SNP-8 and SNP-9 were synonymous and SNP-10 were non-synonymous. Their amino acid changes were K/K, K/K, and D/E, respectively. But these nucleotide substitutions in basic region had not affected the secondary structure of the putative protein. SNP-11 to SNP-13 were located in the region between two conserved regions, TCP and R domains and SNP-14 to SNP-18 after R-domain.

In all the sites detected, polymorphy varied from 5.56% to 38.89% with average of 18.28% within populations (Table 2). The highest frequency of SNPs was found in KH20 and TSh08 populations followed by population QT09 with the number of SNPs of 7, 7 and 6, respectively, while only one SNP was detected in LA18 population. None of eighteen SNPs was shared by all eleven populations (Table 2). However, in JL04 population, one of three detected SNPs was only observed in a local population and other two were fixed in whole JL04 population relative to other populations. Furthermore, one SNP was fixed in whole LA17 and LA18 populations, respectively.

**Genetic diversity.** Heterozygotes were detected in ten of eighteen SNPs (Table 2). The observed heterozygosity was generally lower than expected heterozygosity in ten of eleven populations, indicating an excess of homozygotes in these populations. However, there is no noticeable difference in LA17 population and the He is a little lower than Ho only (Table 5).

**Table 2.** Distribution of SNPs in the 774 bp sequenced of *Gcyc1* gene found in *C. ramondioides* samples from 11 populations. The exact position of the SNP in the r is indicated. Heterozygotes are indicated by K (G/T), R (A/G), S (C/G), W (A/T), Y (C/T) and were detected by double peak in the electropherograms. Consensus sequence is indicated by Con. Syn and Nsyn represented synonymous and non-synonymous substitution, respectively

| Population | SNP and its base position | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 (5) | 2 (18) | 3 (23) | 4 (41) | 5 (78) | 6 (99) | 7 (129) | 8 (222) | 9 (252) | 10 (267) | 11 (404) | 12 (416) | 13 (448) | 14 (507) | 15 (579) | 16 (583) | 17 (733) | 18 (746) | N | Percentage |
| QSh21 | T | A/G | C | T | C | C | T | A | T | T | G | C | T | T | A | C | C | C/S | 2 | 11.11% |
| LAL13 | T | G | C | T | C | A/C | T | A | T | T | G | C/G | T | T | A | C | C | C | 2 | 11.11% |
| LAL14 | T | G | C | T | C | A/C | T | A | T | T | G | C/G/S | T | T | A | C/G/S | C | C | 3 | 11.67% |
| LAN17 | T | G/R | C | T | C | C | T | A | T | T | G | G | T | T | A | C | C | C | 1(2) | 11.11% |
| LAD18 | T | G | C | T | C | C | T | A | T | T | G | G | T | T | A | C | C | C | 1 | 5.56% |
| LASh27 | C/T | G | C | T | C | A/C | T | A | T | T | G | C/G/S | T | T | A | C | C | C | 3 | 16.67% |
| KH20 | C/T | G | C | T | C/G | C | G/T | A/G | T | T | G | C/G/S | T | T/C/Y | A | C/S | C | C | 7 | 38.89% |
| TT10 | T | G | C | T | C | C | T | A | T | T | G | C/G | T | T | A | C/S | C | C | 2 | 11.11% |
| QT09 | T | G | A/C | T | C | C | T | A | G/T | T | A/G/R | C/G | T | T | A/G | C/G/S | C | C | 6 | 33.33% |
| TSh08 & 39 | T | G | C | C/T | C | C | T | A | G/T/K | T | A/G/R | C/G/S | T/W | T | A | C/G/S | C/S | C | 7 | 38.89% |
| JL04 | T | G | C | C | C | C | T | A | T | A/T/W | G | C | T | T | A | G | C | C | 1 (3) | 11.67% |
| CON. | T | G | C | T | C | C | T | A | T | T | G | C | T | T | A | C | C | C | 18 (Total) | 18.28% (Ave.) |
| AA. SUB. | L-P | G-G | T-N | I-T | H-Q | G-G | L-L | K-K | S-S | D-E | R-K | S-C | C-S | D-D | S-S | L-N | R-G | T-S | | |
| SUB. | Nsyn | Syn | Nsyn | Nsyn | Nsyn | Syn | Syn | Syn | Syn | Nsyn | Nsyn | Nsyn | Nsyn | Syn | Syn | Nsyn | Nsyn | Nsyn | | Nsyn/Syn = 11/7 |

