Diversity of Genome Size and Ty1-copia in *Epimedium* Species Used for Traditional Chinese Medicines

Jianjun Chen

Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China

Lijia Li

Key Laboratory of MOE for Plant Development Biology, College of Life Science, Wuhan University, Wuhan, Hubei 430072, China

Ying Wang¹

Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei 430074, China

Additional index words. *Epimedium*, genome size, Ty1-copia, NJ phylogenetic tree, FISH

Abstract. *Epimedium* species are traditional Chinese medicinal plants as well as potential groundcover and ornamental plants. In this study, genome size and genome structures of *Epimedium* species were investigated using flow cytometric and fluorescence in situ hybridization (FISH). The nuclear DNA content of *Epimedium* species ranged from 8.42 pg/2C (8230.7 Mbp) to 9.97 pg/2C (9752.8 Mbp). The pairwise nucleotide diversity (π) of the fragments of the genes for reverse transcriptase (*rt*) of Ty1-copia retrotransposon within a species of *rt* fragments ranged from 0.251 to 0.428 in 10 *Epimedium* species. Phylogenetic analysis of the sequences revealed four major clades with the largest subclade containing 72 sequences of relatively low nucleotide diversity. FISH indicated that Ty1-copia retrotransposons are distributed unevenly along the pachytene chromosomes of *E. wushanense* and *E. sagittatum*, mostly associated with the pericentromeric and terminal heterochromatin. The relatively low sequence heterogeneity of Ty1-copia *rt* sequences implies that the *Epimedium* genomes have experienced a few relatively large-scale proliferation events of copia elements, which could be one of the major forces resulting in the large genome size of *Epimedium* species.

Epimedium L. ($2n = 2x = 12$), referred to as yin yang huo in Chinese, belongs to the basal eudicot plant family, berberidaceae. The genus of *Epimedium* is composed of more than 50 species (Stearn, 2002), most of which are widely distributed in China and commonly used as traditional Chinese medicinal herbs (Ying, 2002) and as ornamental plants (Stearn, 2002). In particular, *Epimedium* species are of

great interest because of their pharmacological properties in the treatment of impotence, spermatorrhea, infertility, amenorrhea, and in improving menopause symptoms (Wu et al., 2003). In China, Herba Epimedii is usually comacerated in wine with other traditional medicines contributing to prevent disease and strengthen immunity (Ma et al., 2011). Four flavonoids, epimedin A, epimedin B, epimedin C, and icarrin, were believed as the major active components in *Epimedium* and were regarded as markers for quality control (Chen et al., 2007; Xie and Sun, 2006; Xie et al., 2010). Five species are officially recorded as medicinal plants in the Chinese Pharmacopoeia, including *E. brevicornum* Maxim, *E. sagittatum* (Sieb. et Zucc), *E. pubescens* Maxim, *E. wushanense* T. S. Ying, and *E. koreanum* Nakai (Chinese Pharmacopoeia Commission, 2005).

Repetitive DNA is the major components of plant genomes (Kubis et al., 1998), which is the primary determinant of genome size and structure and plays an important role in genome evolution. Plant genome size differs as a result of variable amounts of repetitive DNA (Bennett and Leitch, 2011; Flavell et al., 1974). Polyploidization, unequal recombination, and illegitimate recombination leading to plant genome expansion and contraction may be the major driving forces for plant genome size variation (Bennetzen, 2002). Moreover, retrotransposon insertions through a "copy and paste" mechanism also can increase host genome size rapidly (Hawkins et al., 2006; Piegu et al., 2006).

Known as the most abundant repetitive DNA, plant transposable elements (TEs) are classified as RNA-mediated TEs (Class 1)

and DNA-mediated TEs (Class 2) according to their transposition intermediate (Feschotte et al., 2002). Class 1 TEs are divided into long terminal repeat retrotransposons (LTR) and non-LTR retrotransposons. LTR retrotransposons can be further classified as Ty1-copia and Ty3-gypsy elements based on the order of their coding domains. Ty1-copia group retrotransposons have been shown to be present throughout almost all plant genomes with high copy numbers (Flavell et al., 1992a). Sequence analyses of polymerase chain reaction (PCR) fragments of reverse transcriptase conserved domains have revealed very high degrees of sequence heterogeneity in many plants (Flavell et al., 1992b; Kumar et al., 1997). Phylogenetic studies of these LTR retrotransposon families provide information of unknown genomic components and suggest causes of genome size variation.

Most recent studies on *Epimedium* have concentrated on its chemical composition (Jiang et al., 2009; Wu et al., 2011; Zhao et al., 2008), pharmacological properties (Wong et al., 2009), phylogenetic relationship (Sun et al., 2005), karyotype (Sheng et al., 2010; Zhang et al., 2008), and genetic diversity (Xu et al., 2007; Zhou et al., 2007). Despite its great potential value, genomic characteristics, including genome size and genome structure of *Epimedium*, have received little attention. Genomic analyses of *Epimedium* will provide basic information on genome duplication, speciation, and its complex metabolism. In this study, we aimed to characterize the *Epimedium* genome in terms of the nuclear DNA contents, the sequence diversity, and genomic distribution of Ty1-copia retrotransposons.

Received for publication 23 Jan. 2012. Accepted for publication 9 May 2012.

