BMC Genomics

RESEARCH ARTICLE Open Access

Satellite DNA in *Paphiopedilum* subgenus *Parvisepalum* as revealed by high-throughput sequencing and fluorescent in situ hybridization

Yung-I Lee[1,2*], Jing Wei Yap[3,4,5], Shairul Izan[6,7], Ilia J. Leitch[8], Michael F. Fay[4,9], Yi-Ching Lee[1], Oriane Hidalgo[4], Steven Dodsworth[4], Marinus J. M. Smulders[6], Barbara Gravendeel[10,11,12] and Andrew R. Leitch[3]

Abstract

Background: Satellite DNA is a rapidly diverging, largely repetitive DNA component of many eukaryotic genomes. Here we analyse the evolutionary dynamics of a satellite DNA repeat in the genomes of a group of Asian subtropical lady slipper orchids (*Paphiopedilum* subgenus *Parvisepalum* and representative species in the other subgenera/sections across the genus). A new satellite repeat in *Paphiopedilum* subgenus *Parvisepalum*, SatA, was identified and characterized using the RepeatExplorer pipeline in HiSeq Illumina reads from *P. armeniacum* (2n = 26). Reconstructed monomers were used to design a satellite-specific fluorescent in situ hybridization (FISH) probe. The data were also analysed within a phylogenetic framework built using the internal transcribed spacer (ITS) sequences of 45S nuclear ribosomal DNA.

Results: SatA comprises c. 14.5% of the *P. armeniacum* genome and is specific to subgenus *Parvisepalum*. It is composed of four primary monomers that range from 230 to 359 bp and contains multiple inverted repeat regions with hairpin-loop motifs. A new karyotype of *P. vietnamense* (2n = 28) is presented and shows that the chromosome number in subgenus *Parvisepalum* is not conserved at 2n = 26, as previously reported. The physical locations of SatA sequences were visualised on the chromosomes of all seven *Paphiopedilum* species of subgenus *Parvisepalum* (2n = 26–28), together with the 5S and 45S rDNA loci using FISH. The SatA repeats were predominantly localisedin the centromeric, peri-centromeric and sub-telocentric chromosome regions, but the exact distribution pattern was species-specific.

Conclusions: We conclude that the newly discovered, highly abundant and rapidly evolving satellite sequence SatA is specific to *Paphiopedilum* subgenus *Parvisepalum*. SatA and rDNA chromosomal distributions are characteristic of species, and comparisons between species reveal that the distribution patterns generate a strong phylogenetic signal. We also conclude that the ancestral chromosome number of subgenus *Parvisepalum* and indeed of all *Paphiopedilum* could be either 2n = 26 or 28, if *P. vietnamense* is sister to all species in the subgenus as suggested by the ITS data.

Keywords: *Paphiopedilum*, Karyotype, Satellite DNA, Fluorescent in situ hybridization, FISH

* Correspondence: leeyungi@hotmail.com; leeyungi@mail.nmns.edu.tw
[1]Biology Department, National Museum of Natural Science, No 1, Kuan-Chien Rd, 40453 Taichung, Taiwan, Republic of China
[2]Department of Life Sciences, National Chung Hsing University, 40227 Taichung, Taiwan, Republic of China
Full list of author information is available at the end of the article

Background

Nuclear genomes of higher plants are composed of coding and regulatory sequences and substantial amounts of repetitive DNA [including e.g. (retro) transposable elements and tandemly repeated DNA] that are considered to play important roles in genome differentiation, dynamics and evolution [1–4]. Repetitive DNA contributes to the diversity of genome sizes encountered in plants [5–7]. Satellite DNA (Sat) constitutes highly amplified tandemly repeated sequences, that can vary in abundance, sequence and chromosomal distribution between species [7–16]. Often it occurs in heterochromatic, peri-centromeric or sub-telomeric regions of the chromosome, but it can also be in found in interstitial regions. Because of the dynamic nature of the organization and distribution of these types of repeats, the characterization of satellite DNA is particularly useful for chromosome identification and for reconstruction of patterns of species divergence [13, 17–22].