**Table 3.** Nei's Unbiased measures of genetic identity and genetic distance. Estimates of mean genetic identity (above diagonal) and genetic distance (below diagonal) between 11 populations of *C. ramondioides*

| Pop. | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. QSh21 | – | | | | | | | | | | |
| 2. LAL13 | 0.0957 | – | | | | | | | | | |
| 3. LAL14 | 0.0951 | 0.0007 | – | | | | | | | | |
| 4. LAN17 | 0.0991 | 0.0003 | 0.0013 | – | | | | | | | |
| 5. LAD18 | 0.1019 | 0.0003 | 0.0012 | 0.0000 | – | | | | | | |
| 6. LASh27 | 0.0816 | 0.0516 | 0.0519 | 0.0585 | 0.0583 | – | | | | | |
| 7. KH20 | 0.0756 | 0.0716 | 0.0705 | 0.0773 | 0.077 | 0.0131 | – | | | | |
| 8. TT10 | 0.0414 | 0.0448 | 0.0435 | 0.0508 | 0.0507 | 0.0339 | 0.0293 | – | | | |
| 9. QT09 | 0.0485 | 0.0521 | 0.0502 | 0.0582 | 0.058 | 0.0416 | 0.0365 | 0.0047 | – | | |
| 10. TSh08 & 39 | 0.0656 | 0.0662 | 0.0600 | 0.0722 | 0.0719 | 0.0583 | 0.0522 | 0.0183 | 0.0086 | – | |
| 11. JL04 | 0.1715 | 0.1805 | 0.1627 | 0.1872 | 0.1868 | 0.168 | 0.1556 | 0.1185 | 0.1215 | 0.0757 | – |

Mean expected heterozygosity was higher than observed with an average of He = 0.0952 (±0.1300), Ho = 0.0089 (±0.0118), respectively. The TSh08 & 39 populations had the highest expected heterozygosity, i.e. He = 1.076 (±0.1708), whereas the lowest heterozygosity was from the LA18, i.e. He = 0, because only one SNP was detected and had been fixed in whole population (Table 3).

Total genetic diversity (Ht) and gene diversity within populations (Hs) at polymorphic sites averaged 0.097 and 0.042, respectively (Table 4). The mean genetic differentiation between populations (Gst) was 0.564 (Table 4), indicating that about 56.4% of the observed genetic diversity was due to variation between *C. ramondioides* populations and the remaining 43.6% was a function of genetic differentiation among plants within populations. The gene flow and fixation indices are 0.1787 and 0.5832, respectively (Table 6).

**Genetic distance.** The average genetic distance (D) between populations was D = 0.068, with a varied range from 0.0003 to 0.1868 (Table 3). UPGMA analysis between populations using Nei's unbiased genetic distance showed that the JL04 population (from Fujian province) had the farthest genetic distance with other populations, followed by the QSh21 population from Anhui province. In the populations from Zhejiang province, the TSh08 & 39, QT09 and TT10 populations were clustered into one group, and the LA13, LA14, LA17 and LA18 populations, from Longtang mountain in Lin'an in Zhejing province were gathered into another group. Interestingly, the LA27 population from Shunxi in Lin'an and the KH20 population clustered together (Fig. 4).

## Discussion

**Molecular evolution of *Gcyc1*.** Some studies show that *Gcyc*, especially its variable regions, diverges and evolves rapidly among species of Gesneriaceae (Möller 1999, Citerne et al. 2000, Wang et al. 2004a). Based on the SNPs of *Gcyc2* at population level and the morphological traits of flower in *R. myconii*, Picó et al. (2002) suggest that the evolution of the *Gcyc2* gene within species is in accord with the neutral theory. However, the *Gcyc1* gene evolves more rapidly than other known copies of *Gcyc* (Möller 1999, Citerne et al. 2000, Wang et al. 2004a) and has more SNPs than *Gcyc2* (data not shown) within species. The *Gcyc1* has two conserved domains, i.e. a basic-helixI-loop-helixII (BHLH) TCP domain and an Arg-rich R domain (Cubas et al. 1999b). According to Cubas et al. (1999b), the residues in the basic region are the most conserved. A bipartite nuclear localization signal is in the basic region of the Cyc/TB1 subfamily (Luo et al. 1996, Doebley et al. 1997). This region is essential for specific binding to