This paper was part of the colloquium "Research Highlights and Commercial Application of Medicinal Plants" held 27 Sept. 2011 at the ASHS Conference, Waikoloa, HI, and sponsored by the Working Group of Asian Horticulture (WGAH) and the Association of Horticulturists of Indian Origin (AHIO).

This work was supported by the National Natural Science Foundation of China (No. 30800624), CAS/SAFEA International Partnership Program for Creative Research Teams Project, Knowledge Innovation Project of The Chinese Academy of Sciences (KSCX2-EW-J-20), and Distinguished Young Scientist Project in Hubei (2009CDA073). We thank Dr. J. Dolezel for providing seeds of *Vicia faba* cv. Inovec, Dr. Jianfang Gui (Institute of Hydrobiology, Chinese Academy of Sciences) and Dr. Yan Wang (Wuhan University Medicine Division) for sharing flow cytometry, Dr. Yanqin Xu and Ms. Xuejun Zhang for their assistance in material collecting, Dr. Andrew Flavell for helpful advices on designing degenerate primers, and Dr. Alice Hayward and Joao Loureiro for helpful comments on our article. We also gratefully acknowledge the valuable comments by three anonymous reviewers.

¹To whom reprint requests should be addressed; e-mail yingwang@wbgcas.cn.

MATERIALS AND METHODS

Plant materials. Wild populations of 20 to 50 plants of *Epimedium* species were collected from various locations (Tables 1 and 2) and grown in soil outside at Wuhan Botanical Garden, Chinese Academy of Sciences. Young leaves grown 2 weeks of seven species were used for flow cytometric analysis (Table 1). The Ty1-*copia* reverse transcriptase (*rt*) fragments were cloned from 10 species (Table 2). The species of *E. acuminatum*, *E. dolichostemon*, *E. pubescens*, *E. sagittatum*, and *E. wushanense* were included in both studies.

Flow cytometric analysis. Relative DNA content was determined using propidium iodide-stained flow cytometry (Dolezel et al., 2007). The analysis was performed by an EPICS[®] ALTRA[™] flow cytometer (BECKMAN COULTER; Brea, CA) equipped with a water-cooled argon ion laser emitting at 488 nm for flow cytometry. Young leaves of the analyzed individuals and a reference standard were co chopped with a razor blade in a glass petri dish containing 0.5 mL of ice-cold MgSO₄ buffer [9.53 mM MgSO₄·7H₂O, 47.67 mM KCl, 4.77 mM HEPES, 6.48 mM dithiothreitol, 0.25% (w/v) Triton X-100; pH 8.0] (Arumuganathan and Earle, 1991). The crude nuclei suspension was filtered through a 50-μm nylon mesh. Subsequently, 0.5 mL solution containing RNase and propidium iodide (both 50 μg·mL⁻¹) was added. After incubation for 10 min at room temperature, relative fluorescence intensity of nuclei

was analyzed. Young leaves of *Vicia faba* cv. Inovec (2C = 26.9 pg) were used as the internal standards. At least 5000 nuclei were analyzed in each measurement and cv values were below 5.0%. Values obtained were converted to Mbp of nucleotides by the Eq. 1 pg = 978 Mbp (Dolezel et al., 2003).

For a given species, 10 individuals were randomly selected from each population for the flow cytometric analysis. Differences in DNA content between species or populations were tested by one-way analysis of variance, and multiple comparison tests were determined by the Tukey's honestly significant difference test ($P \leq 0.05$) with SSPS 13.0 for Windows (IBM, Chicago, IL).

Cloning of reverse transcriptases and phylogenetic analysis. Total genomic DNA for PCR amplification was extracted from fresh young leaves using the CTAB method (Doyle and Doyle, 1987). The reverse transcriptase (*rt*) domain of Ty1-*copia* retrotransposons was amplified from genomic DNA using degenerate PCR with primers designed against conserved residues from RT domains of retrotransposons. Primer sequences were 5'-ACNGCNTTYTNCAYGG-3', encoding the peptide TAFHGG (amino-terminal), and 5'-ARCATRTCTCNACRTA-3', encoding YVDDML (carboxy-terminal), where R = A/G, Y = C/T, H = A/T/C, and N = A/G/C/T (Kumar et al., 1997). PCR was carried out in 50 μL containing 100 to 200 ng genomic DNA, 1.5 mM MgCl₂, 1 × PCR buffer (10 mM Tris-HCl, pH 8.8, 50 mM KCl), 200 μM dNTPs, 3 μM of each degenerate primer,

and 2 U *Taq* DNA polymerase (Fermentas, Burlington, Ontario, Canada) on a Mastercycler[®] gradient PCR cyclor (Eppendorf; Hamburg-Nord, Hamburg, Germany). The PCR cycling profile included an initial denaturation step at 94 °C for 5 min followed by 35 cycles of 30 s at 94 °C, 50 s annealing at 47 °C, and 45 s at 70 °C with a final extension step at 72 °C for 10 min.

Purified PCR products (E. Z. N. A.[®] Gel Extraction Kit; Omega Bio-tek, Norcross, GA) were ligated into the pMD18-T plasmid vector (TaKaRa Bio; Dalian, Liaoning, China) and transformed into competent *Escherichia coli* (strain DH5α). Clones with positive inserts were detected by PCR using M13 primers. Twenty to 25 positive clones of each species were selected randomly and sequenced using an ABI 3100 (Applied Biosystems) by Shanghai Invitrogen Biotech Co. Ltd.

