Single Nucleotide Polymorphisms in the Coding Region of the Developmental Gene *Gcyc* in Natural Populations of the Relict *Ramonda myconi* (Gesneriaceae)

F. X. Picó [1], M. Möller [2], N. J. Ouborg [1], and Q. C. B. Cronk [2,3]

[1] Department of Ecology, University of Nijmegen, Nijmegen, The Netherlands
[2] Royal Botanic Garden, 20A Inverleith Row, Edinburgh EH3 5LR, UK
[3] Institute of Cell and Molecular Biology, University of Edinburgh, Kings Buildings, Mayfield Road, Edinburgh EH9 3JH, UK

Abstract: *Gcyc* is a developmental gene, present in the Gesneriaceae family, that has both highly conserved and highly variable regions. *Ramonda myconi* (Gesneriaceae) is a paleoendemic plant restricted to mountainous areas of NE Spain. In this study we examine the population variation in the coding region of *Gcyc* in *R. myconi*. The fast-evolving nature of the coding regions of the *Gcyc* gene plus the ancient history of *R. myconi* together provide an appropriate background to obtain the first insights into population-level variation in developmental genes of flowering plants. One locus of *Gcyc* was specifically amplified and sequenced. Four single nucleotide polymorphisms (SNPs) were detected in the 420 sequenced bases of the gene in two *R. myconi* populations. The Pyrenean population had only one SNP while all four SNPs were present in the southern population. Three out of four SNPs were non-synonymous. Such novel results indicate that the detection of SNPs in *R. myconi* over its entire distribution area could be used as an aid to reconstructing the population history of the species, as well as to investigate the relationship between developmental genes and morphological traits.

Key words: Developmental genes, Mediterranean mountains, paleoendemic plants, *Ramonda myconi*, single nucleotide polymorphism.

Abbreviations:

Gcyc: homologue of CYC in Gesneriaceae
CYC: cycloidea
PCR: polymerase chain reaction
SNP: single nucleotide polymorphism

Indroduction

CYCLOIDEA (CYC) is a developmental gene involved in the expression of zygomorphy of *Antirrhinum* flowers (Luo et al., 1996[9]). CYC belongs to a family of genes (TCP family) present in several other angiosperm families (Cubas et al., 1999[4]; Möller et al., 1999[10]). The homologue of CYC in Gesneriaceae is Gcyc that appears to be a double or single copy nuclear gene (Citerne et al., 2000[2]). *Gcyc*, like other CYC-like genes, is interesting in that it has both highly conserved regions (e.g., the TCP box), as well as highly variable regions (Citerne et al., 2000[2]). Such highly variable regions of *Gcyc* have been used for phylogenetic analysis of the Gesneriaceae family, and to investigate the relationship between divergence of *Gcyc* sequences and changes in floral morphology in Gesneriaceae (Möller et al., 1999[10]).

Gesneriaceae is a tropical family of flowering plants with a few extant relict species in Europe (Good, 1974[6]). The genus *Ramonda* includes three of these species: *Ramonda myconi* (L.) Rchb. in the Pyrenees, and *R. serbica* Panc. and *R. nathaliae* Pancic and Petrovic in the Balkans. The paleoendemic *R. myconi* presents a typically fragmented distribution, restricted to mountainous areas, and occurs in clearly isolated populations (Picó and Riba, 2002[11]; Riba et al., 2002[12]). Core populations occur in the eastern Pyrenees whilst peripheral populations are located in more southern calcareous massifs in NE Spain. Nevertheless, the species' distribution range is rather reduced, and the most distant populations are separated by around 300 km. Despite this small distribution range, climatic conditions over the distribution area of *R. myconi* show great differences, ranging from sub-alpine in the north to Mediterranean in the south (Institut Cartogràfic de Catalunya, 1996[7]). These climate types differ greatly in precipitation and temperature, differences that may constitute important forces governing local ecological adaptations of species (Levin, 1992[8]).

