

RESEARCH ARTICLE

# Evidence of Purifying Selection and Co-Evolution at the Fold-Back Arm of the Novel Precursor MicroRNA159 Gene in *Phalaenopsis* Species (Orchidaceae)

Chi-Chu Tsai<sup>1,2</sup>, Yu-Chung Chiang<sup>3\*</sup>, I-Szu Weng<sup>1</sup>, Yu-Shium Lin<sup>1</sup>, Chang-Hung Chou<sup>4\*</sup>

1. Kaohsiung District Agricultural Research and Extension Station, Pingtung, 908, Taiwan, 2. Department of Biological Science and Technology, National Pingtung University of Science and Technology, Pingtung, 912, Taiwan, 3. Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung, 804, Taiwan, 4. Research Center for Biodiversity, China Medical University, Taichung, 404, Taiwan

\*[yuchung@mail.nsysu.edu.tw](mailto:yuchung@mail.nsysu.edu.tw) (YCC); [choumasa@mail.cmu.edu.tw](mailto:choumasa@mail.cmu.edu.tw) (CHC)



CrossMark  
click for updates

**OPEN ACCESS**

**Citation:** Tsai C-C, Chiang Y-C, Weng I-S, Lin Y-S, Chou C-H (2014) Evidence of Purifying Selection and Co-Evolution at the Fold-Back Arm of the Novel Precursor MicroRNA159 Gene in *Phalaenopsis* Species (Orchidaceae). PLoS ONE 9(12): e114493. doi:10.1371/journal.pone.0114493

**Editor:** Tzen-Yuh Chiang, National Cheng-Kung University, Taiwan

**Received:** July 23, 2014

**Accepted:** November 7, 2014

**Published:** December 3, 2014

**Copyright:** © 2014 Tsai et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability:** The authors confirm that all data underlying the findings are fully available without restriction. All of the validated pre-miR159 sequences for the *Phalaenopsis* species in the study have been deposited into GenBank with the accession numbers GU166689-GU166733.

**Funding:** This research was supported by funding from the Ministry of Science and Technology, Taiwan. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Abstract

**Background:** MicroRNAs (miRNAs) are small, endogenously transcribed, non-protein-coding RNAs that play important roles in regulation of gene expression in animals and plants. Here, selective constraints on the novel precursor microRNA159 (pre-miR159) gene were investigated in 42 *Phalaenopsis* species (Orchidaceae).

**Methods/Results:** A novel precursor microRNA159 gene was isolated from 42 *Phalaenopsis* species using a new microRNA-PCR (miR-PCR) approach. Sequencing of pre-miR159 genes revealed differences from the canonical pre-miR159 gene in *Phalaenopsis* species and other plants. Results demonstrated that the 5' and 3' fold-back arms and the terminal loop of the novel pre-miR159 gene have undergone purifying selection and selective constraint for stabilizing the secondary hairpin structure. Two conserved motifs within the 5' fold-back arm had the highest purifying selective pressure within the novel pre-miR159 gene. Evidence of sequence co-evolution between the 5' and 3' fold-back regions was observed.

**Conclusions:** Functional selective pressure might arise from the constraint of forming a hairpin structure and demonstrate co-evolution of sequences between the 5' and 3' fold-back regions of the novel pre-miR159 gene in *Phalaenopsis* species.

## Background

MicroRNAs (miRNAs) are small, endogenously transcribed, non-protein-coding RNAs that regulate gene expression in animals and plants. Most mature miRNA are transcribed as independent transcriptional units and are approximately 20–24 nucleotides (nt) long. During miRNA processing in plants, primary miRNAs (pri-miRNAs) are transcribed by RNA polymerase II. After transcription, pri-miRNAs are processed into precursor miRNAs (pre-miRNAs) by an RNase III-like enzyme called DICER-LIKE 1 (DCL1) [1–4]. Processing by DCL1 releases approximately 22-base-pair (bp) imperfect RNA duplex intermediates (miRNA/miRNA\* duplexes) [5]. Duplexes are exported to the cytoplasm where the RNA-induced silencing complex (RISC) produces one mature miRNA from the miRNA/miRNA\* duplex. The strand selected to produce the mature miRNA within RISC is biased towards the duplex strand with the weakest hydrogen bond at its 5' end. This weakly bonded strand is selectively incorporated into RISC [6]. Once mature, miRNAs down-regulate gene expression by mediating cleavage of mRNA and translational repression. To date, miRNAs have been found in a wide range of eukaryotes, including fruit flies, nematodes, zebrafish, chicken, mice, humans, *Arabidopsis*, maize, and rice [1].

Plant miRNAs recognize target mRNAs with near perfect base pairing. Therefore, computational sequence similarity searches can be used to identify potential targets [7]. In animals and plants, miRNAs are grouped into families where members differ by only a few nucleotides. Although family members are encoded at different loci, they are predicted to regulate similar or identical mRNAs [8]. Plant miRNAs can be encoded by the 5' or 3' fold-back arm of the hairpin. However, when miRNAs are encoded by multiple miRNA genes, miRNAs are always encoded by the same fold-back arm of the hairpin [9]. Early data suggested that plant miRNAs were conserved between monocots and dicots [10]. Indeed, twenty highly conserved miRNA families have been identified in three sequenced plant genomes: *Arabidopsis thaliana*, *Oryza sativa*, and *Populus trichocarpa*. However, deep-sequencing analyses revealed that most miRNAs are not conserved. In addition, although conserved miRNAs are often highly expressed [11, 12], copy numbers of miRNA genes are variable. Some families, including miR156 and miR159, contain numerous members in *A. thaliana*, *O. sativa*, and *P. trichocarpa*, whereas other families, such as miR162 and miR166, contain only a few genes [4].

The miR159 family is found in plants of the Embryophyta, Tracheophyta, Spermatophyta, Angiosperms, and Eudicots, indicating ancient origins [13]. According to miRBase, the miR159 family is encoded by multiple genes in *A. thaliana* [14], *O. sativa* [10], and *Glycine max* [15, 16] and by a single gene in *Arachis hypogaea* [17], *Festuca arundinacea* [18], and *Citrus sinensis* [19]. Single-stranded miR159 is 21 nt long and is always derived from the 3' fold-back arm of the pre-miR159 family in *Arabidopsis*, *Oryza*, and *Populus* [4]. In *Arabidopsis*, miR159 mediates cleavage of MYB transcription factor mRNA in the germinating seed [20] and regulates flowering time and other developmental events [21, 22].

Although miR319 resembles miR159 [4], indicating both miRNA genes evolved from a common ancestor [23], they have distinct target mRNAs [24].

To understand the evolutionary pattern of the fold-back arm of miRNAs in plants, several different pre-miRNA gene sequences and structures have been surveyed [23, 25–27]. Comparing pre-miRNA gene sequences between closely related species should help determine molecular evolution patterns [25] and address the role of selective constraint [28]. In addition, both pre-miR159 and pre-miR319 are firstly processed from the loop of hairpin structure by DCL1. This separates them from other plant miRNAs, which are processed via first cutting at the base of hairpin structure [29]. Here, molecular evolution patterns and functional constraint of the pre-miR159 gene might help to determine the origins of the unique miRNA processing pattern of pre-miR159. A new analytical approach, microRNA-PCR (miR-PCR), was developed to examine the different regions of the pre-miR159 gene in 42 *Phalaenopsis* species (Orchidaceae), an ornamental flowering plant found distributed throughout tropical Asia and the Pacific Islands [30] for which molecular phylogenies were reconstructed [31, 32]. Analyses were designed to determine whether selective pressure has acted on the sequence or the hairpin structure of the pre-miR159 gene during evolution.

## Results and Discussion

### Isolation of the novel pre-miR159 gene from *Phalaenopsis* species

A single band was amplified from each *Phalaenopsis* species using miR-PCR. Five clones were randomly selected for clone-based sequencing. Sequences suggested the pre-miR159 gene might be encoded at a single locus in *Phalaenopsis* species with the exception of *P. sumatrana*, *P. lindenii*, and *P. gibbosa*, which had distinct sequences in the hairpin region. To further validate sequences of the miRNA/miRNA\* duplexes for each *Phalaenopsis* species, five clones were randomly selected for amplification by inverse PCR (iPCR) and clone-based sequencing. Results showed a single miRNA/miRNA\* duplex for each species. To validate these results, thirty clones were randomly selected from *P. amabilis* and products from miR-PCR and iPCR were sequenced. Results supported the claim that the PCR product amplified by miR-PCR was homogeneous.