BLASTX searches of the cloned sequences against the National Center for Biotechnology Information plant protein database were performed using default parameters. All sequences were translated into peptides using BioEdit 7.0.1 to find conserved residues (Hall, 1999). Nucleotide diversity π (π), between the sets of sequences from each species including the primer regions was calculated using DnaSP 4.10 (Rozas et al., 2003). It is given by the following formula:

$$\pi = \sum_{ij} x_i x_j \pi_{ij}$$

where x_i and x_j are the respective frequencies of the i th and j th sequences and π_{ij} is the proportion of the non-identical nucleotides between the i th and j th sequences.

All phylogenetic analyses were carried out in MEGA 3 (Kumar et al., 2004). Neighbor-joining (NJ) analyses of the nucleotide data sets were conducted using p -distances with 1000 bootstrap iterations.

Fluorescence in situ hybridization analysis. Young panicles of *E. wushanense* and *E. sagittatum* were harvested and fixed in 3:1 (100% ethanol:glacial acetic acid) Carnoy's solution. Microsporocytes at the pachytene stage were incubated in mixed

Table 1. Nuclear 2C DNA content of *Epimedium* species using *V. faba* cv. Inovec (2C = 26.9 pg) as an internal standard.

Species	Locality	DNA content ± SD (pg/2C)	2C genome size (Mbp) ^y
<i>E. acuminatum</i>	Wufeng, Hubei	8.42 ± 0.15 a ^z	8230.7 ± 151.4
<i>E. boreali-guizhouense</i>	Yanhe, Guizhou	8.70 ± 0.62 b	8510.5 ± 610.1
<i>E. dolichostemon</i>	Lichuan, Hubei	9.05 ± 0.08 c	8853.2 ± 78.2
<i>E. franchetii</i>	Shennongjia, Hubei	8.42 ± 0.30 a	8238.9 ± 297.0
<i>E. pubescens</i>	Dujiangyan, Sichuan	9.28 ± 0.21 d	9076.4 ± 202.2
<i>E. sagittatum</i>	Huangshan, Anhui	8.87 ± 0.52 b	8678.3 ± 504.9
<i>E. wushanense</i>	Badong, Hubei	9.40 ± 0.09 e	9197.3 ± 91.7

^zThe same letter indicates no significant difference at $P \leq 0.05$.

^y1 pg = 978 Mbp.

Table 2. List of taxa, collection details, voucher information, clones analyzed, nucleotide diversity, and GenBank accession numbers of *rt* sequences of Ty1-*copia* isolated from *Epimedium* species.

Species	Locality (voucher) ^z	Number of sequences			Nucleotide diversity π (π) ^v	GenBank accession numbers
		Total ^y	Indels/stop ^x	Full length ^w		
<i>E. acuminatum</i>	Wufeng, Hubei (WBG: YX and JC WF1-18)	23	6	17	0.251 (l)	GQ852896-GQ852915
<i>E. coactum</i>	Jianhe, Guizhou (WBG: XZ Coa2-12)	22	11	11	0.351 (m)	GQ852879-GQ852895
<i>E. diphyllum</i>	Sanhui, Guizhou (WBG: XZ Ce940544)	18	10	8	0.381 (m)	GQ852972-GQ852983
<i>E. dolichostemon</i>	Lichuan, Hubei (WBG: YX and JC LC2-12)	22	10	12	0.378 (m)	GQ852933-GQ852952
<i>E. ecalcaratum</i>	Songtao, Guizhou (WBG: XZ 09172-8)	20	10	10	0.391 (m)	GQ852953-GQ852971
<i>E. koreanum</i>	Fusong, Liaoning (WBG: YX and JZ CY8-23)	19	7	12	0.428 (h)	GQ853003-GQ853016
<i>E. leptorrhizum</i>	Lichuan, Hubei (WBG: YX and JC LC1-11)	21	7	14	0.276 (l)	GQ852916-GQ852932
<i>E. pubescens</i>	Dujiangyan, Sichuan (WBG: YX, XZ, and JC SC1-46)	21	5	16	0.364 (m)	GQ852984-GQ853002
<i>E. sagittatum</i>	Huangshan, Anhui (WBG: YX and XZ HS1-33)	25	12	13	0.342 (m)	GQ852860-GQ852878
<i>E. wushanense</i>	Badong, Hubei (WBG: YX WY234)	22	11	11	0.286 (l)	GQ852843-GQ852859

^zWBG = Wuhan Botanical Garden; Collectors: YX = Y. Xu; XZ = X. Zhang; JC = J. Chen; JZ = J. Zhou.

^yThe total number of sequences obtained.

^xNumber of sequences with disrupted reading frames or putative stop codons.

^wNumber of sequences without putative stop codons.

^vNucleotide diversity, π (π), the average number of nucleotide differences per site between two sequences: l = low (0.200 to 0.300), m = medium (0.300 to 0.400), h = high (> 0.400) nucleotide diversity.

enzymes (pectinase:macerozyme = 1:1) for 30 min and squashed in 45% glacial acetic acid solution. Slides were immediately dipped into liquid nitrogen. Coverslips were removed, and slides were stored at -70°C until use.