Single nucleotide polymorphisms (SNPs) in the coding regions of genes may alter the function or structure of the encoded proteins, and may even cause serious disorders (Syvänen, 2001[14]). SNPs are commonly analysed for diagnostic purposes in humans because of the impact they have on many of the known inherited monogenic diseases. In fact, there is a catalogue of more than 1.4 million SNPs (approximately one SNP every 2000 or 3000 bases) pinpointing the exact location of each in the human genome (Sachidanandam et al., 2001[13]). However, SNPs may also provide fundamental new insights into the evolution of species, as changes at the DNA level allow populations to diverge and ultimately to form new species (Barraclough, 2001[1] and references therein). So far, there have been relatively few studies of SNPs in plants, and the impact of SNPs on gene expression and evolution in plants is not yet fully understood (see Cronk, 2001[3]).
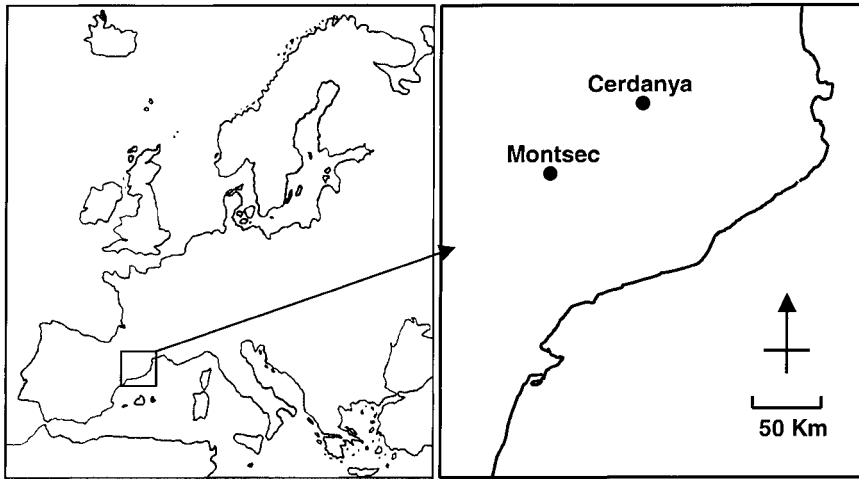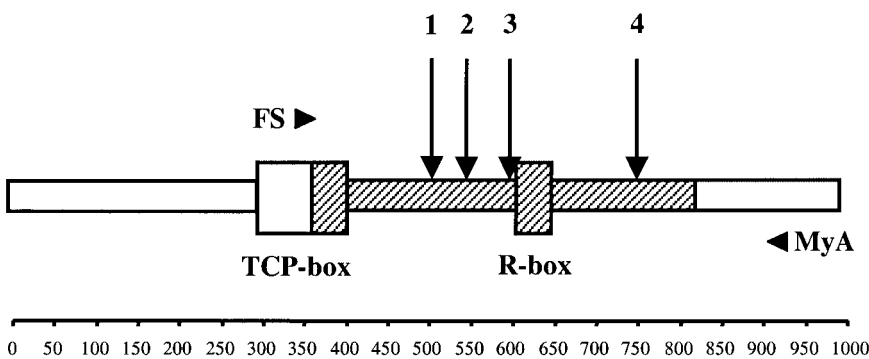
Fig. 2 Schematic structure and map of the *Gcyc* locus. Triangles indicate approximately the position of the FS and MyA primers and arrows the position of the four SNPs found within the analysed region of 420 bases indicated by the hatched area. The conserved TCP and R boxes are indicated.

The goal of this paper is to determine the approximate level of between- and within-population variation in *Gcyc* sequence of *R. myconi.* The fast-evolving nature of the coding regions of the *Gcyc* gene, plus the ancient history of *R. myconi* provide together an appropriate background to obtain the first insights into population-level variation in developmental genes of flowering plants. Furthermore, SNPs on the coding regions of *Gcyc* might also be used as markers to analyse re-colonization patterns of the relict *R. myconi* between glacial episodes. The evolutionary implications of the results are also discussed and future research lines presented.