Paphiopedilum (Orchidaceae: Cypripedioideae) is the most diverse genus of terrestrial slipper orchids, containing about 80 species, nearly all of which are rare and threatened [23]. The genus consists of seven subgroups, comprising three subgenera (*Parvisepalum*, *Brachypetalum* and *Paphiopedilum*) and five sections in subgenus *Paphiopedilum* (*Paphiopedilum*, *Cochlopetalum*, *Coryopedilum*, *Pardalopetalum* and *Barbata*) based on morphological, cytological and molecular phylogenetic data [24–28]. Patterns of speciation in some sections are complex and potentially involve recurrent patterns of interspecific hybridization, arising from the redistribution of taxa with changing sea levels across South East Asia during the glacial cycles of the late Cenozoic [28, 29].

The genus is characterized by considerable chromosome number variation (2n = 26–42) and a relatively wide range of genome sizes (2.2-fold, 1C = 16.5–35.9 pg, mean 1C = 25.4 pg) [27]. Previous cytological studies have suggested that Robertsonian translocations have contributed to the diversity of chromosome numbers observed between *Paphiopedilum* species, involving the fission of metacentric chromosomes at or near the centromere to generate telocentric chromosomes [30–37]. In addition, it is clear from more recent cytological studies that other types of complex chromosomal rearrangements (e.g. inversions and duplications) may also have contributed to the karyotypic diversity observed [38].

Little is known about the composition, diversity and evolutionary dynamics of repetitive DNA sequences in the genomes of *Paphiopedilum* species. In a study of ribosomal DNA (rDNA) sequence evolution by Lan and Albert [38], no clear relationships were uncovered between the number of rDNA signals, chromosome number and genome size. Duplications of the nuclear 45S rDNA locus occurred independently in subgenus *Parvisepalum* and sections *Coryopedilum* and *Pardalopetalum* of subgenus *Paphiopedilum*, whereas duplications of 5S rDNA loci were only observed in subgenus *Paphiopedilum*.

Recently, with developments in high-throughput sequencing approaches, non-model species have become more amenable to in-depth analyses of the repetitive DNA component of their genomes [7, 39, 40]. Even for species with large genomes (1C > 10 Gbp), it is possible to gain insights into the types, amounts, diversity and evolution of the most abundant repetitive elements using low-coverage sequencing data [14, 15, 41–47].

In this study, we use low-coverage genomic DNA sequence data from the Illumina next-generation sequencing platform to characterize the satellite DNA component of seven *Paphiopedilum* species, selected to represent the phylogenetic diversity of the genus. We undertook an in-depth analysis of the most abundant satellite DNA identified in the genus, which was identified in *P. armeniacum*. This species belongs to subgenus *Parvisepalum*, a lineage that is considered to have diverged from the rest of *Paphiopedilum* early in the evolution of the genus. In addition, we examined the chromosomal distribution of SatA in closely related species of subgenus *Parvisepalum* and representative species belonging to the other two subgenera, to provide a phylogenetic perspective of its distribution and evolution across the genus. Finally, we explored the utility of SatA as a chromosomal marker for characterizing karyotype evolution in species belonging to subgenus *Parvisepalum*.

Results

Phylogenetic relationships in *Paphiopedilum* subgenus *Parvisepalum*

Nuclear ribosomal ITS sequences were used to reconstruct phylogenetic relationships of the seven *Paphiopedilum* species belonging to subgenus *Parvisepalum* (Additional file 1: Fig. S1). The analysis re-confirmed the monophyly of the subgenus and resolved *P. vietnamense* as sister to the other species (bootstrap = 93%; PP = 1). In addition, two clades within the subgenus were recovered with strong to moderate support, the first consisting of *P. hangianum* and *P. emersonii* (bootstrap = 93%; PP = 1) and the other comprising *P. armeniacum*, and *P. malipoense*, (bootstrap = 69%; PP = 0.97). However, relationships between these clades and the two remaining species of the subgenus, i.e. *P. delenatii* and *P. micranthum,* remained unresolved.

Satellite DNA identification and characterization

Using RepeatExplorer to individually cluster the Illumina HiSeq data for seven *Paphiopedilum* taxa (corresponding to between c. 0.84 and 7.6% of the genome depending on the taxon, Additional file 2: Table S1), we identified four distinct types of satellite DNA (SatA, SatB, SatG, and SatJ) based on the shape of the output graphs. The abundance of each repetitive DNA type varied between species (Table 1), with none containing all four satellites. The amount of satellite DNA was also estimated for the sister genus *Phragmipedium*, in which only SatG was identified, occurring in low abundance (0.04%) in the *P. longifolium* genome.