The sequence of the pre-miR159 gene from *Phalaenopsis* species differed from other plants, and the second miR159, miR159.2, present in other plants, was not found [33, 34]. To characterize the differences, the *Phalaenopsis amabilis* canonical pre-miR159 gene was isolated by genome walking; miR159 primers were used for targeting. Secondary structures of the canonical and novel pre-miR159 genes from *P. amabilis* are provided in Figure S1. Higher base pairing in the secondary structure (65.8%) of the canonical pre-miR159 gene was observed compared to the novel pre-miR159 gene (53.9%). This observation may explain why the canonical pre-miR159 gene was not amplified by miR-PCR [35]. Secondary structures of the novel pre-miR159 for all *Phalaenopsis* species were predicted in

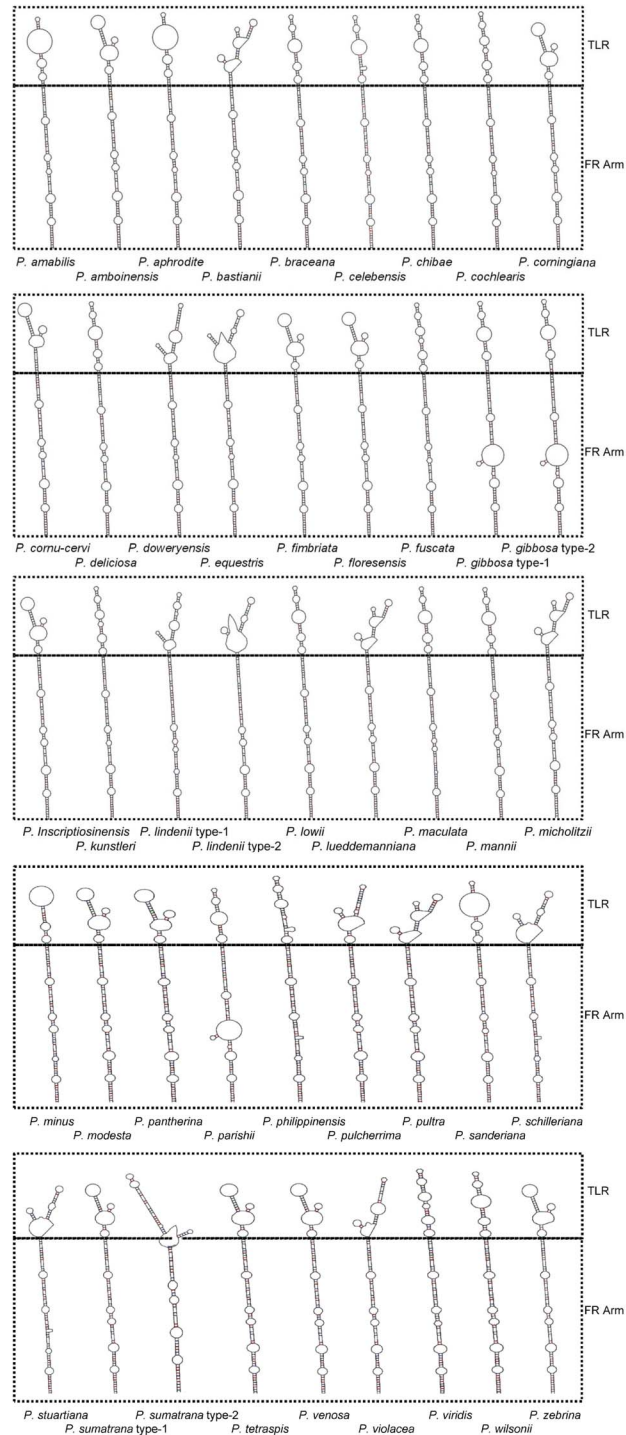
[Figure 1](#), and a close-up view of the pre-miRNA secondary structure is illustrated in [Figure S2](#).

### Nucleotide sequence diversity of the novel pre-miR159 gene from *Phalaenopsis* species

Nucleotide sequences of the novel pre-miR159 genes from 42 *Phalaenopsis* species were aligned. Of 214 sequenced bases, 42 sites were variable, including 19 single mutations ([Figure 2](#), [Table 1](#)). Based on sequence alignments ([Figure 2](#)) and secondary structure predictions ([Figure 1](#)), positions 1 to 73 and 140 to 214 were critical for stabilizing the hairpin structure. Therefore, segments 1 to 73, 74 to 139, and 140 to 214 represented the 5' fold-back arm, the terminal loop region, and the 3' fold-back arm, respectively ([Figure 2](#), [Table 1](#)).

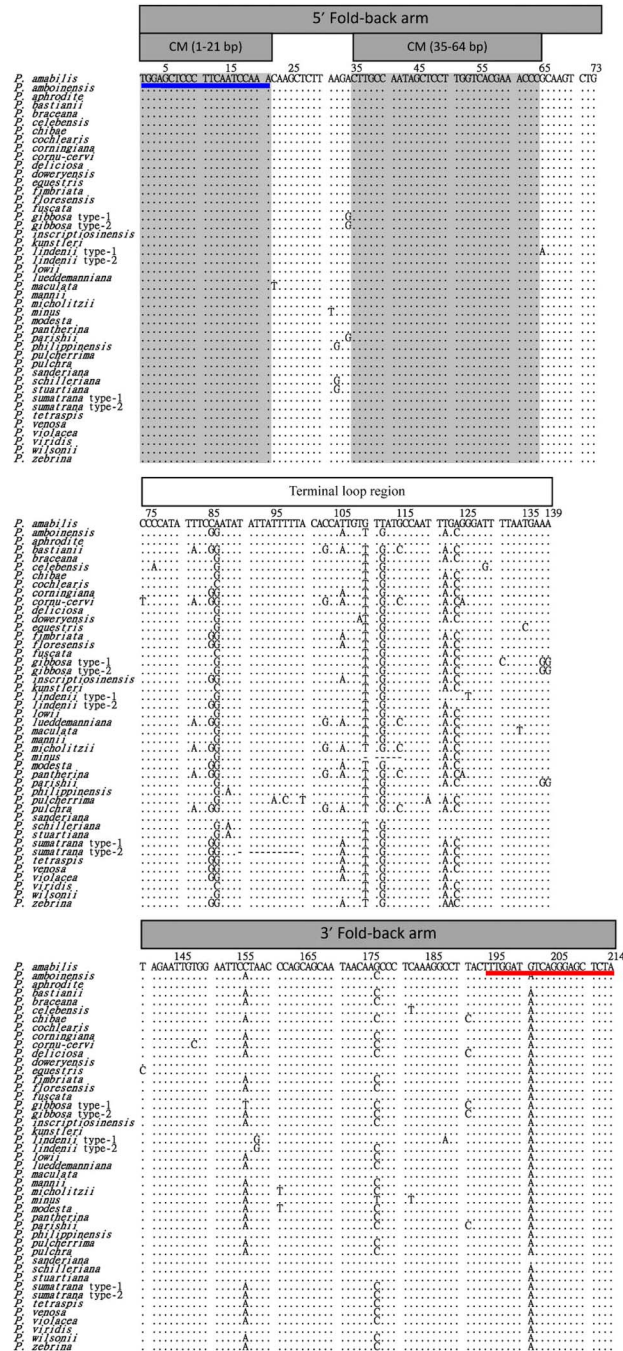
The nucleotide diversity was  $\theta=0.044$  for the entire novel pre-miR159 gene. The nucleotide diversity in the 5' fold-back arm, the terminal loop, and the 3' fold-back arm were  $\theta=0.016$ , 0.093, and 0.031, respectively. The nucleotide diversity in the terminal loop region was higher compared to the 5' and 3' fold-back arms, but it was close to that of the internal transcribed spacer 1 (ITS1) of nuclear ribosomal DNA (nrDNA) ([Table 1](#)). Substitution rates along the novel pre-miR159 gene revealed major differences between the three regions ([Figure 3](#)). High, variable substitution rates produced diversity in the terminal loop region, whereas two conserved motifs, 1 to 21 and 35 to 64, were present in the 5' fold-back arm. Compared to the 3' fold-back arm or the terminal loop region, the 5' fold-back arm was more conserved. In the terminal loop region, one-third of the nucleotides were variable (27 variable sites/66 sites). These results suggest variability in the novel pre-miR159 gene was non-random.