The PCR products of *copia* fragments from *E. wushanense* were labeled as probes using the DIG-Nick Translation Mix Kit (Roche, Basel, Switzerland) following the manufacturer's instructions. FISH was conducted according to Jiang et al. (1995). Digoxigenin-labeled probes were detected with sheep anti-digoxigenin (Roche) and amplified with fluorescein isothiocyanate-conjugated anti-sheep IgG (Vector Laboratories, Burlingame, CA). Chromosomes were counterstained with 30 μL (2 $\mu\text{g}\cdot\text{mL}^{-1}$) of 4',6-diamidino-phenylindole (DAPI; Sigma, St. Louis, MO) in an antifade solution (Vector Laboratories).

Chromosomes and FISH signal images were captured with an Olympus BX61 fluorescence microscope conjunct with a Photometrics SenSys CCD (charge-coupled device) 1400E. Gray-scale images were captured for each color channel and then merged using Metamorph software. Final image adjustments were processed using Adobe Photoshop CS.

RESULTS AND DISCUSSION

Genome size of *Epimedium* species. To avoid inaccurate estimates of genomic DNA content an MgSO_4 isolation buffer containing dithiothreitol was used to prevent interference on fluorescent DNA staining from the high level of secondary metabolites in leaves of *Epimedium*. DNA content was calculated based on the peaks with good cv values

showing very low background noise (Fig. 1). Nuclear DNA content of seven *Epimedium* species varied by more than 1.12-fold, from $2C = 8.42 \text{ pg}/8238.9 \text{ Mbp}$ (*E. acuminatum*) to $9.40 \text{ pg}/9197.3 \text{ Mbp}$ (*E. wushanense*) (Table 1), suggesting that interspecific genome size variation is continuous. Statistical analyses using the Tukey's honestly significant difference test revealed significant interspecies differences in genome sizes (Table 1).

Previous studies revealed the chromosome numbers of all *Epimedium* species to be $2n = 2x = 12$ (Zhang et al., 2008). However, in the present study, it was clarified that *Epimedium* species have large genome sizes in comparison with model plants such as *Arabidopsis thaliana* ($2n = 12$, $1C = 157 \text{ Mbp}$), *Oryza sativa* ($2n = 24$, $1C = 490 \text{ Mbp}$), *Medicago truncatula* ($2n = 16$, $1C = 466 \text{ Mbp}$), *Solanum lycopersicum* ($2n = 24$, $1C = 950 \text{ Mbp}$), and *Zea mays* ($2n = 20$, $1C = 2460 \text{ Mbp}$) (Bennett and Leitch, 2004). Moreover, mean genome sizes of the other two sister genera *Berberis* L. ($2n = 2x = 28$) and *Mahonia* Nutt. ($2n = 2x = 28$) within berberidaceae were 1.46 pg/C and 1.22 pg/C , respectively (Rounsaville and Ranney, 2010). By contrast, *Epimedium* has a large genome size divided into relatively few chromosomes. We searched the Plant C-value database (Bennett and Leitch, 2004) for other angiosperms having the same chromosome number ($2n = 12$) and comparable genome sizes of 4.0 to 5.0 pg/C . Similar genome configurations were found for *Vicia elegans* (4.08 pg/C), *V. incana* (4.10 pg/C), *V. cassubica* (4.13 pg/C), *V. dalmatica* (4.13 pg/C), *V. cuspidata* (4.63 pg/C), *V. ramulifera* (4.68 pg/C), *V. neglecta* (4.73 pg/C), *V. unijuga* (4.83 pg/C), *Crepis*

paludosa (4.16 pg/C), and *C. viscidula* (4.37 pg/C). Few chromosome numbers and relatively large genome is the common features in these species. The chromosomes in the Mexican axolotl, an amphibian species having relatively few chromosomes and a gigantic genome, arose from fission of ancestral chromosomes (Voss et al., 2011). Although there is no certain relationship between the chromosome number and DNA contents, analysis of the chromosome evolution history is a way to understand the causes resulting in few chromosomes and large genome in these specially plant species.

An abundance of repetitive elements is the most remarkable characteristic of large plant genomes (Kubis et al., 1998). Amplification of retrotransposon populations is associated with the large genome size of *Vicia* species (Hill et al., 2005; Neumann et al., 2006). This is also true in maize where the repetitive elements, especially retrotransposons, have played key roles in the evolution of the species (Baucom et al., 2009; Rabinowicz and Bennetzen, 2006). Therefore, it is expected that repetitive elements could be the major contributors to large genome size in *Epimedium* genomes.

Although it is not known whether complex metabolism correlates to increased genome sizes in medicinal plants, larger transcriptional regulation networks may be involved in synthesizing the diverse chemical components and responding to environmental stress. Therefore, it is essential to study function, diversity, and regulation of important plant metabolic pathways along with genomic analyses.