SatA appeared to be specific to subgenus *Parvisepalum*, and it accounted for 14.4% of the *P. armeniacum* genome (Table 1). In contrast, SatB was found in all *Paphiopedilum* subgroups, ranging from 0.1% in *P. armeniacum* and *P. concolor* (subgenus *Parvisepalum* and section *Brachypetalum*, respectively) to 7.5% in *P. lowii* (section *Pardalopetalum*), although it was absent in the outgroup *P. longifolium*. SatG was found in both *Phragmipedium* and most *Paphiopedilum* species analysed with the exception of *P. armeniacum* (subgenus *Parvisepalum*) and *P. concolor* (subgenus *Brachypetalum*), in which it appears to have been lost or to be present in amounts below the threshold of detection used here. SatJ was found exclusively in just two sections of subgenus *Paphiopedilum*, comprising 2.5 and 1.8% of the *P. villosum* (section *Paphiopedilum*) and *P. appletonianum* (section *Barbata*) genomes, respectively (Table 1).

Characterization of SatA in subgenus *Parvisepalum*

The characteristics of the top four most abundant SatA monomers in *P. armeniacum* are summarized in Additional file 3: Table S2 and Additional files 4, 5 and 6: Figs. S2-S4. All have high AT content (c. 66%) and contain multiple, often long stretches, of inverted repeat regions which form multiple hairpin loop motifs interspersed with unpaired bases (Additional file 4: Fig. S2). The largest (359 bp) monomer (CL1_965) is made up of three highly similar 146, 146 and 67 bp long repeat subunits (Additional file 6: Fig. S4) and is the most abundant high BLAST similarity hit in an all-to-all sequence comparison to nearly 50% of all SatA reads. The remaining three monomers (CL1_940, CL1_393 and CL1_886) are distinct from CL1_965. They ranged in length from 235 to 307 bp and possessed high sequence similarity (above 90%) to each other and as such can be aligned easily.

Chromosomal organization of SatA and rDNA

SatA distribution patterns

The physical locations of long (> 1000 bp = the lower threshold of FISH sensitivity) [48] stretches of SatA sequence were visualized using FISH and showed that SatA hybridized to all species of subgenus *Parvisepalum* (Fig. 1). In contrast, none of the representative species belonging to the other subgenera or sections of *Paphiopedilum* had any hybridization signal. Both 45S and 5S rDNA probes hybridized to the same chromosome preparations (Fig. 2).

Species of *Paphiopedilum* subgenus *Parvisepalum* have 2n = 2× = 26 (except for *P. vietnamense*, 2n = 2× = 28), with mostly metacentric or sub-metacentric chromosomes. However, despite this apparent karyotypic uniformity, the hybridization pattern of the SatA probe was seen to differ considerably between species (Fig. 3). Thus to compare the physical distribution of the SatA hybridization sites in a phylogenetic context, karyotypes (Fig. 3) and ideograms (Fig. 4) of all seven species were prepared and arranged according to the nrITS phylogenetic tree. The arrangement of the chromosomes shown in these figures assumes that the SatA and rDNA signals are most likely carried on homologous chromosomes in the most closely related species, using approaches developed in Lim et al. [49] (Figs. 3 and 4).

The greatest abundance of SatA signals (which broadly reflects the genome proportion of 14.4% estimated by RepeatExplorer; see Table 1) was found in *P. armeniacum*, where all chromosomes had at least one site of SatA sequences. The signals were mainly located in the centromeric (defined here to include both the centromere and peri-centromeres) and sub-telomeric regions, although the precise position, strength and frequency of signals varied between chromosomes. Broadly similar patterns were also observed in *P. malipoense*. In contrast, clear SatA signals in *P. hangianum* and *P. emersonii* were only found in the centromeric regions of about half the chromosomes. For the remaining species analysed, both centromeric and sub-telomeric SatA signals were detected on at least some of the chromosomes, although the exact position, and strength of signal differed between species. Overall, the most abundant and intense centromeric and sub-telomeric SatA signals were observed in *P. armeniacum* and *P. malipoense*, whereas the weakest signals were found in *P. vietnamense* (Fig. 1).