Pair-wise nucleotide differences between species ranged from 0 to 0.0758 (average: 0.0276), 0 to 0.0282 (average: 0.0052), 0 to 0.0701 (average: 0.0231), 0 to 0.1891 (average: 0.0627), and 0 to 0.3688 (average: 0.0965) for the entire novel pre-miR159 gene, the 5' fold-back arm, the 3' fold-back arm, the terminal loop region, and the ITS1 of nrDNA, respectively. Nucleotide differences in the novel pre-miR159 genes were compared to the ITS1 of nrDNA, a neutral locus. Matched-pairs t-tests revealed significantly lower values for pair-wise nucleotide differences between species across the entire gene and within the three regions of the novel pre-miR159 genes ( $p<0.001$ ). The maximum-likelihood relative rate test rejected the null hypothesis of rate constancy for 118 of 990 comparisons between *Paraphalaenopsis* outgroups and novel pre-miR159 genes ([Table S1](#)). Zero comparisons and few comparisons rejected the null hypothesis of rate constancy in the 5' fold-back arm and 3' fold-back arm (0 of 990 and 20 of 990 comparisons), respectively, but 140 of 990 comparisons rejected the null hypothesis of rate constancy in the terminal loop region ([Table S2–S4](#)). In instances where the null hypothesis was rejected, the terminal loop regions exhibited higher rates of nucleotide substitutions compared to the 5' and 3' fold-back arms. According to the definition of selective constraint described by Kimura and Takahata, these results indicate stronger purifying selection acted during



**Figure 1. The hairpin secondary structure of the novel pre-miR159 from the 42 *Phalaenopsis* species.** FR Arm means the fold-back arm and TLR means the terminal loop region.

doi:10.1371/journal.pone.0114493.g001



**Figure 2. Nucleotide polymorphisms in the novel pre-miR159 gene among the 42 *Phalaenopsis* species.** Nucleotides identical to the first line are indicated by a dot. Only base substitutions are indicated; deletion polymorphisms are indicated by dashes. The numbers at the top of the sequences represent the nucleotide positions. Positions 1 to 73, 74 to 139, and 140 to 214 represent the 5' fold-back arm, the terminal loop region, and the 3' fold-back arm, respectively. The gray regions are the longer highly conserved motifs among the 42 *Phalaenopsis* species. Blue and red lines, represent the predicted miRNA\* and miRNA.

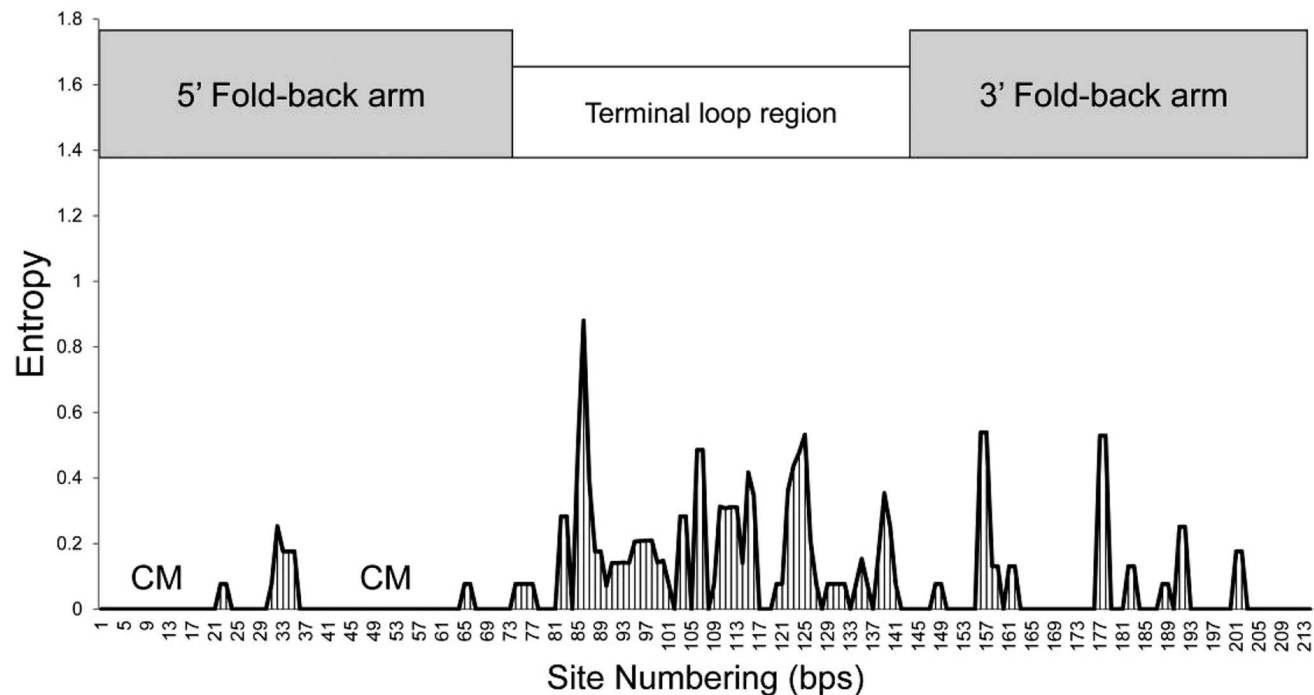
doi:10.1371/journal.pone.0114493.g002

**Table 1.** Pattern of nucleotide substitutions, including number of variable sites, nucleotide diversity ( $\theta$ ), and number of single mutations in the novel pre-miR159 gene (including 5' fold-back arm, terminal loop region, and 3' fold-back arm) and the internal transcribed spacer 1 (ITS1) of nuclear ribosomal DNA among 42 *Phalaenopsis* species.

DNA region	Positions	No. of sites	No. of variable sites	Nucleotide diversity ( $\theta$ )	No. of single mutations
pre-miR159	1..214	214	42	0.044	19
5' fold-back arm	1..73	73	5	0.016	3
Terminal loop region	74..139	66	27	0.093	13
3' fold-back arm	140..214	75	10	0.031	3
ITS1 of ribosomal DNA	1..266	266	87	0.110	33

doi:10.1371/journal.pone.0114493.t001

evolution of the 5' fold-back arm compared to the 3' fold-back arm or the terminal loop region [36]. Five, 27, and 10 variable sites were detected in the 5' fold-back arm, terminal loop region, and 3' fold-back arm, respectively (Table 1). Considering that evolutionary divergence of functional DNA is reduced by purifying selection [36], DNA regions that have undergone purifying selection should have fewer segregating sites compared with a linked neutral region [37–39]. These results demonstrate that the 5' and 3' fold-back arms have undergone purifying selection during evolution.



**Figure 3.** Substitution rate at each site within the novel pre-miR159 gene of 42 *Phalaenopsis* species. For each alignment of 42 sequences, the nucleotide substitution rate at each site was estimated by calculating entropy using the DAMBE software. A schematic of the gene is illustrated above the graph; fold-back arms are indicated as gray boxes and the terminal loop region is indicated as an open box. Two spikes in apparent diversity can be observed over the terminal loop region. Two conserved motifs (CM) are present in the 5' fold-back arm.

doi:10.1371/journal.pone.0114493.g003

## Functional constraints of the novel pre-miR159 gene fold-back arms in *Phalaenopsis* species

According to sequence alignments of the novel pre-miR159 genes of 42 *Phalaenopsis* species (Figure 2), 15 segregating sites exist within the 5' and 3' fold-back arms (Table 2). Comparing nucleotide sequence substitutions (Figure 2) and secondary hairpin structures (Figure 1), five newly formed base pairings which can increase the stability of the hairpin structure, and five synonymous base-pair substitutions (A-U→G-U, G-U→A-U, G-C→G-U, and G-U→G-C) were found [40]. Other substitutions located in internal loops, that did not affect the hairpin structure, were also discovered (Table 2). These results indicate that the substitutions found within the 5' and 3' fold-back arms do not destroy secondary structure. These results suggest that the novel pre-miR159 gene was randomly mutated during evolution, but that only substitution events that did not destroy secondary structure were retained and inherited. These results also indicated that functional constraints were present during evolutionary processing of the fold-back regions and demonstrate co-evolution of the 5' and 3' fold-back regions.

Secondary structure of the terminal loop in the novel pre-miR159 gene was variable among 42 *Phalaenopsis* species examined, although the stem structure of fold-back arms was conserved (Figure 1). No insertions/deletions (indels) were observed within the 5' or 3' fold-back arms of the novel pre-miR159 gene in each of 42 *Phalaenopsis* species examined (Figures 1 and 2). In contrast, indels were found in the terminal loop region in two taxa, *P. minus* and *P. sumatrana* (Figure 2). The novel pre-miR159 gene in *P. minus* contained a 6 nt deletion in the terminal loop, the fold-back structure resembled the novel pre-miR159 gene from other *Phalaenopsis* species (Figures 1 and 2; Figure S3). However, the conserved stem structure of fold-back arms was destroyed in *P. sumatrana* by a 10 nt deletion (positions 90 to 99 in the alignment sequence) (Figures 1 and 2; Figure S3). It is unclear whether the novel pre-miR159 gene in *P. sumatrana* can produce a mature miR159. However, several studies have shown that efficiency of miRNA production is reduced when the fold-back structure is destroyed [25, 41, 42]. These results indicate that the buffering capacity of indels in the terminal loop is higher compared to the arms. This may be because the structure of both arms is necessary for pre-miRNA processing [1–4].