Materials and Methods

Plant species and study sites

R. myconi is a long-lived iteroparous gesneriad that forms a basal rosette of leaves and flowers from late May to July. A detailed description of the species can be found in Picó and Riba (2002[11]) and Riba et al. (2002[12]). Populations of this paleoendemic plant are restricted to shady ravines on northern montane slopes of the Pyrenees and other massifs in NE Spain. For our study, we examined two contrasting populations from extremes of the species' distribution area (Fig. 1). One was a Pyrenean population located at La Cerdanya valley (IN population; 42°26′N, 1°55′E; 1500 m a.s.l.) and the other a southern population at the Montsec massif (CAR population; 41°59′N, 0°45′E;

1300 m a.s.l.). The climate at La Cerdanya is sub-alpine (total annual precipitation and mean annual temperature: 1250 mm and 6 °C) whereas at Montsec is continental (650 mm and 13 °C), but both with Mediterranean influence. Both populations have more than 1300 flowering plants. At each population, we collected leaf samples from 10 widely separated individuals across the population. Leaf samples were immediately stored in liquid nitrogen.

DNA extraction, primers, amplification, sequencing and analysis

DNA isolation was carried out using standard procedures (Doyle and Doyle, 1987[5]). For this study we chose one of the two loci of the *Gcyc* gene (Fig. 2), *Gcyc2,* which was amplified using standard and specific primers. The general forward primer *Gcyc*FS (ATG CTA GGT TTC GAC AAG CC) was designed and modified from *Antirrhinum* CYC sequences (Luo et al., 1996[9]) and a *Saintpaulia* cloned sequence (Möller et al., 1999[10]). The specific reverse primer *Gcyc*MyA (ATA GTT CGG ATT GTG CTG TTA A) was designed to exclusively amplify *Gcyc2.* Polymerase chain reaction (PCR) was as follows and made up to 50 µl with sterile distilled water: 0.2 µl *Taq* polymerase (5U µl) (Promega, Madison, WI, USA), 5 µl 10 × PCR buffer (Promega, Madison, WI, USA), 5 µl dNTP at 2 mM (Boehringer Mannheim, GmbH, Germany), 2.5 µl MgCl$_2$ at 50 mM (Promega, Madison, WI, USA), 5 µl of each primer at 10 µM

```
                    10        20        30        40        50        60
                     .         .         .         .         .         .
                =========  ========== ========== ========== ========== ==========
Ramonda_RBGE    ATGCTAGGTT TCGACAAGCC AAGTAAAACC CTTGAATGGC TTCTTACAAA ATCGAAAGCA
Ramonda_IN4     ATGCTAGGTT TCGACAAGCC AAGTAAAACC CTTGAATGGC TTCTTACAAA ATCGAAAGCA
Ramonda_CAR3    ATGCTAGGTT TCGACAAGCC AAGTAAAACC CTTGAATGGC TTCTTACAAA ATCGAAAGCA

                    70        80        90       100       110       120
                     .         .         .         .         .         .
                =========  =====----- ---------- ---------- ---------- --*-------
Ramonda_RBGE    GCCATTAAGG AGCTTGTGCA GATGAAGAAA AATGACTCCA CTACTTGCAC TAATATGAGT
Ramonda_IN4     GCCATTAAGG AGCTTGTGCA GATGAAGAAA AATGACTCCA CTACTTGCAC TAWTATGAGT
Ramonda_CAR3    GCCATTAAGG AGCTTGTGCA GATGAAGAAA AATGACTCCA CTACTTGCAC TAATATGAGT

                   130       140       150       160       170       180
                     .         .         .         .         .         .
                ---------- --------*- ---------- ---------- ---------- ----------
Ramonda_RBGE    GTTGTTTCTT CCCCTTCAGA CTGCGATGGG AACTACATAG ATCCGGATTC TAATGCACCG
Ramonda_IN4     GTTGTTTCTT CCCCTTCAGA CTGCGATGGG AACTACATAG ATCCGGATTC TAATGCACCG
Ramonda_CAR3    GTTGTTTCTT CCCCTTCAGA CTGCGATGGG AACTACATAG ATCCGGATTC TAATGCACCG

                   190       200       210       220       230       240
                     .         .         .         .         .         .
                ---------- ---------- ---------- ---------- ---------- ----------
Ramonda_RBGE    GCCAATTATT TTGCTTACGG TTGTGGAAGA ACAAAAGATC CACAGCAGGA TGTAATGAAC
Ramonda_IN4     GCCAATTATT TTGCTTACGG TTGTGGAAGA ACAAAAGATC CACAGCAGGA TGTAATGAAC
Ramonda_CAR3    GCCAATTATT TTGCTTACGG TTGTGGAAGA ACAAAAGATC CACAGCAGGA TGTAATGAAC

                   250       260       270       280       290       300
                     .         .         .         .         .         .
                ---*--+++ +++++++++++ ++++++++++ ++++++++++ ++++++++++ ++++++++++
Ramonda_RBGE    CTTGCCAAAG AATCGAGGGC TAAGGCGAGG GCGAGGGCTA GGGAAAGAAC TAGAGAGAAA
Ramonda_IN4     CTTGCCAAAG AATCGAGGGC TAAGGCGAGG GCGAGGGCTA GGGAAAGAAC TAGAGAGAAA
Ramonda_CAR3    CTTRCCAAAG AATCGAGGGC TAAGGCGAGG GCGAGGGCTA GGGAAAGAAC TAGAGAGAAA

                   310       320       330       340       350       360
                     .         .         .         .         .         .
                ---------- ---------- ---------- ---------- --*------- ----------
Ramonda_RBGE    ATGTGCAAGA AGAAGCTTAA TGATTCTAGA AACATGGCCT CTCCTTCTAG CCTCAATTTT
Ramonda_IN4     ATGTGCATGA AGAAGCTTAA TGATTCTAGA AACATGGCCT CTCCTTCTAG CCTCAATTTT
Ramonda_CAR3    ATGTGCATGA AGAAGCTTAA TGATTCTAGA AACATGGCCT CYCCTTCTAG CCTCAATTTT

                   370       380       390       400       410       420
                     .         .         .         .         .         .
                ---------- ---------- ---------- ---------- ---------- ----------
Ramonda_RBGE    CCCTTAGCTA ATAATACAGC TGCTGCTGCT GCTGCAACTG AAGAACTGAT TCATGAATCT
Ramonda_IN4     CCCTTAGCTA ATAATACAGC TGCTGCTGCT GCTGCAACTG AAGAACTGAT TCATGAATCT
Ramonda_CAR3    CCCTTAGCTA ATAATACAGC TGCTGCTGCT GCTGCAACTG AAGAACTGAT TCATGAATCT
```