**Table 4.** G-statistic calculated from 18 polymorphic loci over eleven populations in *C. ramondioides.* Ho, proportion of heterozygotes. Hs, within population gene diversity. Ht, total genetic diversity. Dst, average gene diversity among populations. Gst, mean genetic differentiation between populations

| Site | Ho | Hs | Ht | Dst | Gst |
|---|---|---|---|---|---|
| SNP1 | 0.000 | 0.088 | 0.210 | 0.122 | 0.583 |
| SNP2 | 0.006 | 0.037 | 0.14 | 0.103 | 0.738 |
| SNP3 | 0.000 | 0.018 | 0.019 | 0.001 | 0.042 |
| SNP4 | 0.000 | 0.034 | 0.198 | 0.164 | 0.83 |
| SNP5 | 0.000 | 0.016 | 0.016 | 0.000 | 0.028 |
| SNP6 | 0.000 | 0.066 | 0.085 | 0.019 | 0.224 |
| SNP7 | 0.000 | 0.023 | 0.025 | 0.002 | 0.072 |
| SNP8 | 0.000 | 0.023 | 0.025 | 0.002 | 0.072 |
| SNP 9 | 0.003 | 0.061 | 0.070 | 0.009 | 0.129 |
| SNP10 | 0.021 | 0.031 | 0.037 | 0.006 | 0.157 |
| SNP11 | 0.022 | 0.079 | 0.102 | 0.023 | 0.227 |
| SNP12 | 0.038 | 0.118 | 0.483 | 0.365 | 0.755 |
| SNP13 | 0.007 | 0.007 | 0.007 | 0.000 | 0.006 |
| SNP14 | 0.004 | 0.012 | 0.012 | 0.000 | 0.016 |
| SNP15 | 0.000 | 0.010 | 0.010 | 0.000 | −0.010 |
| SNP16 | 0.027 | 0.094 | 0.254 | 0.160 | 0.631 |
| SNP17 | 0.010 | 0.010 | 0.010 | 0.000 | 0.024 |
| SNP18 | 0.018 | 0.040 | 0.053 | 0.013 | 0.246 |
| Overall | 0.009 | 0.042 | 0.097 | 0.055 | 0.564 |

promoter elements of the proliferating cell nuclear antigen (PCNA) genes (Kosugi and Ohashi 1997). Therefore, it is very possible that non-synonymous substitution of this region may affect the function of the putative protein and ultimately result in the change of phenotype.

In our results, eleven of eighteen SNPs (one SNP per 43 bp) detected in 774 bp of *Gcyc1* sequence are non-synonymous substitutions at population level in *C. ramondioides* (Table 2). Although one non-synonymous substitution (SNP-10, GAT→GAA or D→E) takes place in the basic region of TCP domain of *Gcyc1*, there is no change in the carried charge and the hydrophobicity of the two amino acids. Similar results can be drawn from other ten non-synonymous substitution in the highly variable regions of the gene (Table 2). Therefore, the secondary structure of putative protein seems not to be affected by these non-synonymous substitutions. Morphological observation of flower also shows that there is no noticeable differentiation in shape and size across the dorsoventral axis within each whorl. These facts allow us to suggest that the SNPs presented in the coding region of *Gcyc1* have no effect on this gene function in controlling floral symmetry in *C. ramondioides.* These results implied that the variations in DNA sequence of *Gcyc1* gene has a neutral evolutionary pattern at population level. Therefore, SNPs of the coding region of *Gcyc1* gene is an ideal molecular marker for reconstructing the evolutionary history of *C. ramondioides* populations.