To determine whether the secondary structure or the sequence of the novel pre-miR159 gene has any highly conserved regions, all predicted secondary structures and sequence alignments of the novel pre-miR159 genes from *Phalaenopsis* were compared (Figures 1 and 2). Two conserved internal loops, one within the miR159/miR159\* duplex and one near the 5'-end of miR159, were observed. Two highly conserved motifs within the 5' fold-back arm were also found. *Arabidopsis* DCL1, a miRNA processing protein, contains two double-strand RNA binding domains (dsRBDs); one plays a major role in pri-miRNA binding [43]. Selective constraint of the novel pre-miR159 5' fold-back arm may be involved in DCL1 targeting, suggesting that conserved structures and sequences in the novel pre-



miR159 gene may play important roles in the unique processing that has been described [29].

The first conserved motif is the complementary sequence (21 nt) of miR159. The second is located within the 5' fold-back arm at positions 35 to 64 (30 nt) (Figures 1 and 2). This result is consistent with the canonical pre-miR159 and pre-miR319 found in other plants [44]. The second conserved motif, found within the 5' fold-back arm, is only found in the pre-miR159 and pre-miR319 families [44]. Also revealed by sequence alignments, the nucleotide sequence of the second miRNA derived from the canonical pre-miR159 gene is not conserved among plants. The second miRNA derived from pre-miR159 is also processed differentially among plants. In *Phaseolus vulgaris*, the second miRNA is processed in response to stress [27, 34]. It has also been observed from the canonical pre-miR159 in *Phalaenopsis aphrodite* subspecies *formosana* by deep sequencing [45]. Because miRNA\* is selectively constrained for miRNA biogenesis [3], selective constraint for the second conserved motif of the novel pre-miR159 gene may come from processing of the second miR159.

### Evolution of the novel miR159 gene in *Phalaenopsis* species

In *Arabidopsis*, *Oryza*, and *Populus*, the miR159 family is derived from the 3' fold-back arm of the pre-miRNA [4]. The predicted miRNA159 derived from the novel pre-miR159 gene is 5'-UUUGGAUAUCAGGGAGCUCUA-3'; however, three *Phalaenopsis* species (*P. amabilis*, *P. aphrodite*, and *P. sandariana*) have a substitution at position 8 from the 5' end (i.e., 5'-UUUGGAUGUCAGGGAGCUCUA-3', the substitution is underlined). The morphological characteristics of these *Phalaenopsis* species are distinct from other members of the section *Phalaenopsis*, including *P. philippinensis*, *P. schilleriana*, and *P. stuartiana*, which have marbling on the upper surface of their leaves and bear anthocyanins in the leaves [30]. The two predicted miR159 derived from the novel pre-miR159 genes in *Phalaenopsis* was aligned with other members of the miR159 family from miRBase (Figure 4). According to alignments, the length of miR159 was 21 or 20 nt, and the 5' and 3' ends were more variable. In addition, alignments revealed no correlation between sequence divergence and phylogenetic relationships, suggesting that miR159 family members have experienced high levels of selective constraint during evolution.

The predicted miR159 from three species, *P. sumatrana*, *P. lindenii*, and *P. gibbosa*, may be processed from two copies of the novel pre-miR159 gene. This was indicated by variations (substitutions or deletions) observed in the terminal loop region of the novel pre-miR159 gene in these species. These three non-conserved paralogs of the novel pre-miRNAs may be considered young copies [13], which would be consistent with the model that miRNAs are created and destroyed continuously during evolution [11, 46]. This result also indicates that a duplication event occurred in these three species.

The canonical pre-miR159 gene from *P. amabilis* was further isolated by genome walking (Figure S1). The miR159 (5'-

**Table 2.** Nucleotide substitutions within the fold-back arms of the novel pre-miR159 genes and the corresponding pairing in the secondary structure among 42 *Phalaenopsis* species.

Taxa	Sequence substitution in the novel pre-miR159 gene fold-back regions		
	Position <sup>a</sup>	Substitution type	Nucleotide of corresponding pair in the secondary structure
<i>P. maculata</i>	22	C→U*	A
<i>P. minus</i>	31	A→U*	A
<i>P. schilleriana</i> , <i>P. stuartiana</i> , <i>P. philippinensis</i>	32	A→G**	U
<i>P. gibbosa</i> -type-1, <i>P. gibbosa</i> -type-2, <i>P. parishii</i>	34	A→G***	Internal loop
<i>P. lindenii</i> -type-1	65	G→A**	U
<i>P. equestris</i>	140	U→C**	G
<i>P. cornu-cervi</i>	148	U→C**	G
14 species ( <i>P. gibbosa</i> -type-1)	156	A→C*** (U***)	Internal loop
<i>P. lindenii</i> -type-1, <i>P. lindenii</i> -type-2	158	A→G*	C
<i>P. modesta</i> , <i>P. micholitzii</i>	161	C→U**	G
14 species ( <i>P. minus</i> )	177	C→G*(U***)	C (Internal loop)
<i>P. celebensis</i> , <i>P. minus</i>	182	C→U*	A
<i>P. lindenii</i> -type-1	188	C→A***	C
<i>P. chibae</i> , <i>P. deliciosa</i> , <i>P. parishii</i> , <i>P. gibbosa</i> -type-1, <i>P. gibbosa</i> -type-2	191	U→C***	Internal loop
<i>P. amabilis</i> , <i>P. aphrodite</i> , <i>P. sanderiana</i>	201	A→G***	Internal loop

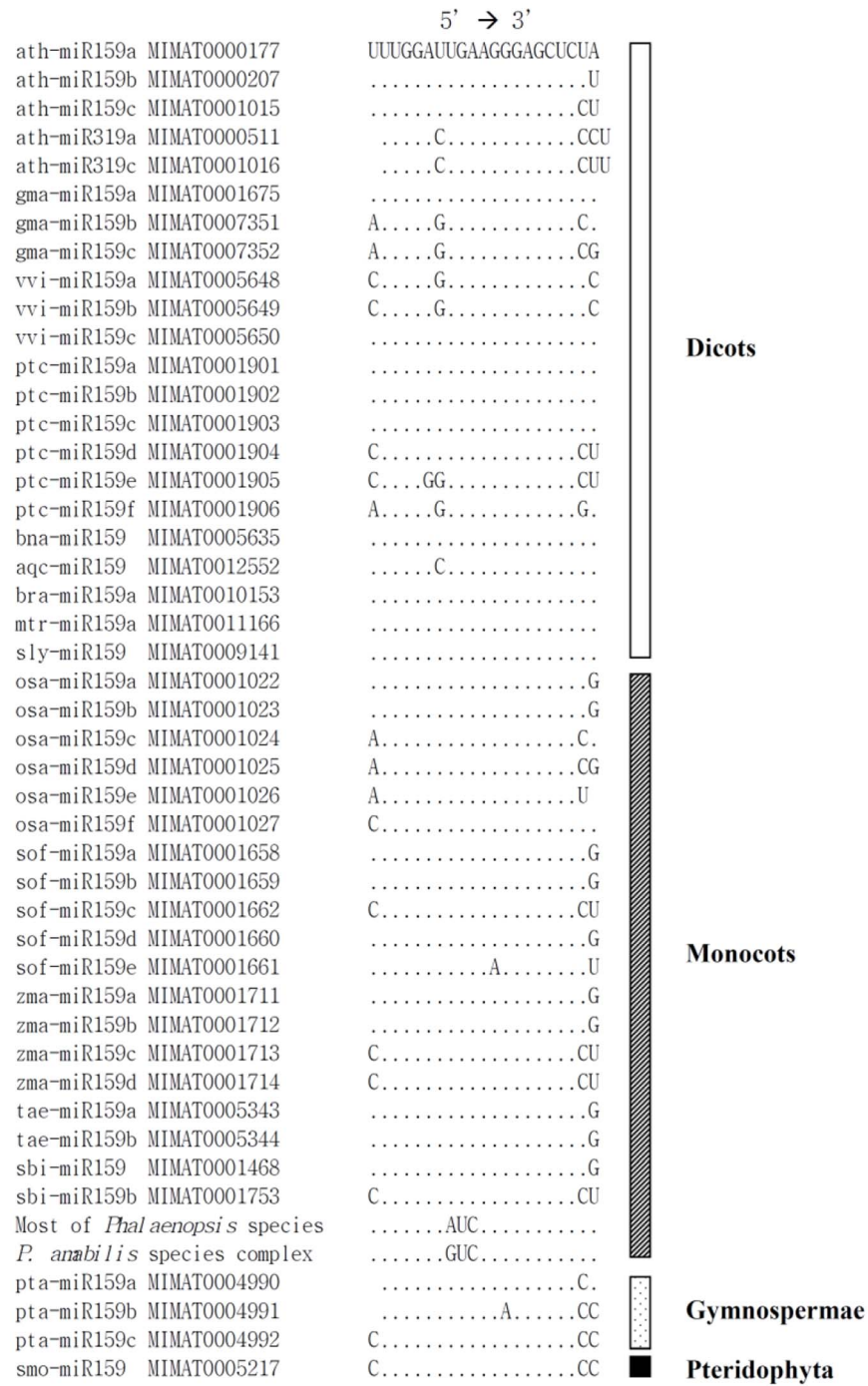
<sup>a</sup>Nucleotide position according to [Figure 1](#). Positions 1–73, 74–139, and 140–214 represent the 5' fold-back arm, the terminal loop region, and the 3' fold-back arm, respectively.