Isolation and characterization of *rt* sequences of *Ty1-copia* in *Epimedium* genomes. We

Fig. 1. Representative histograms of relative fluorescence intensity obtained using MgSO_4 isolation buffer after simultaneous analysis of nuclei isolated from *V. faba* cv. Inovec ($2C = 26.9 \text{ pg DNA}$, as an internal reference standard) and *Epimedium* species: (A) *E. sagittatum*, (B) *E. pubescens*, (C) *E. wushanense*, (D) *E. dolichostemon*, and (E) *E. acuminatum*. The mean channel number and cv (%) value of each peak is also given.

Fig. 2. Neighbor-joining (NJ) phylogenetic tree generated by the *rt* conserved sequences in 10 *Epimedium* species. Bootstrap probabilities more than 50% are shown. Sequences that possess frameshift mutations resulting from disrupted in this study were excluded from the analysis. Ea = *E. acuminatum*; Ec = *E. coactum*; Edi = *E. diphyllum*; Edo = *E. dolichostemon*; Ee = *E. ecalcaratum*; Ek = *E. koreanum*; El = *E. leptorrhizum*; Ep = *E. pubescens*; Es = *E. sagittatum*; Ew = *E. wushanense*.

Fig. 3. Fluorescence in situ hybridization (FISH) analysis of Ty1-*copia* retrotransposons spread on pachytene chromosomes of *E. wushanense* (A–C) and *E. sagittatum* (D–F). (A, D) 4',6-Diamidino-phenylindole (DAPI) staining chromosomes; (B, E) dispersed signal; (C, F) merged images. Bar = 10 μ m.

retrotransposons in many higher plant genomes (Belyayev et al., 2005; Flavell et al., 1992a; Hill et al., 2005; Park et al., 2007; Rogers and Pauls, 2000; Ruas et al., 2008). The correct target *rt* sequences in this work were acquired by PCR amplification. The downstream priming region specifying the peptide sequence YVDDML is well conserved, but the upstream primer region TAFLHG in Ty1-*copia* is much less conserved (Flavell et al., 1992a; Xiong and Eickbush, 1990). Our results are mostly consistent with these findings.

In the present study, the diversity of *copia-rt* fragments in *Epimedium* was surveyed using a relatively large number of sequences, i.e., ≈ 20 sequences for each species analyzed, which might adequately represent the diversity of *rt* fragments within the genome (Park et al., 2007; Ruas et al., 2008). Compared with *copia-rt* sequence heterogeneity in plants having relatively small genome sizes, $\pi = 0.76$ in *A. thaliana*, 0.54 in *O. sativa*, and 0.55 in *Z. mays* (Navarro-Quezada and Schoen, 2002), relatively low to medium sequence heterogeneity (0.251 to 0.428) of *rt* fragments was observed in *Epimedium* genomes using Jukes-Cantor-corrected estimates of nucleotide diversity (π). This fact suggests that a few events of amplification of retrotransposable elements amplification might be responsible for the large genome sizes of *Epimedium* species.

LTR retrotransposons are one of the main driving forces of plant genome evolution, influencing genome size and genome structure (Oliver and Greene, 2009; Oliver et al., 2007). According to the patterns of long-term changes in transposable elements, bursts of retrotransposable element activity leading to genomic expansions over a short time might have been followed by rapid DNA loss through unequal homologous recombination or illegitimate recombination (Le Rouzic et al., 2007; Vitte et al., 2007). Only one or a few LTR retrotransposon families have been responsible for doubling genome size in a short time in *Gossypium* (Hawkins et al., 2006) and *Oryza* (Piegu et al., 2006). A recent study has shown

isolated and characterized the *rt* sequences of Ty1-*copia* in *Epimedium* genomes to investigate the contribution of them to the genome sizes of *Epimedium* species. PCR amplification of the Ty1-*copia* *rt* sequences from the 10 *Epimedium* species yielded products of expected sizes of ≈ 270 bp. Amplified fragments were cloned and ≈ 20 randomly selected clones were sequenced for each species, resulting in a total of 213 non-redundant sequences. These sequences were deposited in GenBank under accession numbers GQ852843–GQ853016 (Table 2). Eighty-nine sequences (41.8%) containing indels and putative stop codons were identified across all species, suggesting these elements may have lost the ability to transpose autonomously. The sequence variation for each species was observed, and it was clarified that the intraspecific nucleotide diversity (π) ranged from 0.251 for *E. acuminatum* to 0.428 for *E. koreanum* (Table 2).

A phylogenetic tree was constructed for the 174 *copia* sequences without disrupted sequence using NJ methods. This revealed four major clades according to bootstrap (BS) values [Fig. 2; clades labeled A (BS 78), B (BS 99), C (BS 99), and D (BS 98)]. Most of the sequences with low nucleotide diversity were grouped into clade A, which was divided into five subclades [A1–A5 (BS 99 to all subclades); Fig. 2]. Subclades A1 contained 72 (41.4%) sequences from all species with a nucleotide diversity (π) of 0.048 (data not shown). Similar tree topological structures are found in the subclades A3 and D4 (Fig. 2). The sequences in *E. koreanum* with high sequence heterogeneity (0.428) were dispersed across the tree. Except for *E. koreanum*, the individual phylogenetic trees for each species showed a similar topological structure with most sequences grouped in clade A (Fig. S1).