Fig. 3 Complete sequences of the amplified region of *Gcyc2* (420 bases) for two samples of each population of study (*Ramonda_*IN4 and *Ramonda_*CAR3) plus one cloned sequence (*Ramonda_*RBGE) from a plant in the Royal Botanic Garden of Edinburgh (accession no. 1971 1477) deposited in the GenBank Nucleotide Sequence Database (GenBank accession no. AF208318). SNPs are indicated by (*), the TCP-box by (=) and the R-box by (+).

(Operon, DNA Services, Southampton, UK) and 2 µl aliquots of template DNA. Cycling parameters were as follows: 1 × 94 °C, 3 min; 30 × 94 °C, 60 s; 55 °C, 60 s; 72 °C, 90 s; 1 × 72 °C, 5 min. PCR products were run on 1% agarose gels (Gibco BRL, Life Technologies, Paisley, UK), and purified using Qiagen PCR purification kits (Qiagen GmbH, Hilden, Germany). Sequencing reactions (20 µl total volume) were performed using Thermo sequenase II reagent premix (Amersham Pharmacia Biotech Inc., Piscataway, NJ, USA). Products of sequencing were visualized with the program Factura™ version 1.2.0r6 (Applied Biosystems, Inc.). We detected heterozygotes by examining raw data from the electropherograms in which heterozygotes were indicated by double peaks of the same intensity. When checked, double peaks were detectable in both forward and reverse sequencing directions. The cloned sequence (RBGE) was obtained as described by Citerne et al. (2000[2]).