**Genetic differentiation integrating morphologic traits and habitat features.** Our results show that in general, the level of genetic diversity of populations in *C. ramondioides* (P = 18.28%, He = 0.0952 ± 0.13) is lower compared with a tropical long-lived tree *Eurycoma longifolia* (He = 0.226 ± 0.029) reported from SNPs marker (Osman et al. 2003). The mean genetic differentiation between populations (Gst) is 0.564 in *C. ramondioides*, indicating that there are more variations among populations (56.4%) than

**Table 5.** Genetic diversity in *Conandron ramondioides* populations. Ho, observed heterozygosity. He, expected heterozygosity. Values in parentheses are standard deviation

| Pop. | No. of samples | Ho | He |
|---|---|---|---|
| QSh21 | 5 | 0.0111 (0.0471) | 0.0411 (0.1209) |
| LAL13 | 17 | 0.0000 (0.0000) | 0.0123 (0.0358) |
| LAL14 | 19 | 0.0058 (0.0170) | 0.0263 (0.0646) |
| LAN17 | 32 | 0.0035 (0.0147) | 0.0034 (0.0143) |
| LAD18 | 23 | 0.0000 (0.0000) | 0.0000 (0.0000) |
| LASh27 | 16 | 0.0139 (0.0589) | 0.0760 (0.1749) |
| KH20 | 22 | 0.0101 (0.0249) | 0.0825 (0.1256) |
| TT10 | 17 | 0.0033 (0.0139) | 0.0093 (0.0287) |
| QT09 | 19 | 0.0058 (0.0170) | 0.0645 (0.1093) |
| TSh08 & 39 | 27 | 0.0288 (0.0517) | 0.1076 (0.1708) |
| JL04 | 22 | 0.0126 (0.0536) | 0.0181 (0.0767) |
| Overall | 219 | 0.0089 (0.0118) | 0.0950 (0.1297) |

within population (43.6%) (Table 4). This shows that there is higher genetic differentiation among populations than within populations in *C. ramondioides*. Isolation of populations usually decreases gene flow and increases genetic differentation among populations as well as loss of the genetic variability within populations (Templeton et al. 1990, Hanski 1991, Fore and Guttman 1992, Husband and Barrett 1996, Young et al. 1996). In *C. ramondioides*, a very low gene flow is shown between populations (Nm = 0.1787) (Table 6), which is probably mainly due to the complete isolation of populations generated by long distances between populations that are usually small in size at isolated patches of habitats. As Wright (1931) suggested if the gene

**Table 6.** F-Statistics and Gene flow for all sites. Nm = Gene flow estimated from Fst = 0.25(1 − Fst)/Fst

| Site | Fis | Fit | Fst | Nm* |
|---|---|---|---|---|
| SNP1 | 1.0000 | 1.0000 | 0.6081 | 0.1611 |
| SNP2 | 0.8358 | 0.9593 | 0.7524 | 0.0823 |
| SNP3 | 1.0000 | 1.0000 | 0.0966 | 2.3375 |
| SNP4 | 1.0000 | 1.0000 | 0.8409 | 0.0473 |
| SNP5 | 1.0000 | 1.0000 | 0.0833 | 2.7500 |
| SNP6 | 1.0000 | 1.0000 | 0.2696 | 0.6773 |
| SNP7 | 1.0000 | 1.0000 | 0.1255 | 1.7417 |
| SNP8 | 1.0000 | 1.0000 | 0.1255 | 1.7417 |
| SNP9 | 0.9414 | 0.9518 | 0.1781 | 1.1534 |
| SNP10 | 0.3016 | 0.4339 | 0.1895 | 1.0694 |
| SNP11 | 0.7091 | 0.7866 | 0.2663 | 0.6887 |
| SNP12 | 0.6581 | 0.9205 | 0.7674 | 0.0758 |
| SNP13 | −0.0385 | −0.0034 | 0.0338 | 7.1500 |
| SNP14 | 0.6423 | 0.6646 | 0.0624 | 3.7583 |
| SNP15 | 1.0000 | 1.0000 | 0.0481 | 4.9500 |
| SNP16 | 0.7002 | 0.8952 | 0.6506 | 0.1343 |
| SNP17 | −0.0588 | −0.0051 | 0.0508 | 4.6750 |
| SNP18 | 0.5238 | 0.6573 | 0.2804 | 0.6417 |
| Mean | 0.7824 | 0.9093 | 0.5832 | 0.1787 |