A PCR-based approach has been used to amplify the *rt* sequence of the Ty1-*copia*

that regions of the maize genome with the highest LTR retrotransposon density exhibited the lowest LTR retrotransposon diversity (Baucom et al., 2009). Indeed, the degree of nucleotide sequence heterogeneity of Ty1-*copia* is rather variable among different plant genomes and is likely correlated to copy numbers of retrotransposons with high copy numbers associated with low heterogeneity (Flavell et al., 1992a; Navarro-Quezada and Schoen, 2002). The relatively low *copia-rt* sequence heterogeneity in *Epimedium* suggests that a few Ty1-*copia* retrotransposons have undergone massive proliferation events in recent evolutionary time, leading to rapidly increasing copy numbers and finally to increased genome size of *Epimedium*.

Chromosomal localization of Ty1-*copia* retrotransposons in *Epimedium* genomes. The abundance and distribution of Ty1-*copia* elements in *E. wushanense* and *E. sagittatum* were investigated by FISH to pachytene chromosomes using digoxin-labeled *rt* sequences as probes that were amplified from the *E. wushanense* genome. Compared with metaphase-FISH, pachytene-FISH provides increased resolution (De Jong et al., 1999). Our analysis demonstrated that Ty1-*copia* retrotransposons spread unevenly along the pachytene chromosomes of *E. wushanense* and *E. sagittatum* and concentrate mostly at the pericentromeric and telomeric heterochromatin regions, which are highly stained by DAPI (Fig. 3).

The FISH results suggest that Ty1-*copia* elements make up a major fraction of *Epimedium* genomes. Similar chromosomal localizations of retrotransposons were revealed in other plant species (Benabdelmouna and Darmency, 2003; Heslop-Harrison et al., 1997; Lamb et al., 2007; Pearce et al., 1997). For example, clusters of signals of repetitive DNA sequences in *Arabidopsis* were not evenly dispersed on all chromosome pairs but were found in particular regions of the paracentromeric heterochromatin (Brandes et al., 1997). BAC sequences containing Ty3-*gypsy*-like retrotransposons from tomato were shown to spread throughout the pericentromeric heterochromatin (Wang et al., 2006). The enrichment of retrotransposons in heterochromatin might be partly the result of targeted bias integration, in which the heterochromatic mark triggers epigenetic modification by RNA interference or methylation, leading to transposon inactivation (Gao et al., 2008; Lippman et al., 2004). The chromatic distribution pattern of *copia* retrotransposable elements in *Epimedium* also suggests that they could play a key role in establishing and maintaining heterochromatin structure.

CONCLUSION

The present study is the first to report on the large genome sizes of the traditional Chinese medicinal *Epimedium* species, which have been valued pharmacologically for the secondary metabolite composition. The large genome sizes of *Epimedium* species make it relatively difficult for whole genome

sequencing. It was shown in the present study that high copy number of *copia* retrotransposable elements with low sequence heterogeneity comprises a major portion of *Epimedium* genomes. The abundance, diversity, and distribution of these elements make it a good choice to develop molecular markers useful for detecting genetic diversity in natural populations and to construct genetic linkage maps. Further study will be conducted for isolating the complete LTRs and developing retrotransposon-based molecular markers for these medicinal plants.