Results

A total of four SNPs were detected in the 420 sequenced bases of the *Gcyc2* (Table **1** and Fig. **3**). Bases near the specific reverse primer *Gcyc*MyA were not included in the analysis because of uncertainty in the sequence. The sequences of *Gcyc2* obtained for the two *R. myconi* populations were identical in length, differing only in the SNPs, to the cloned sequence of a plant grown in the Royal Botanic Garden of Edinburgh (Fig. **3**). This represents proof that we only amplified and sequenced *Gcyc2* with the specific primers used and that SNPs only occurred in that locus. Each SNP included two bases only: A/T (SNP-1), G/A (SNP-2), G/A (SNP-3) and T/C (SNP-4). SNP-1, SNP-2 and SNP-3 represented non-synonymous substitutions, whilst SNP-4 implied a synonymous substitution. The SNPs altered the predicted nucleotide and amino acid composition as follows. For the non-synonymous substitutions SNP-1, SNP-2 and SNP-3, nucleotide changes were AAT/ATT, GAC/AAC and GCC/ACC, respectively, and amino acid (standard codes) changes were N/I, D/N and A/T, respectively. For the synonymous substitution SNP-4, nucleotide changes were TCT/TCC maintaining the resulting amino acid (S/S). All four SNPs were found in the CAR population, whilst only SNP-1 was found in the IN population (Table **1**). Heterozygotes were detected in three of the four SNPs. In particular, SNP-3 showed four heterozygotes, SNP-1 two and SNP-4 only one (Table **1**). Levels of heterozygosity were generally lower than equilibrium expectations, but differences were significant only for SNP-1 ($\chi^2_{2,0.05} = 6.3$, $p < 0.05$) and SNP-2 ($\chi^2_{2,0.05} = 10.0$, $p < 0.01$) in the CAR population.

Table 1 Distribution of SNPs in the 420 bases sequenced of the *Gcyc2* gene found in *Ramonda myconi* samples from IN and CAR populations. The exact position of the SNP in the amplified region is indicated in parenthesis. Heterozygotes are indicated by W (A/T), R (A/G) and Y (C/T) and were detected by double peaks in the electropherograms

| Sample | SNP | | | | Sample | SNP | | | |
	1 (113)	2 (139)	3 (244)	4 (342)		1 (113)	2 (139)	3 (244)	4 (342)
IN3	T	G	G	T	CAR3	A	G	R	Y
IN4	W	G	G	T	CAR5	A	A	A	T
IN6	T	G	G	T	CAR6	A	G	R	T
IN7	T	G	G	T	CAR8	A	G	G	T
IN8	A	G	G	T	CAR9	W	G	R	T
IN9	T	G	G	T	CAR10	T	G	G	T
IN12	A	G	G	T	CAR18	A	G	R	T
IN13	T	G	G	T	CAR25	A	A	A	T
IN15	T	G	G	T	CAR27	T	G	G	T
IN16	T	G	G	T	CAR30	T	G	G	T

Discussion

The *Gcyc* gene contains several variable regions that differ considerably among species within the Gesneriaceae, indicating that part of this gene evolves rapidly (Möller et al., 1999[10]; Citerne et al., 2000[2]). Two contrasting hypotheses can explain the existence of fast-evolving coding regions. On the one hand, changes in these regions may not affect the functionality of the gene, so the gene would not be under purifying selection, while on the other hand, these interspecific differences could be intensely selected for during the speciation process during modification of floral morphology within Gesneriaceae. The fact that significant polymorphism can be detected at the population level indicates that the neutral hypothesis may be correct. In addition, the fact that three out of four of the SNPs are non-synonymous substitutions gives extra support for the reduced functionality hypothesis in *R. myconi* (Möller et al., 1999[10]).

Levels of *Gcyc* polymorphism in *Ramonda myconi* differed between populations. SNP-1 was detected in both populations, whilst the other three SNPs were found only in the CAR population. These results may indicate that SNP-1 is more ancient than the rest of the SNPs, which might have evolved later in the CAR population. Therefore, population history and the founder effect might play a central role in explaining between-population variation in *Gcyc*. Given the fact that *R. myconi* is considered a pre-glacial relict in the Mediterranean, the effect of glaciations in extant populations could have been very different throughout the species' distribution area. Hence, northern populations (i.e., IN population) could have been more affected by ice than southern populations (i.e., CAR population), so that genetic variation could have been lost due to either dramatic reductions in population size and subsequent genetic drift or a re-colonization process from south to north during interglacial periods.