\*Represents increasing stability (new base pairing) of the fold-back structure; \*\* Represents a minor change in fold-back stability (increase or decrease) without destroying the nucleotide pairing (i.e., A-U→G-U, G-C→G-U, G-U→G-C, or G-U→A-U); \*\*\* Represents no change in the fold-back stability or substitutions located in the internal loops

doi:10.1371/journal.pone.0114493.t002

UUUGGAUUGAAGGGAGCUCUA-3') derived from the canonical pre-miR159 gene of *P. amabilis* is typical [9] ([Figure 4](#)) and abundantly expressed in *P. aphrodite* subspecies *formosana* [45]. The novel pre-miR159 genes isolated from all *Phalaenopsis* species can form hairpin structures and are subject to selective constraint for stabilizing the fold-back structure. Therefore, the novel pre-miR159 gene may elicit biological function by generating miRNAs that down-regulate target mRNA.

Although miRNA and its target mRNA require near-perfect base pairing in plants [7], three substitutions between novel miR159 and canonical miR159 were discovered. These changes, located in the seed region (defined as the second to the seventh nucleotides in the mature miRNA) which are critical for target recognition [47]. Indeed, both canonical miR159 and its target site can be found on the *MYB* transcript in *Phalaenopsis* species [45, 48]. These data indicate new miRNA genes may evolve by point mutation and selection against inadequate miRNA/mRNA pairing [49]. Therefore, the predicted novel miR159 might target and down-regulate other unknown mRNAs. Two other possibilities may explain the novel miR159 sequence observed in *Phalaenopsis* species. First, novel miR159 cannot be accurately processed from the novel pre-miR159 gene due to tissue-specific control of miRNA processing [50, 51]. This is supported by the absence of



**Figure 4. A comparison of the predicted miR159 derived from the novel pre-miR159 genes of the *Phalaenopsis* species and members of the miR159 family from diverse plants.** Positions 8–10 from the 5' end of the miR159 derived from the *Phalaenopsis* species indicate three distinct similarities with other members of the miR159 family. Two types of miR159 can be observed across the *Phalaenopsis* species. Position 8 is adenosine (A) in most of the *Phalaenopsis* species, whereas in other species (the *Phalaenopsis amabilis* species complex, which includes *P. amabilis*, *P. aphrodite*, and *P. sanderiana*) it is guanosine (G).

doi:10.1371/journal.pone.0114493.g004

novel miR159 in leaf tissues of *P. aphrodite* subspecies *formosana* [45]. Second, RNA editing of pre-miRNA/miRNA has been observed in several studies [52–54]. Therefore, novel pre-miR159 might undergo RNA editing to generate canonical miR159 in *Phalaenopsis* species.

## Conclusions

A novel pre-miR159 gene was isolated from *Phalaenopsis* species. The nucleotide sequence of the novel pre-miR159 gene differed from the canonical pre-miR159 gene in *Phalaenopsis* species and other plants. Regions of the novel pre-miR159 gene were associated with distinct purifying selective pressures. The 5' fold-back arm displayed evidence of strong purifying selection during evolution, and the 5' and 3' fold-back arms were subject to selective constraints. Selective constraints were also indicated for the stem of the hairpin structure in the novel pre-miR159 gene, and evidence of co-evolution of the 5' and 3' fold-back regions was uncovered. Strong purifying selection of the 5' fold-back arm implied that motifs in the region may be critical for miR159 processing and biogenesis. Moreover, it appears that the novel pre-miR159 gene has undergone duplication events.

## Methods

### Plant materials

Forty-two species were selected from the subgenera and sections of the genus *Phalaenopsis*. Leaf materials were collected from living plants cultivated in the Kaohsiung District Agricultural Research and Extension Station (KDARES) in Taiwan. Voucher specimens were deposited at the herbarium of the National Museum of Natural Science, Taiwan (TNM). Details of the materials, their distributions, and systematic classifications are listed in [Table 3](#).

### Primer design and PCR amplification of the novel pre-miR159 gene

Genomic DNA was extracted from fresh *Phalaenopsis* leaves using the cetyltrimethylammonium bromide protocol [55]. To investigate sequence variation of the pre-miR159 gene between *Phalaenopsis* species, a new analytical approach, based on near-perfect base pairing and inverted repeats located at both ends of the pre-miRNA [7], were developed. Taking into account that the miRNA is located on the same fold-back arm (5' or 3') among diverse plants [4] and that the length of the mature miRNA is approximately 20–24 nt [1, 3, 4], a single primer was designed to amplify the pre-miRNA region. This approach was named microRNA-PCR (miR-PCR). Primers derived from the conserved miR159 region in *Arabidopsis* (ath-miR159a: 5'-UUUGGAUUGAAGGGAGCUCUA-3') [9, 56] and *Oryza* (osa-miR159a: 5'-UUUGGAUUGAAGGGAGCUCUG-3') [10] were designed to amplify the pre-miR159 region from *Phalaenopsis*. The conserved sequence of miR159a, 5'-UUUGGAUUGAAGGGAGCUCU-3', is located on the

**Table 3.** Names of the specimens, geographical distributions, and sources for the plant materials in the study.

Taxa and systematic classification <sup>a</sup>	Geographical distribution	Voucher <sup>b</sup>
Genus <i>Phalaenopsis</i>		
<b>Subgenus <i>Proboscidioides</i> (Rolfe) E. A. Christ.</b>		
<i>Phalaenopsis lowii</i> Rchb.f. <sup>b</sup>	Myanmar and adjacent western Thailand	C. C. Tsai 1714
<b>Subgenus <i>Aphyllae</i> (Sweet) E. A. Christ.</b>		
<i>Phalaenopsis wilsonii</i> Rolfe	China (Sichuan, Yunnan, and eastern Tibet)	C. C. Tsai 1645
<i>Phalaenopsis minus</i> (Seidenf.) E. A. Christ.	Endemic to Thailand	C. C. Tsai 1791
<i>Phalaenopsis braceana</i> (J. D. Hook.) E. A. Christ.	Bhutan and China	C. C. Tsai 1747
<b>Subgenus <i>Parishianae</i> (Sweet) E. A. Christ.</b>		
<i>Phalaenopsis gibbosa</i> Sweet	Vietnam and Laos	C. C. Tsai 1783
<i>Phalaenopsis parishii</i> Rchb.f.	Eastern Himalayas, India, Myanmar, and Thailand	C. C. Tsai 1316
<b>Subgenus <i>Polychilos</i> (Breda) E. A. Christ.</b>		
<b>Section <i>Polychilos</i> (Breda) Rchb.f.</b>		
<i>Phalaenopsis manni</i> Rchb.f.	Northeast India, Nepal, and China to Vietnam	C. C. Tsai 1796
<i>Phalaenopsis cornu-cervi</i> (Breda) Bl. & Rchb.f.	Northeast India and the Nicobar Islands to Java and Borneo	C. C. Tsai 1562
<i>Phalaenopsis pantherina</i> Rchb.f.	Endemic to Borneo	C. C. Tsai 1395
<b>Section <i>Fuscatae</i> Sweet</b>		
<i>Phalaenopsis cochlearis</i> Holtt.	Malaysia (Malay Peninsula) and Indonesia (Sarawak)	C. C. Tsai 1722
<i>Phalaenopsis viridis</i> J. J. Sm.	Endemic to Indonesia (Sumatra)	C. C. Tsai 1141
<i>Phalaenopsis fuscata</i> Rchb.f.	Malaysia (Malay Peninsula), Borneo (West Koetai)	C. C. Tsai 1733
<i>Phalaenopsis kunstleri</i> J. D. Hook.	Myanmar and Malay Peninsula	C. C. Tsai 1139
<b>Section <i>Amboinenses</i> Sweet</b>		
<i>Phalaenopsis pulchra</i> (Rchb.f.) Sweet	Endemic to the Philippines (Luzon and Leyte)	C. C. Tsai 1281
<i>Phalaenopsis violacea</i> Witte	Indonesia (Sumatra) and Malaysia (Malay Peninsula)	C. C. Tsai 1598
<i>Phalaenopsis micholitzii</i> Rolfe	Philippines (Mindanao)	C. C. Tsai 1674
<i>Phalaenopsis fimbriata</i> J. J. Sm.	Indonesia (Java, Sarawak, and Sumatra)	C. C. Tsai 1732
<i>Phalaenopsis floresensis</i> Fowlie	Endemic to the island of Flores	C. C. Tsai 1610
<i>Phalaenopsis doweryensis</i> Garay & E. A. Christ.	East Malaysia (Sabah), without a precise locality	no voucher
<i>Phalaenopsis modesta</i> J. J. Sm.	Endemic to the island of Borneo in East Malaysia (Sabah) and Indonesia (Kalimantan)	C. C. Tsai 1756
<i>Phalaenopsis maculata</i> Rchb.f.	Malaysia (Pahang), East Malaysia (Sabah and Sarawak), and Indonesia (Kalimantan Timur)	C. C. Tsai 1774
<i>Phalaenopsis amboinensis</i> J. J. Sm.	Indonesia (Molucca Archipelago and Sulawesi)	C. C. Tsai 1605
<i>Phalaenopsis lueddemanniana</i> Rchb.f.	Endemic to the Philippines	C. C. Tsai 1162
<i>Phalaenopsis bastianii</i> Gruss & Rollke	Endemic to the Philippines	C. C. Tsai 1412
<i>Phalaenopsis venosa</i> Shim & Fowlie	Endemic to Indonesia (Sulawesi)	C. C. Tsai 1014
<b>Section <i>Zebrinae</i> Pfitz.</b>		
<i>Phalaenopsis inscriptiosinensis</i> Fowlie	Endemic to Indonesia (Sumatra)	no voucher
<i>Phalaenopsis tetraspis</i> Rchb.f.	India (Andaman and Nicobar Islands) and Indonesia (Sumatra)	C. C. Tsai 1693
<i>Phalaenopsis comingiana</i> Rchb.f.	Borneo (Sarawak and elsewhere on the island)	C. C. Tsai 1776
<i>Phalaenopsis sumatrana</i> Korth. & Rchb.f.	Widespread from Myanmar, Thailand, and Vietnam, to Indonesia (Java and Sumatra), Malaysia (Perak and Johore), East Malaysia (Sabah), and the Philippines (Palauan)	C. C. Tsai 1650
<i>Phalaenopsis zebrina</i> Witte		no voucher