Literature Cited

- Arumuganathan, K. and E.D. Earle. 1991. Estimation of nuclear DNA content of plants by flow cytometry. *Plant Mol. Biol. Rpt.* 9:229–241.
- Baucom, R.S., J.C. Estill, C. Chaparro, N. Upshaw, A. Jogi, J.M. Deragon, R.P. Westerman, P.J. Sannmiquel, and J.L. Bennetzen. 2009. Exceptional diversity, non-random distribution and rapid evolution of retroelements in the B73 Maize genome. *PLoS Genet.* 5:e1000732.
- Belyayev, A., O. Raskina, and E. Nevo. 2005. Variability of the chromosomal distribution of Ty3-*gypsy* retrotransposons in the populations of two wild Triticeae species. *Cytogenet. Genome Res.* 109:43–49.
- Benabdelmouna, A. and H. Darmency. 2003. *Copia*-like retrotransposons in the genus *Setaria*: Sequence heterogeneity, species distribution and chromosomal organization. *Plant Syst. Evol.* 237:127–136.
- Bennett, M.D. and I.J. Leitch. 2004. Plant DNA C-values database (release 3.0, Dec. 2004). <<http://www.kew.org/cvalues/>>.
- Bennett, M.D. and I.J. Leitch. 2011. Nuclear DNA amounts in angiosperms: Targets, trends and tomorrow. *Ann. Bot. (Lond.)* 107:467–590.
- Bennetzen, J.L. 2002. Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115:29–36.
- Brandes, A., H. Thompson, C. Dean, and J.S. Heslop-Harrison. 1997. Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. *Chromosome Res.* 5:238–246.
- Chen, X.J., B.L. Guo, S.P. Li, Q.W. Zhang, P.F. Tu, and Y.T. Wang. 2007. Simultaneous determination of 15 flavonoids in *Epimedium* using pressurized liquid extraction and high-performance liquid chromatography. *J. Chromatography* 1163:96–104.
- Chinese Pharmacopoeia Commission. 2005. Pharmacopoeia of the People's Republic of China. Chemical Industry Press, Beijing, China.
- De Jong, J.H., P. Fransz, and P. Zabel. 1999. High resolution FISH in plants - techniques and applications. *Trends Plant Sci.* 4:258–263.
- Dolezel, J., J. Bartos, H. Voglmayr, and J. Greilhuber. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry* 51A:127–128.
- Dolezel, J., J. Greilhuber, and J. Suda. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protoc.* 2:2233–2244.
- Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–15.
- Feschotte, C., N. Jiang, and S.R. Wessler. 2002. Plant transposable elements: Where genetics meets genomics. *Nat. Rev. Genet.* 3:329–341.
- Flavell, A.J., E. Dunbar, R. Anderson, S.R. Pearce, R. Hartley, and A. Kumar. 1992a. Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res.* 20:3639–3644.
- Flavell, A.J., D.B. Smith, and A. Kumar. 1992b. Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. *Mol. Gen. Genet.* 231:233–242.
- Flavell, R.B., M.D. Bennett, J.B. Smith, and D.B. Smith. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12:257–269.
- Gao, X., Y. Hou, H. Ebina, H.L. Levin, and D.F. Voytas. 2008. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res.* 18:359–369.
- Hall, T.A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp.* 41:95–98.
- Hawkins, J.S., H. Kim, J.D. Nason, R.A. Wing, and J.F. Wendel. 2006. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res.* 16:1252–1261.
- Heslop-Harrison, J.S., A. Brandes, S. Taketa, T. Schmidt, A. Vershinina, E. Alkhimova, A. Kamm, R. Doudrick, T. Schwarzacher, A. Katsiotis, S. Kubis, A. Kumar, S. Pearce, A. Flavell, and G. Harrison. 1997. The chromosomal distributions of Ty1-*copia* group retrotransposable elements in higher plants and their implications for genome evolution. *Genetica* 100:197–204.
- Hill, P., D. Burford, D.M. Martin, and A.J. Flavell. 2005. Retrotransposon populations of *Vicia* species with varying genome size. *Mol. Genet. Genomics* 273:371–381.
- Jiang, F., X.L. Wang, N.L. Wang, and X.S. Yao. 2009. Two new flavonol glycosides from *Epimedium koreanum* Nakai. *J. Asian Nat. Prod. Res.* 11:401–409.
- Jiang, J., B.S. Gill, G. Wang, P.C. Ronald, and D.C. Ward. 1995. Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* 92:4487–4491.
- Kubis, S., T. Schmidt, and J.S. Heslop-Harrison. 1998. Repetitive DNA elements as a major component of plant genomes. *Ann. Bot. (Lond.)* 82:45–55.
- Kumar, A., S.R. Pearce, K. McLean, G. Harrison, J.S. Heslop-Harrison, R. Waugh, and A.J. Flavell. 1997. The Ty1-*copia* group of retrotransposons in plants: Genomic organization, evolution, and use as molecular markers. *Genetica* 100:205–217.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* 5:150–163.
- Lamb, J.C., J.M. Meyer, B. Corcoran, A. Kato, F. Han, and J.A. Birchler. 2007. Distinct chromosomal distributions of highly repetitive sequences in maize. *Chromosome Res.* 15:33–49.
- Le Rouzic, A., T.S. Boutin, and P. Capy. 2007. Long-term evolution of transposable elements. *Proc. Natl. Acad. Sci. USA* 104:19375–19380.
- Lippman, Z., A.V. Gendrel, M. Black, M.W. Vaughn, N. Dedhia, W.R. McCombie, K. Lavine, V. Mittal, B. May, K.D. Kasschau, J.C. Carrington, R.W. Doerge, V. Colot, and R. Martienssen. 2004. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476.
- Ma, H., X. He, Y. Yang, M. Li, D. Hao, and Z. Jia. 2011. The genus *Epimedium*: An ethnopharmacological and phytochemical review. *J. Ethnopharmacol.* 134:519–541.
- Navarro-Quezada, A. and D.J. Schoen. 2002. Sequence evolution and copy number of Ty1-*copia* retrotransposons in diverse plant genomes. *Proc. Natl. Acad. Sci. USA* 99:268–273.