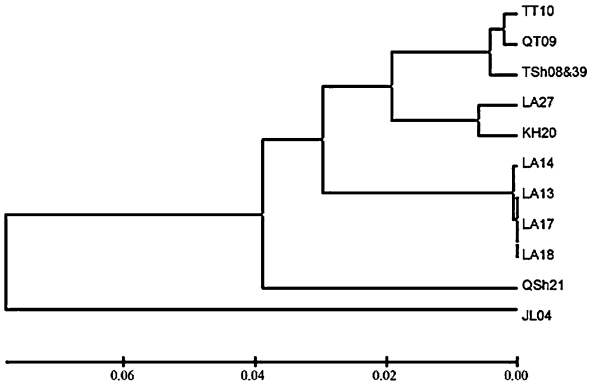


**Fig. 4.** Dendrogram based on Nei's (1978) genetic distance method = UPGMA. Modified from NEIGHBOR procedure of PHYLIP Version 3.5

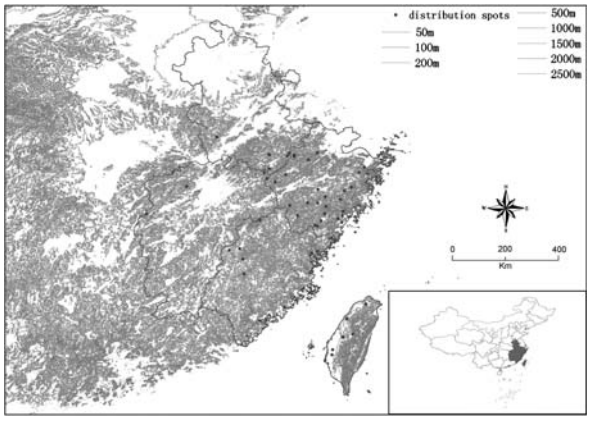


**Fig. 5.** Topography of distribution of *C. ramondioides* in SE China

flow is less than 1 (Nm < 1), the finite gene flow is the major reason causing the differentiation between populations.

For all eighteen SNPs, homozygotes outnumber heterozygotes. This homozygote excess would be expected if *C. ramondiodies* had an appreciable selfing rate. Hamrick and Godt (1990) suggest that there are some 51% genetic variations between populations in self-fertilizing plants (Gst = 0.510), equvilant to 5 times of that in bee-pollinated outbreeding plants. The value of Gst (0.564) in *C. ramondioides* implies that this species might have an appreciable selfing or inbreeding rate, as that reported from *R. myconi* (Picó et al. 2002) and *T. oldhamii* (Wang et al. 2004a), or that the vegetative reproduction by rhizome might become increasingly predominant over sexual reproduction. In addition, observation on greenhouse cultivated plants shows that flowers of *C. ramondiodies* are not capable of autopollination. Therefore, the high value of GST should be mainly due to the fragmentation of continuous habitats into smaller and more isolated patch resulting in a decrease in level of gene flow, particularly over longer distances. Further field investigations and studies for breeding system are necessary to reveal its reproductive features in link with the genetic differentiation of populations in *C. ramondiodies*.

**Population history of *C. ramondioides* in SE China.** Our results demonstrate that the level of *Gcyc1* polymorphism varies among populations and no SNP is shared by all eleven populations in *C. ramondioides*. High percentages of SNP are found in the populations, KH20 and TSh08 & 39 and QT09 (Table 2), while only one SNP is detected in the LA18 population. In addition, the pair of genetic distance of UPGMA cluster analysis between populations shows that the genetic distances among populations are not fully correlated with the geographic distances among eleven populations of *C. ramondioides* in SE China. Based on SNPs of *Gcyc2* in two populations of *R. myconii*, Picó et al. (2002) suggest that the effect of glaciations are very different in extant populations of the pre-glacial relict, and the genetic variation could have been lost due to either dramatic reductions in population size and subsequent genetic drift or a recolonization process from south to north during interglacial periods. According to Wang et al. (2004b), the distribution of *T. oldhamii* seems to be consistent with former land connection between Taiwan and mainland of China, and between continental of China and the Ryukyu islands at a glacial maximum during the Quaternary, followed by progressive fragmentation of the populations. In *C. ramondioides*, the distribution area might have originally been continuous and wide-ranging gene exchange would have been existed. However, the continuous distribution area could have then fragmented into small and isolated populations

because of the climate change during glacial periods, such as temperature, air humidity and aridity change, and human activity, such as deforestation and crop planting. The affect of glaciations and human activity could have been very different throughout the species' distribution area, especially in SE China.