By arranging karyotypes in a phylogenetic framework as predicted from nrITS sequence data (Fig. 3), the distribution of the SatA sequences is more similar in closely related species than more distantly related species, as illustrated by the following examples. (1) The distribution of the SatA sites in *P. armeniacum* were broadly similar to those in *P. malipoense* to which it is closely related. Nevertheless, although SatA sites were detected in centromeric and sub-telocentric regions of all eight chromosome pairs in *P. malipoense*, six chromosome pairs lacked prominent SatA signals in the centromeric region. These strong SatA signals coincided with heterochromatic bands seen by DAPI staining (data not shown) which became apparent after the denaturation step of the FISH protocol. (2) *P. emersonii* and *P. hangianum* are closely related sister species and both possessed SatA sites primarily in the peri-centromeric region.

Table 1 Major types of repetitive DNA in *Paphiopedilum*

Subgenus/Section	Outgroup	Parvisepalum	Brachypetalum	Cochlopetalum	Coryopedilum	Pardalopetalum	Paphiopedilum	Barbata
	Phragmipedium longifolium	*P. armeniacum*	*P. concolor*	*P. primulinum*	*P. rothschildianum*	*P. lowii*	*P. villosum*	*P. appletonianum*
	%	%	%	%	%	%	%	%
Satellites								
SatA	0.00	14.39	0.00	0.00	0.00	0.00	0.00	0.00
SatB	0.00	0.10	0.13	6.18	2.81	7.45	1.07	4.11
SatG	0.04	0.00	0.00	0.46	0.65	1.63	0.69	0.42
SatJ	0.00	0.00	0.00	0.00	0.00	0.00	2.51	1.80
Total Satellites	0.04	14.49	0.13	6.64	3.47	9.08	4.27	6.33
LTR elements								
Ty3/gypsy								
Ogre/Tat	30.04	31.80	48.86	38.03	44.98	39.06	46.32	40.48
Chromovirus	7.82	2.49	1.47	0.59	1.44	0.81	0.89	0.66
Athila	0.20	4.41	1.21	0.21	1.19	0.18	0.28	0.14
Total Ty3/gypsy	38.06	38.70	51.54	38.83	47.61	40.05	47.49	41.28
Ty1/copia								
Maximus/SIRE	3.05	2.24	1.56	0.97	1.49	0.48	0.99	0.62
Ivana	0.00	3.14	2.14	1.14	1.92	0.81	2.74	1.82
Tork	7.17	0.00	0.16	0.05	0.04	0.08	0.05	0.07
Total Ty1/copia	10.22	5.38	3.86	2.16	3.45	1.37	3.78	2.51
Other repetitive elements								
TRIM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LINE	1.09	0.57	0.49	0.10	0.76	0.53	0.36	0.26
DNA transposon	0.60	2.41	1.97	1.31	2.02	1.25	1.63	1.83
MITE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
rDNA	0.14	0.04	0.04	0.06	0.00	0.00	0.02	0.00
SSR	0.23	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pararetrovirus	0.00	0.08	0.00	0.00	0.16	0.00	0.00	0.59
Unclassified repetitive	17.31	9.86	10.60	11.94	10.36	9.90	10.63	11.38
Low and single copy	32.26	28.46	31.37	38.94	32.14	37.80	31.82	35.83
Total % repetitive DNA	61.18	62.04	59.54	50.43	59.26	53.62	58.64	53.88