- Neumann, P., A. Koblikova, A. Navratilova, and J. Macas. 2006. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. *Genetics* 173:1047–1056.
- Oliver, K.R. and W.K. Greene. 2009. Transposable elements: Powerful facilitators of evolution. *Bioessays* 31:703–714.
- Oliver, M.J., D. Petrov, D. Ackerly, P. Falkowski, and O.M. Schofield. 2007. The mode and tempo of genome size evolution in eukaryotes. *Genome Res.* 17:594–601.
- Park, J.M., G.M. Schneeweiss, and H. Weiss-Schneeweiss. 2007. Diversity and evolution of Ty1-*copia* and Ty3-*gypsy* retroelements in the non-photosynthetic flowering plants *Orobanchaceae* and *Phelipanche* (Orobanchaceae). *Gene* 387:75–86.
- Pearce, S.R., G. Harrison, P.J. Heslop-Harrison, A.J. Flavell, and A. Kumar. 1997. Characterization and genomic organization of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). *Genome* 40:617–625.
- Piegu, B., R. Guyot, N. Picault, A. Roulin, A. Saniyal, H. Kim, K. Collura, D.S. Brar, S. Jackson, R.A. Wing, and O. Panaud. 2006. Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Res.* 16:1262–1269.
- Rabinowicz, P.D. and J.L. Bennetzen. 2006. The maize genome as a model for efficient sequence analysis of large plant genomes. *Curr. Opin. Plant Biol.* 9:149–156.
- Rogers, S.A. and K.P. Pauls. 2000. Ty1-*copia*-like retrotransposons of tomato (*Lycopersicon esculentum* Mill.). *Genome* 43:887–894.
- Rounsaville, T.J. and T.G. Ranney. 2010. Ploidy levels and genome sizes of *Berberis* L. and *Mahonia* Nutt. species, hybrids, and cultivars. *HortScience* 45:1029–1033.
- Rozas, J., J.C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Ruas, C.F., H. Weiss-Schneeweiss, T.F. Stuessy, M.R. Samuel, A. Pedrosa-Harand, K. Tremetsberger, P.M. Ruas, P.M. Schlüter, M.A. Ortiz Herrera, C. König, and N.I. Matzenbacher. 2008. Characterization, genomic organization and chromosomal distribution of Ty1-*copia* retrotransposons in species of *Hypochoeris* (Asteraceae). *Gene* 412:39–49.
- Sheng, M.Y., L.J. Wang, and X.J. Tian. 2010. Karyomorphology of eighteen species of genus *Epimedium* (Berberidaceae) and its phylogenetic implications. *Genet. Resources Crop Evol.* 57:1165–1176.
- Stearn, W.T. 2002. The genus *Epimedium* and other herbaceous berberidaceae including the genus *Podophyllum*. Timber Press, Portland, OR.
- Sun, Y., K.P. Fung, P.C. Leung, and P.C. Shaw. 2005. A phylogenetic analysis of *Epimedium* (Berberidaceae) based on nuclear ribosomal DNA sequences. *Mol. Phylogenet. Evol.* 35: 287–291.
- Vitte, C., O. Panaud, and H. Quesneville. 2007. LTR retrotransposons in rice (*Oryza sativa*, L.): Recent burst amplifications followed by rapid DNA loss. *BMC Genomics* 8:218.
- Voss, S.R., D.K. Kump, S. Putta, N. Pauly, A. Reynolds, R.J. Henry, S. Basa, J.A. Walker, and J.J. Smith. 2011. Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Res.* 21:1306–1312.
- Wang, Y., X. Tang, Z. Cheng, L. Mueller, J. Giovannoni, and S.D. Tanksley. 2006. Euchromatin and pericentromeric heterochromatin: Comparative composition in the tomato genome. *Genetics* 172:2529–2540.
- Wong, S.P., P. Shen, L. Lee, J. Li, and E.L. Yong. 2009. Pharmacokinetics of prenylflavonoids and correlations with the dynamics of estrogen action in sera following ingestion of a standardized *Epimedium* extract. *J. Pharm. Biomed. Anal.* 50:216–223.
- Wu, C., J. Zhang, T. Zhou, B. Guo, Y. Wang, and J. Hou. 2011. Simultaneous determination of seven flavonoids in dog plasma by ultra-performance liquid chromatography-tandem mass spectrometry and its application to a bioequivalence study of bioactive components in *Herba Epimedii* and Er-Xian Decoction. *J. Pharm. Biomed. Anal.* 54:186–191.
- Wu, H., E.J. Lien, and L.L. Lien. 2003. Chemical and pharmacological investigations of *Epimedium* species: A survey. *Prog. Drug Res.* 60:1–57.
- Xie, J.P. and W.J. Sunday. 2006. Progress of chemical materials and pharmacy of genus *Epimedium* plants. *Strait Pharm. J.* 18:17–20.
- Xie, P.S., Y.Z. Yan, B.L. Guo, C.W.K. Lam, S.H. Chui, and Q.X. Yu. 2010. Chemical pattern-aided classification to simplify the intricacy of morphological taxonomy of *Epimedium* species using chromatographic fingerprinting. *J. Pharm. Biomed. Anal.* 52:452–460.
- Xiong, Y. and T.H. Eickbush. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9: 3353–3362.
- Xu, Y., Z. Li, Y. Wang, and H. Huang. 2007. Allozyme diversity and population genetic structure of three medicinal *Epimedium* species from Hubei. *J. Genet. Genomics* 34:56–71.
- Ying, T.S. 2002. Petal evolution and distribution patterns of *Epimedium* L. (Berberidaceae). *Acta Phytotaxon. Sin.* 40:481–489 [in Chinese].
- Zhang, Y.J., H.S. Dang, A.P. Meng, J.Q. Li, and X.D. Li. 2008. Karyomorphology of *Epimedium* (Berberidaceae) and its phylogenetic implications. *Caryologia* 61:283–293.
- Zhao, H.Y., J.H. Sun, M.X. Fan, L. Fan, L. Zhou, Z. Li, J. Han, B.R. Wang, and D.A. Guo. 2008. Analysis of phenolic compounds in *Epimedium* plants using liquid chromatography coupled with electrospray ionization mass spectrometry. *J. Chromatography* 1190:157–181.
- Zhou, J., Y. Xu, H. Huang, and Y. Wang. 2007. Identification of microsatellite loci from *Epimedium koreanum* and cross-species amplification in four species of *Epimedium* (Berberidaceae). *Mol. Ecol. Notes* 7:467–470.