Table 3. Cont.

Taxa and systematic classification <sup>a</sup>	Geographical distribution	Voucher <sup>b</sup>
<b>Subgenus <i>Phalaenopsis</i></b>		
<b>Section <i>Phalaenopsis</i> Benth.</b>		
<i>Phalaenopsis philippinensis</i> Golamco ex Fowlie & Tang	Endemic to the Philippines	C. C. Tsai 1333
<i>Phalaenopsis amabilis</i> (L.) Blume	Widespread from Sumatra and Java to the southern Philippines, and east to New Guinea and Queensland, Australia	C. C. Tsai 193
<i>Phalaenopsis aphrodite</i> Rchb.f.	Northern Philippines and southeastern Taiwan	C. C. Tsai 1420
<i>Phalaenopsis sanderiana</i> Rchb.f.	Endemic to the Philippines	C. C. Tsai 1526
<i>Phalaenopsis schilleriana</i> Rchb.f.	Endemic to the Philippines	C. C. Tsai 1004
<i>Phalaenopsis stuartiana</i> Rchb.f.	Endemic to the island of Mindanao in the southern Philippines	C. C. Tsai 1419
<b>Section <i>Deliciosae</i> E. A. Christ.</b>		
<i>Phalaenopsis chibae</i> Yukawa	Endemic to Vietnam	C. C. Tsai 1792
<i>Phalaenopsis deliciosa</i> Rchb.f.	Widespread from Sri Lanka and India to the Philippines and Sulawesi	C. C. Tsai 1664
<b>Section <i>Esmeralda</i> Rchb.f.</b>		
<i>Phalaenopsis pulcherrima</i> (Lindl.) J. J. Sm.	Widespread from northeast India and southern China throughout Indochina to Malaysia (Malay Peninsula), Indonesia (Sumatra), and East Malaysia (Sabah)	C. C. Tsai 1020
<b>Section <i>Stauroglottis</i> (Schauer) Benth.</b>		
<i>Phalaenopsis equestris</i> (Schauer) Rchb.f.	Philippines and Taiwan	C. C. Tsai 1689
<i>Phalaenopsis celebensis</i> Sweet	Endemic to Indonesia (Sulawesi)	C. C. Tsai 1671
<i>Phalaenopsis lindenii</i> Loher	Endemic to the Philippines	C. C. Tsai 1574

<sup>a</sup>The systematic characterizations of *Phalaenopsis* are based on Christenson (2001).

<sup>b</sup>Their voucher specimens were deposited at the herbarium of the National Museum of Natural Science, Taiwan (TNM).

doi:10.1371/journal.pone.0114493.t003

3' fold-back arm. Consequently, the sequence of the single primer used for amplifying the novel pre-miR159 region of the *Phalaenopsis* species was 5'-AGAGCTCCCTTCAATCCAAA-3'.

PCR reactions (25 µl) contained 40 mM Tricine-KOH (pH 8.7), 15 mM KOAc, 3.5 mM Mg(OAc)<sub>2</sub>, 3.75 µg/ml BSA, 0.005% Tween 20, 0.005% Nonidet-P40, four dNTPs (0.2 mM each), primers (0.4 µM each), 1.25 U of Advantage 2 DNA polymerase (Clontech Laboratories, Inc., CA, USA), and 10 ng of genomic DNA. Cycling was performed in a thermocycler (Biometra, Germany) under the following conditions: 94°C for 5 minutes followed by 40 cycles of denaturation at 94°C for 40 seconds, annealing at 50°C for 35 seconds, and extension at 72°C for 50 seconds, with a final extension at 72°C for 7 minutes. PCR products were visualized on a 1% agarose gel. A product of the expected size was amplified from each of the samples. Amplified products were purified using Qiagen columns (Valencia, CA, USA), and purified PCR products were cloned into pGEM-T Easy Vectors (TaKaRa, Japan). Five independent clones were sequenced using the dideoxy chain-termination method and an ABI3730 automated sequencer with the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, CA, USA). Sequencing reactions were performed according to the manufacturer's recommendations.

## Inverse PCR

Inverse PCR (iPCR) was performed as described by Ochman *et al.* [57]. Universal primers for iPCR were designed within the hairpin structure, excluding the miRNA/miRNA\* duplex: F, 5'-GTGGAATTCATAACCCAGTAGTA-3' and R, 5'-GGGTTTCGTGACCAAGGAGCTA-3'. Nested primers for the second PCR were F, 5'-ATTCATAACCCAGCAGCAATAACA-3', and R, 5'-CTATTGGCAAGTCTTAAGAGCTTG-3'. Genomic DNA from 42 *Phalaenopsis* species was digested with *Dra*I. DNA fragments were ligated to obtain intramolecular circularized DNA for iPCR amplification. PCR reactions (25  $\mu$ l) contained 40 mM Tricine-KOH (pH 8.7), 15 mM KOAc, 3.5 mM Mg(OAc)<sub>2</sub>, 3.75  $\mu$ g/ml BSA, 0.005% Tween 20, 0.005% Nonidet-P40, four dNTPs (0.2 mM each), primers (0.4  $\mu$ M each), 1.25 U Advantage 2 DNA polymerase (Clontech Laboratories, Inc., CA, USA), and 10 ng genomic DNA. Cycling was performed in a thermocycler (Biometra) using the following conditions: 94°C for 5 minutes followed by 35 cycles of denaturation at 94°C for 40 seconds, annealing at 56°C for 40 seconds, and extension at 72°C for 2 minutes, with a final extension at 72°C for 7 minutes. PCR products were detected by agarose gel electrophoresis (1.0% w/v in TBE), stained with 0.5  $\mu$ g/ml ethidium bromide, and photographed under UV light exposure. A product of the expected size was amplified from each of the samples, and amplified products were purified using Qiagen columns (Valencia, CA, USA). Purified PCR products were cloned into *pGEM-T Easy Vectors* (TaKaRa, Japan), and five independent clones were sequenced. Cloned DNA was sequenced following the dideoxy chain-termination method using an ABI3730 automated sequencer with the BigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems, CA, USA). Sequencing reactions were performed according to the manufacturer's recommendations. All validated pre-miR159 sequences have been deposited into GenBank with the accession numbers GU166689-GU166733. To compare nucleotide substitutions between neutral and novel pre-miR159 gene sequences, sequences of the internal transcribed spacer 1 (ITS1) of nuclear ribosomal DNA (nrDNA) for these *Phalaenopsis* species published by Tsai *et al.* [31] were re-aligned. To estimate the relative rate test, two *Paraphalaenopsis* taxa (*Paraphalaenopsis laycockii* and *Paraphalaenopsis serpentina*) were included in the study as outgroups.