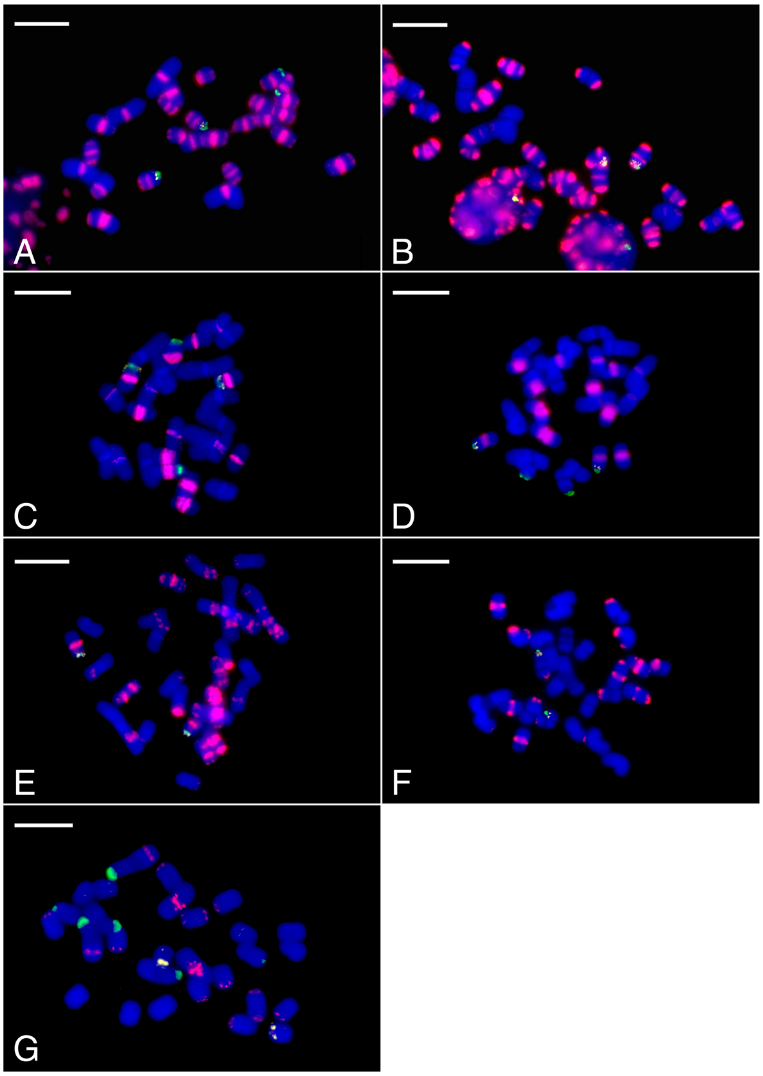

Fig. 1 Fluorescent in situ hybridization (FISH) of root tip metaphase chromosomes of species belonging to subgenus *Parvisepalum* with the SatA (red), 45S (green) and 5S (white) rDNA probes, counterstained with DAPI (blue): (**a**) *P. armeniacum*, (**b**) *P. malipoense*, (**c**) *P. emersonii*, (**d**) *P. hangianum*, (**e**) *P. micranthum*, (**f**) *P. delenatii*, (**g**) *P. vietnamense*. Bar = 10 μm

Nevertheless, in *P. hangianum*, the SatA signals co-localized exclusively with the heterochromatic bands in the centromeric region of six chromosome pairs, whereas in *P. emersonii*, SatA signals were restricted to the centromeric regions of just eight chromosome pairs, despite heterochromatic bands being present in the centromeric region of all chromosomes (data not shown).

In contrast to these species, *P. delenatii* showed only weak SatA signals in the peri-centromeric, sub-telocentric and interstitial regions. The distribution of SatA and rDNA signals in *P. vietnamense* is highly distinctive in comparison with the other species. In *P. micranthum,* multiple strongly dispersed interstitial signals across several chromosomes were observed. In addition, several hemizygous sub-telomeric *Sat*A signals, including one which spanned almost the entire length of one chromosome, were found in *P. micranthum* (Figs. 1e, 3 and 4).

rDNA distribution patterns

The 45S and 5S rDNA sequences in *Paphiopedilum* subgenus *Parvisepalum* were consistently localized in sub-telomeric positions although the number and strength of signals varied between species (Figs. 1 and 3, Table 2). All species had at least one pair of chromosomes with co-localized 45S + 5S rDNA sites (see ideograms in Fig. 4). In *P. micranthum* and *P. delenatii* (Figs. 1e and f), no further 45S or 5S rDNA sites were detected, whereas *P. armeniacum*, *P. emersonii* and *P. hangianum* (Figs. 1a, c and d) had an additional chromosome pair with 45S rDNA signals in sub-telomeric positions. The most