The present distribution of *C. ramondioides* in the Huanglong Mountain in Xiushui county of Jiangxi province provides a good case for explaining the population fragment affected by climatic changes and human activities. According to specimen documents, populations of *C. ramondioides* existed in the Huanglong Mountain area in the past (Fig. 5). But now they cannot be found there. Loss of the habitat may have to the extinction of most individuals in this aera owing to climate change and human activity. Our field survey shows that about 90% of this area has been covered with man-made conifer forests and the habitats have become more and now than before. Another case is the QSh21 populations distributed in the Tianzhu Mountain in Qianshan county of Anhui province, which is completely isolated from other extant populations by a plain with cultivated crops (Fig. 5). It is very possible that plants of *C. ramondioides* originally distributed in the lowland below 500 m might have become completely extincted because of human agricultural activities.

**Potential implication of SNPs marker on population history of non-model organism.** Zhang et al. (2003) suggested that SNPs will become the marker-of-choice in nuclear DNA analysis of populations because of their simplicity in character state and the ease of large-scale automated detection. And these markers will only be really valuable when they are employed in numbers. At present, as a result of extensive genetic studies and genome sequencing, a large numbers of genome wide SNPs have been identified in model organisms. However, there is only a few individual cases for SNPs markers to be systematically employed in population genetic study in non-model species. Our study indicates that, similar to that in *R. myconi* ( Picó et al. 2002 ), detailed SNPs on between- and within-population in *Gcyc* over the whole distribution area of *C. ramondioides*, on account of their neutral evolutionary pattern, could potentially be used to reconstruct the population history of this species. Although we can also extract this information from other markers, such as SSR, a comparison between methods would be very informative. Therefore, SNPs from known genes could be used as marker to carry out population genetic and evolutionary studies, especially in non-model species, such as *Gcyc* in *C. ramondioides*. Moreover, SNPs from the coding region of a developmental gene would contribute to our understanding the relationship between variation at the gene level and subsequent changes in biological functions and morphological traits.

We thank Prof. Song Ge (State Key Laboratory of Systematic and Evolution Botany, Institute of Botany, the Chinese Academy of Sciences, and Prof. Ting Wang (Wuhan Botanical Garden, the Chinese Chinese Academy of Sciences) for their advice and revision. We also thank Dr. Rui Wang for help in ArcGIS maping. This study was supported by National Natural Science Foundation of China, grant 30121003.

## References

Almeida J, Rocheta M, Gallego L (1997) Genetic control of flower shape in *Antirrhinum majus*. Development 124: 1387–1392

Barraclough TG (2001) Evolutionary rates and species divergence in flowering plants. Evolution 55: 677–683

Barrett SCH, Kohn JR (1991) Genetic and evolutionary consequences of small population size in plants: implication for convervation. In: Falk D A (ed) Genetic, conservation of rare plants. Blackwell Science, Oxford University Press, New York, Oxford, pp 3–30

Brookes AJ (1999) The essence of SNPs. Gene 234: 177–186

Brumfield RT, Beerli P, Nickerson DA, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol Evol 18(5): 249–256

Citerne HL, Luo D, Pennington RT, Coen E, Cronk QCB (2003) A phylogenomic investigation of *Cycloidea*-like TCP genes in the Leguminosae. Pl Physiol 131: 1042–1053

Citerne HL, Möller HLM, Cronk QCB (2000) Diversity of cyc-like in Gesneriaceae in relation to floral symmetry. Ann Bot 86: 167–176

Cubas P, Lauter N, Doebley J, Coen E (1999b) The TCP domain: a motif found in proteins regulating plant growth and development. Pl J 18: 215–222

Cubas P, Vincent C, Coen E (1999a) An epigenetic mutation responsible for natural variation in floral symmetry. Nature 401: 157–161

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in Maize. Nature 386: 485–488