## Genome walking

Using miR159.2 as a primer targeting region, the canonical pre-miR159 gene was isolated using the Genome Walker Universal Kit (Clontech Laboratories, Inc., CA, USA). PCR was performed with Advantage 2 DNA polymerase (Clontech Laboratories, Inc., CA, USA). DNA fragments were purified with the QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA). Recovered PCR products were ligated into T-vectors (Promega, Wisconsin, USA), and recombinants were transformed into *Escherichia coli* DH5 $\alpha$  (RBC, Taipei, Taiwan). Plasmid DNA was purified using Qiagen spin mini prep kits. Plasmid DNA was sequenced with vector-specific primers (SP6 and T7) using the dideoxy chain-termination

method, an ABI3730 automated sequencer, and the Ready Reaction Kit (PE Biosystems, CA, USA) for BigDye Terminator Cycle Sequencing. Each sample was sequenced three times. Reactions were performed according to the manufacturer's recommendations.

### Sequence alignment, secondary structure prediction, and nucleotide variability

Sequences of novel pre-miR159 genes were aligned using the Clustal W multiple alignment program in BioEdit [58]. The hairpin structure of the novel pre-miR159 was predicted using RNA folding software [59]. Alignment results coupled with the secondary structure of the novel pre-miR159 genes were used to guide the division of pre-miR159 into three regions: the 5' fold-back arm, the terminal loop region, and the 3' fold-back arm. To detect sequence polymorphisms for the different regions of pre-miR159 gene, the number of variable sites, nucleotide diversity ( $\theta$ ), and single mutations were estimated using DNASP version 4.10 [60].

### Substitution rate at each site (entropy) calculation

To evaluate variability and complexity of each nucleotide site, entropy for each nucleotide site was estimated using the Shannon entropy formula:

$H_i = -(\sum_{j=1}^4 P_{ij} \log_2 P_{ij})$  [61], where  $H_i$  corresponds to the entropy of each site  $i$ ;  $j$  is equal to 1, 2, 3 and 4, corresponding to the A, C, G, and T nucleotides, respectively; and  $P_{ij}$  is the proportion of nucleotide  $j$  in site  $i$  [62]. For entropy analyses, aligned sequences of the novel pre-miR159 genes were estimated using Data Analysis in Molecular Biology and Evolution (DAMBE) v. 5.2.76 [63].

### Determination of nucleotide substitutions per site ( $D_{XY}$ ) and the relative rate test

The number of nucleotide substitutions per site between species ( $D_{XY}$ ) was estimated using the six-parameter method [64]. Pair-wise nucleotide differences were calculated using DAMBE v.5.2.76 [63] for the entire novel pre-miR159 genes, the 5' fold-back arm, the 3' fold-back arm, the terminal loop region, and the ITS1 of nrDNA, respectively. To compare significance between the novel pre-miR159 gene and the ITS1 of nrDNA, statistical analyses were performed using matched-pairs t-tests for paired groups;  $p$  values of  $<0.05$  were considered statistically significant. Maximum-likelihood relative rate tests were estimated using HyPhy version 2.10 [65]. Nucleotide substitution models were evaluated by hierarchical likelihood ratio tests implemented in Modeltest version 3.7 [66]. The Jukes and Cantor 1969 (JC69) model [67] was determined the best model by having the highest Bayesian Information Criterion (BIC) value. The relative rate test compares the number of nucleotide substitutions per site between two ingroup species by exploiting outgroups to classify those substitutions that can be



unambiguously assigned to one of the ingroup taxa [68]. To test significance between different regions of pre-miR159 genes, statistical analyses using t-tests for paired groups were performed.

## Supporting Information

**Figure S1. The hairpin secondary structure of the novel pre-miR159 (A) and canonical pre-miR159 (B) in *Phalaenopsis amabilis*.**

[doi:10.1371/journal.pone.0114493.s001](https://doi.org/10.1371/journal.pone.0114493.s001) (PDF)

**Figure S2. A close-up view of the secondary structure of the novel pre-miR159 from the 42 *Phalaenopsis* species.**

[doi:10.1371/journal.pone.0114493.s002](https://doi.org/10.1371/journal.pone.0114493.s002) (PDF)

**Figure S3. The hairpin secondary structure of the novel pre-miR159 from (A) *Phalaenopsis minus*; and (B) *P. sumatrana*-type 2 with a 10 nt deletion within the terminal loop region.** The blue line region represents the fold-back arm of secondary structure for all *Phalaenopsis* species. The red line region represents the mature miR159.

[doi:10.1371/journal.pone.0114493.s003](https://doi.org/10.1371/journal.pone.0114493.s003) (PDF)

**Table S1. The maximum-likelihood relative rate test in the novel pre-miR159 gene of 42 *Phalaenopsis* species.** Pairwise comparisons of Nucleotide substitution rate (above diagonal) and P-value (below diagonal) between species determined from the novel pre-miR159 gene.

[doi:10.1371/journal.pone.0114493.s004](https://doi.org/10.1371/journal.pone.0114493.s004) (PDF)

**Table S2. The maximum-likelihood relative rate test in the 5' Fold-back arm of the novel pre-miR159 gene of 42 *Phalaenopsis* species.** Pairwise comparisons of Nucleotide substitution rate (above diagonal) and P-value (below diagonal) between species deduced from the 5' Fold-back arm of the novel pre-miR159 gene.

[doi:10.1371/journal.pone.0114493.s005](https://doi.org/10.1371/journal.pone.0114493.s005) (PDF)

**Table S3. The maximum-likelihood relative rate test in the terminal loop region of the novel pre-miR159 gene of 42 *Phalaenopsis* species.** Pairwise comparisons of Nucleotide substitution rate (above diagonal) and P-value (below diagonal) between species deduced from the terminal loop region of novel the pre-miR159 gene.

[doi:10.1371/journal.pone.0114493.s006](https://doi.org/10.1371/journal.pone.0114493.s006) (PDF)

**Table S4. The maximum-likelihood relative rate test in the 3' Fold-back arm of the novel pre-miR159 gene of 42 *Phalaenopsis* species.** Pairwise comparisons of nucleotide substitution rate (above diagonal) and P-value (below diagonal) between species deduced from the 3' Fold-back arm of the novel pre-miR159 gene.

[doi:10.1371/journal.pone.0114493.s007](https://doi.org/10.1371/journal.pone.0114493.s007) (PDF)

## Acknowledgments

We thank Dr. Katrina Bogan for revising, editing, and polishing our paper and two reviewers for their helpful comments.

## Author Contributions

Conceived and designed the experiments: CCT YCC CHC. Performed the experiments: CCT YCC CHC. Analyzed the data: CCT YCC CHC. Contributed reagents/materials/analysis tools: ISW YSL. Wrote the paper: CCT YCC CHC.

## References

1. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116: 281–297.
2. Kurihara Y, Watanabe Y (2004) *Arabidopsis* micro-RNA biogenesis through Dicer-like 1 protein functions. *Proc Natl Acad Sci USA* 101: 12753–12758.
3. Parizotto EA, Dunoyer P, Rahm N, Himber C, Vionnet O (2004) In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes Dev* 18: 2237–2242.
4. Jones-Rhoades MW, Bartel DP, Bartel B (2006) MicroRNAs and their regulatory roles in plants. *Annu Rev Plant Biol* 57: 19–53.
5. Kurihara Y, Takashi Y, Watanabe Y (2006) The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. *RNA* 12: 206–212.
6. Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, et al. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* 115: 199–208.
7. Rhoades MW, Reinhart BJ, Lim LP, Burge CB, Bartel B, et al. (2002) Prediction of plant microRNA targets. *Cell* 110: 513–520.
8. Meyers BC, Souret FF, Lu C, Green PJ (2006) Sweating the small stuff: microRNA discovery in plants. *Curr Opin Biotechnol* 17: 139–146.
9. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP (2002) MicroRNAs in plants. *Genes Dev* 16: 1616–1626.
10. Jones-Rhoades MW, Bartel DP (2004) Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Mol Cell* 14: 787–799.
11. Fahlgren N (2007) High-throughput sequencing of *Arabidopsis* microRNAs: Evidence for frequent birth and death of MIRNA genes. *PLoS ONE* 2: e219.
12. Moxon S, Jing R, Szittyá G, Schwach F, Pilcher RLR, et al. (2008) Deep sequencing of tomato short RNAs identifies microRNAs targeting genes involved in fruit ripening. *Genome Res* 18: 1602–1609.
13. Cuperus JT, Fahlgren N, Carrington JC (2011) Evolution and functional diversification of miRNA genes. *Plant Cell* 23: 431–442.
14. Rajagopalan R, Vaucheret H, Trejo J, Bartel DP (2006) A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes Dev* 20: 3407–3425.
15. Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, et al. (2008) Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics* 9: 160.
16. Kulcheski FR, de Oliveira LF, Molina LG, Almerao MP, Rodrigues FA, et al. (2011) Identification of novel soybean microRNAs involved in abiotic and biotic stresses. *BMC Genomics* 12: 307.
17. Zhao CZ, Xia H, Frazier TP, Yao YY, Bi YP, et al. (2010) Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). *BMC Plant Biol* 10: 3.