For all four SNPs, homozygotes outnumbered heterozygotes. Although consistent conclusions cannot be drawn from sample sizes of 10 plants per population, such homozygote excess would be expected if *R. myconi* had an appreciable selfing rate. Hand-pollination experiments carried out on natural populations of *R. myconi* revealed that flowers are self-compatible but not capable of auto-pollination (F. X. Picó *unpublished result*). However, detailed data on pollinator visitation rate (mainly by bumblebees and syrphids), pollinator behaviour and selfing rate in the field are still lacking. It is possible that an appreciable amount of insect-mediated self-pollination (geitonogamy) occurs in *R. myconi.*

Understanding the relationship between variation at the gene level and subsequent changes in biological functions proves fundamental to obtain new insights into the evolution of plants (Cronk, 2001[3]). Given the importance that SNPs have in causing disorders in humans, several SNP genotyping methods have been recently developed to detect SNPs that are used as markers to identify genes that predispose individuals to certain disease (see Syvänen, 2001[14] for a comprehensive review). We suggest that some of these advanced genotyping methods could also be applied to known genes in natural plant populations, such as *Gcyc2* in *R. myconi,* to detect SNPs to be used as markers in population genetics and evolutionary studies.

Detailed data on between- and within-population variation in *Gcyc* over the whole distribution range of *R. myconi* could potentially be used to reconstruct the population history of this pre-glacial relict species. Although this information can also be extracted from allozyme or cpDNA studies, a comparison between methods would be very informative. Moreover, because *Gcyc* is a developmental gene, the relationship between *Gcyc* sequence variation and flower development could be investigated to study the link between developmental genes and morphological traits in wild populations.

Acknowledgements

References

[1] Barraclough, T. G. (2001) Evolutionary rates and species divergence in flowering plants. Evolution 55, 677 – 683.

[2] Citerne, H. L., Möller, M., and Cronk, Q. C. B. (2000) Diversity of cycloidea-like genes in Gesneriaceae in relation to floral symmetry. Annals of Botany 86, 167 – 176.

[3] Cronk, Q. C. B. (2001) Plant evolution and development in a postgenomic context. Nature Reviews Genetics 2, 607 – 620.

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

[5] Doyle, J. J. and Doyle, J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19, 11 – 15.

[6] Good, R. (1974) The Geography of Flowering Plants, 4th edn. London: Longman.

[7] Institut Cartogràfic de Catalunya (1996) Atles Climàtic de Catalunya. Termopluviometria. Barcelona: Institut Cartògrafic de Catalunya i Departament de Medi Ambient de la Generalitat de Catalunya.

[8] Levin, S. A. (1992) The problem of pattern and scale in ecology. Ecology 73, 1943 – 1967.

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

[10] Möller, M., Clokie, M., Cubas, P., and Cronk, Q. C. B. (1999) Integrating molecular phylogenies and developmental genetics: a Gesneriaceae case study. In Molecular Systematics and Plant Evolution (Hollingsworth, P. M., Bateman, R. M., and Gornall, R. J., eds.), London: Taylor and Francis, pp. 375 – 402.

[11] Picó, F. X. and Riba, M. (2002) Regional-scale demography of *Ramonda myconi:* remnant population dynamics in a pre-glacial relict species. Plant Ecology 161, 1 – 13.

[12] Riba, M., Picó, F. X., and Mayol, M. (2002) Effects of regional climate and small-scale habitat quality on performance in the relict species *Ramonda myconi.* Journal of Vegetation Science 13, 259 – 268.

[13] Sachidanandam, R., Weissman, D., Schmidt, S. C., et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928 – 933.

[14] Syvänen, S.-A. (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. Nature Reviews Genetics 2, 930 – 942.

F. X. Picó

Department of Ecology
University of Nijmegen
Toernooiveld 1
6525 ED Nijmegen
The Netherlands

E-mail: xavier.pico@sci.kun.nl