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11–15

Fore AF, Guttman SI (1992) Genetic structure after forest fragmentation: a landscape ecology perspective of *Acer saccharum*. Canad J Bot 70: 1659–1668

Fukuda T, Yokoyama J, Maki M (2003) Molecular evolution of *CYCLOIDEA*-like genes in Fabaceae. Molec Evol 57: 588–597

Halushka MK, Fan JB, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nature Genetics 22: 239–247

Hamrick JK, Godt MJW (1990) Allozyme diversity in plant species. In: Brown AHD, Clegg MT, Kahler AL et al. (eds) Plant population genetics, breeding, and genetic resources, Mass. SinauerSunderland, pp 43–63

Han B, Xue YB (2003) Genome-wide intra-specific DNA-sequence in rice. Curr Opin Pl Biol 6: 134–138

Hanski I (1991) Metapopulation dynamics: brief history and conceptual domain. Biol J Linn Soc 42: 3–16

Husband BC, Barrett SCH (1996) A metapopulation perspective in plant population biology. J Ecol 84: 461–469

Kosugi S, Ohashi Y (1997) PCF1 and PCF2 specially bind to *cis* elements in the rice proliferating cell nuclear antigen gene. Pl Cell 9: 1607–1619

Luo D, Carpenter R, Vincent C, Copsey L, Coen E (1996) Origin of floral asymmetry in *Antirrhinum*. Nature 383: 794–799

Luo D, Carpenter R, Vincent C, Copsey L, Coen RS (1999) Control of organ asymmetry in flowers of *Antirrhimum*. Cell 99: 367–376

Marshall E (1997) Mapping away at genome patenting. Science 277: 1752–1753

Möller NJ (1999) Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study. Chapter 17: 395–402

Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Minobe Y (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. DNA Research Int J Rapid Publ Reports Genes Genomes 9: 163–171

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583–590

Osman A, Jordan B, Lessard PA, Muhannad N, Haron MR, Riffin NM, Sinskey AJ, Rha C, Housman DE (2003) Genetic diversity of *Eurycoma longifolia* inferred from single nucleotide polymorphisms. Pl Physiol 131: 1294–1301

Picó FX, Möller M, Ouborg NJ, Cronk QCB (2002) Single Nucleotide Polymorphisms in the coding region of the developmental gene Gcyc natural populations of the relict *Ramonda myconi* (Gesneriaceae). Pl Biol 4: 625–629

Rafalski JA (2002) Application of single nucleotide polymorphisms in crop genetics. Curr Opin Pl Biol 5: 94–100

Syvänen AC (2001) Accessing genetic variation: genotyping single nucleotide polymorphism. Nature Rev Genet 2: 930–942

Templeton AR, Shaw K, Routman E, Davis SK (1990) The genetic consequences of habitat fragmentation. Ann Missouri Bot Gard 77: 13–27

Tenaillon M, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci 98: 9161–9166

Wang CN, Möller M, Cronk QCB (2004a) Phylogenetic position of *Titanotrichum oldhamii* (Gesneriaceae) inferred from four different gene regions. Syst Bot 29(2): 407–407

Wang CN, Möller M, Cronk QCB (2004b) Population genetic structure of *Titanotrichum oldhamii* (Gesneriaceae), a subtropical bulbiluferous plant with mixed sexual and asexual reproduction. Ann Bot 93: 201–209

Wang L, Gao Q, Wang YZ, Lin QB (2006) Isolation and sequence analysis of two CYC-like genes, SiCYC1A and SiCYC1B, from zygomorphic and actinomorphic cultivars of *Saintpaulia ionantha* (Gesneriaceae). Acta Phytotax Sin 44(4): 353–361

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Molec Genet 2: 1123–1128

Wright S (1931) Evolution in mendelian populations. Genetics 16: 97–195

Young AG, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. Trends Ecol Evol 11: 413–419

Zhang DX, Hewitt GM (2003) Nuclear DNA analyses in geneti studies of populations: practice, problems and prospects. Molec Ecol 12: 563–584

Zhu YL, Song QJ, Hyten DL, Tassell CPV, Matukumalli LK, Grimm DR, Hayatt SM, Fickus EW, Young ND, Cregan PB (2003) Single-nucleotide polymorphisms in soybean. Genetics 163: 1123–1134