18. **Unver T, Bakar M, Shearman RC, Budak H** (2010) Genome-wide profiling and analysis of *Festuca arundinacea* miRNAs and transcriptomes in response to foliar glyphosate application. *Mol Genet Genomics* 283: 397–413.
19. **Xu Q, Liu Y, Zhu A, Wu X, Ye J, et al.** (2010) Discovery and comparative profiling of microRNAs in a sweet orange red-flesh mutant and its wild type. *BMC Genomics* 11: 246.
20. **Reyes JL, Chua NH** (2007) ABA induction of miR159 controls transcript levels of two MYB factors during *Arabidopsis* seed germination. *Plant J* 49: 592–606.
21. **Achard P, Herr A, Baulcombe DC, Harberd NP** (2004) Modulation of floral development by a gibberellin-regulated microRNA. *Development* 131: 3357–3365.
22. **Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, et al.** (2005) Specific effects of microRNAs on the plant transcriptome. *Dev Cell* 8: 517–527.
23. **Li Y, Li C, Ding G, Jin Y** (2011) Evolution of MIR159/319 microRNA genes and their post-transcriptional regulatory link to siRNA pathways. *BMC Evol Biol* 11: 122.
24. **Palatnik JF, Wollmann H, Schommer C, Schwab R, Boisbouvier J, et al.** (2007) Sequence and expression differences underlie functional specialization of arabidopsis microRNAs miR159 and miR319. *Dev Cell* 13: 115–125.
25. **Warthmann N, Das S, Lanz C, Weigel D** (2008) Comparative analysis of the MIR319a microRNA locus in *Arabidopsis* and related Brassicaceae. *Mol Biol Evol* 25: 892–902.
26. **de Meaux J, Hu JY, Tartler U, Goebel U** (2008) Structurally different alleles of the ath-MIR824 microRNA precursor are maintained at high frequency in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 105: 8994–8999.
27. **Contreras-Cubas C, Rabanal FA, Arenas-Huertero C, Ortiz M, Covarrubias AA, et al.** (2012) The *Phaseolus vulgaris* miR159a precursor encodes a second differentially expressed microRNA. *Plant Mol Biol* 80: 103–115.
28. **Zeng Y, Cullen BR** (2003) Sequence requirements for microRNA processing and function in human cells. *RNA* 9: 112–123.
29. **Bologna NG, Mateos JL, Bresso EG, Palatnik JF** (2009) A loop-to-base processing mechanism underlies the biogenesis of plant microRNAs miR319 and miR159. *EMBO J* 28: 3646–3656.
30. **Christenson EA** (2001) *Phalaenopsis*: A Monograph. Timber Press, Portland, OR, USA, pp.27–35.
31. **Tsai CC, Huang SC, Chou CH** (2006) Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the internal transcribed spacer of the nuclear ribosomal DNA. *Plant Syst Evol* 256: 1–16.
32. **Tsai CC, Chiang YC, Huang SC, Chen CH, Chou CH** (2010) Molecular phylogeny of *Phalaenopsis* Blume (Orchidaceae) based on the plastid and nuclear DNA. *Plant Syst Evol* 288: 77–98.
33. **Axtell MJ, Snyder JA, Bartel DP** (2007) Common functions for diverse small RNAs of land plants. *Plant Cell* 19: 1750–1769.
34. **Arenas-Huertero C, Perez B, Rabanal F, Blanco-Melo D, De la Rosa C, et al.** (2009) Conserved and novel miRNAs in the legume *Phaseolus vulgaris* in response to stress. *Plant Mol Biol* 70: 385–401.
35. **Suzuki MT, Giovannoni SJ** (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *App Environ Microbiol* 62: 625–630.
36. **Kimura M, Takahata N** (1983) Selective constraint in protein polymorphism: study of the effectively neutral mutation model by using an improved pseudosampling method. *Proc Natl Acad Sci USA* 80: 1048–1052.
37. **Guo X, Wang Y, Keightley PD, Fan L** (2007) Patterns of selective constraints in noncoding DNA of rice. *BMC Evol Biol* 7: 208.
38. **Chen CN, Chiang YC, Ho THD, Schaal BA, Chiang TY** (2004) Coalescent processes and relaxation of selective constraints leading to contrasting genetic diversity at paralogs AtHVA22d and AtHVA22e in *Arabidopsis thaliana*. *Mol Phyl Evol* 32: 616–626.
39. **Chiang YC, Schaal BA, Ge XJ, Chiang TY** (2004) Range expansion leading to departures from neutrality in the nonsymbiotic hemoglobin gene and the cpDNA *trnL-trnF* intergenic spacer in *Trema dielsiana* (Ulmaceae). *Mol Phyl Evol* 31: 929–942.

40. Mathews DH, Sabina J, Zuker M, Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 288: 911–940.
41. Song L, Axtell MJ, Fedoroff NV (2010) RNA Secondary Structural Determinants of miRNA Precursor Processing in *Arabidopsis*. *Curr Biol* 20: 37–41.
42. Werner S, Wollmann H, Schneeberger K, Weigel D (2010) Structure determinants for accurate processing of miR172a in *Arabidopsis thaliana*. *Curr Biol* 20: 42–48.
43. Liu Q, Yan Q, Liu Y, Hong F, Sun Z, et al. (2013) Complementation of Hyponastic Leaves1 by Double-Strand RNA-Binding Domains of Dicer-Like1 in Nuclear Dicing Bodies. *Plant Physiol* 163: 108–117.
44. Dezulian T, Palatnik JF, Huson D, Weigel D (2005) Conservation and divergence of microRNA families in plants. *Genome Biol* 6: P13.
45. An FM, Hsiao SR, Chan MT (2011) Sequencing-based approaches reveal low ambient temperature-responsive and tissue-specific microRNAs in *Phalaenopsis* orchid. *PLoS One*, 6: e18937
46. Lu J, Shen Y, Wu Q, Kumar S, He B, et al. (2008) The birth and death of microRNA genes in *Drosophila*. *Nat Genet* 40: 351–355.
47. Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. *Proc Natl Acad Sci USA* 104: 3300–3305.
48. Chao YT, Su CL, Jean WH, Chen WC, Chang YCA, et al. (2014) Identification and characterization of the microRNA transcriptome of a moth orchid *Phalaenopsis aphrodite*. *Plant Mol Biol* 84: 529–548.
49. Chen K, Rajewsky N (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 38: 1452–1456.
50. Lee EJ, Baek M, Gusev Y, Brackett DJ, Nuovo GJ, et al. (2008) Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA* 14: 35–42.
51. Choudhury NR, de Lima Alves F, de Andres-Aguayo L, Graf T, Caceres JF, et al. (2013) Tissue-specific control of brain-enriched miR-7 biogenesis. *Genes Dev* 27: 24–38.
52. Luciano DJ, Mirsky H, Vendetti NJ, Maas S (2004) RNA editing of a miRNA precursor. *RNA* 10: 1174–1177.
53. Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, et al. (2006) RNA editing of human microRNAs. *Genome Biol* 7: R27.
54. de Hoon MJ, Taft RJ, Hashimoto T, Kanamori-Katayama M, Kawaji H, et al. (2010) Cross-mapping and the identification of editing sites in mature microRNAs in high-throughput sequencing libraries. *Genome Res* 20: 257–264.
55. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11–15.
56. Xie Z, Allen E, Fahlgren N, Calamar A, Givan SA, et al. (2005) Expression of *Arabidopsis* MIRNA genes. *Plant Physiol* 138: 2145–2154.
57. Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* 120: 621–623.
58. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41: 95–98.
59. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31: 3406–3415.
60. Rozas J, Sanchez-De I, Barrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
61. Xia X, Xie Z, Salemi M, Chen L, Wang Y (2003) An index of substitution saturation and its application. *Mol Phy Evol* 26: 1–7.
62. Xia X, Lemey P (2009) Assessing substitution saturation with DAMBE. In *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*, 2nd ed.; , Lemey P, Salemi M, Vandamme AM, , Eds. , ; Cambridge University Press: Cambridge, UK, pp. 611–626.
63. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. *J Hered* 92: 371–373.

64. **Gojobori T, Ishii K, Nei M** (1982) Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J Mol Evol* 18: 414–423.
65. **Pond SLK, Frost SDW, Muse SV** (2005) Hypothesis testing using phylogenetics (HyPhy). *Bioinformatics* 21: 676–679.
66. **Posada D, Crandall KA** (1998) ModelTest: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
67. **Jukes TH, Cantor CR** (1969) Evolution of Protein Molecules. In *Mammalian Protein Metabolism*; Munro HN Eds.; Academic Press: New York, USA, pp. 21–32.
68. **Nei M, Kumar S** (2000) *Molecular Evolution and Phylogenetic*; Oxford University Press: Oxford, UK, pp. 191–195.