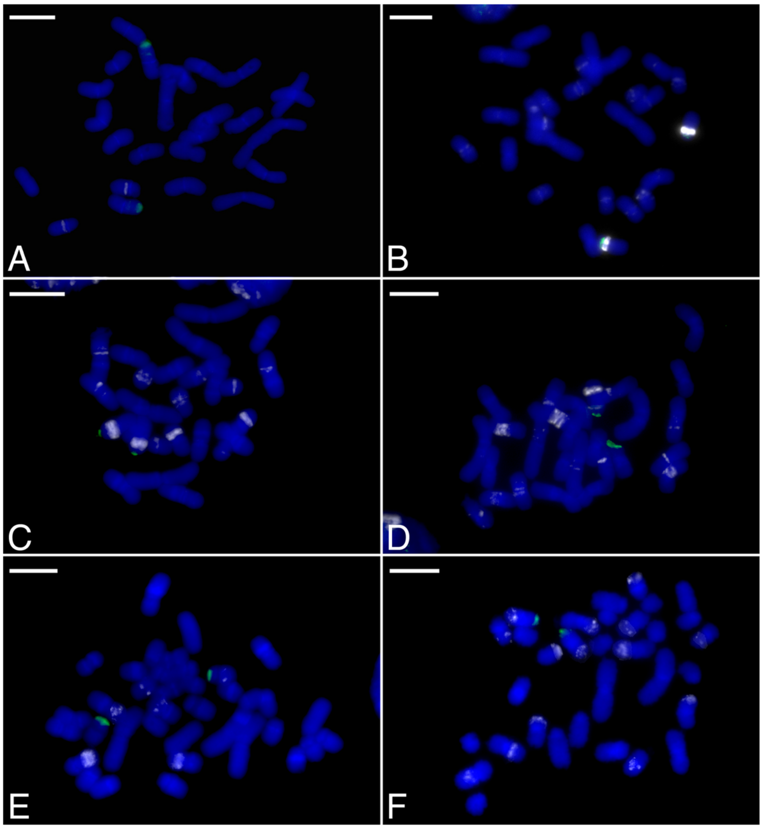

Fig. 2 Fluorescent in situ hybridization (FISH) of root tip metaphase chromosomes with the SatA, 45S (green) and 5S (white) rDNA probes, counterstained with DAPI (blue): (**a**) *P. concolor* (subgenus *Brachypetalum*), (**b**) *P. villosum* (section *Paphiopedilum*), (**c**) *P. rothschildianum* (section *Coryopedilum*), (**d**) *P. lowii* (section *Pardalopetalum*), (**e**) *P. appletonianum* (section *Barbata*), and (**f**) *P. primulinum* (section *Cochlopetalum*). The absence of FISH signals confirms that SatA is indeed specific to subgenus *Parvisepalum*. Bar = 10 μm

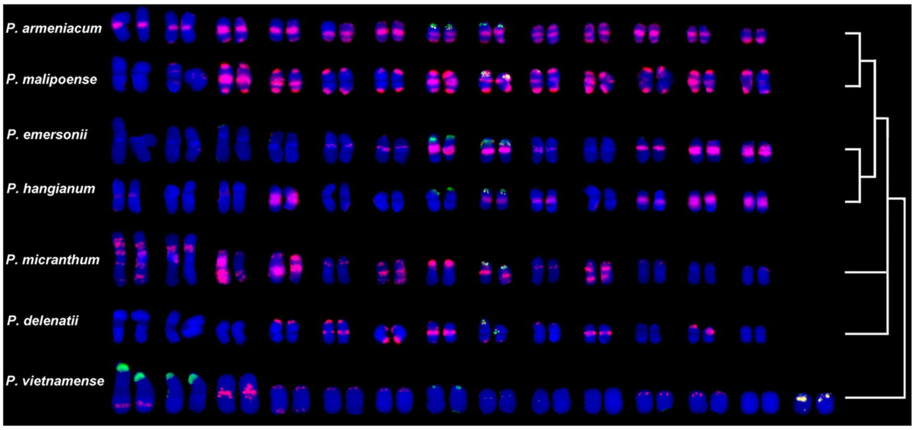

Fig. 3 Fluorescent in situ hybridization of karyotypes of *Paphiopedilum* subgenus *Parvisepalum* with the SatA (red), 45S (green) and 5S (white) rDNA probes, counterstained with DAPI (blue). Phylogenetic relationships between these species shown on the right-hand side of the figure (see also Additional file 1: Fig. S1)

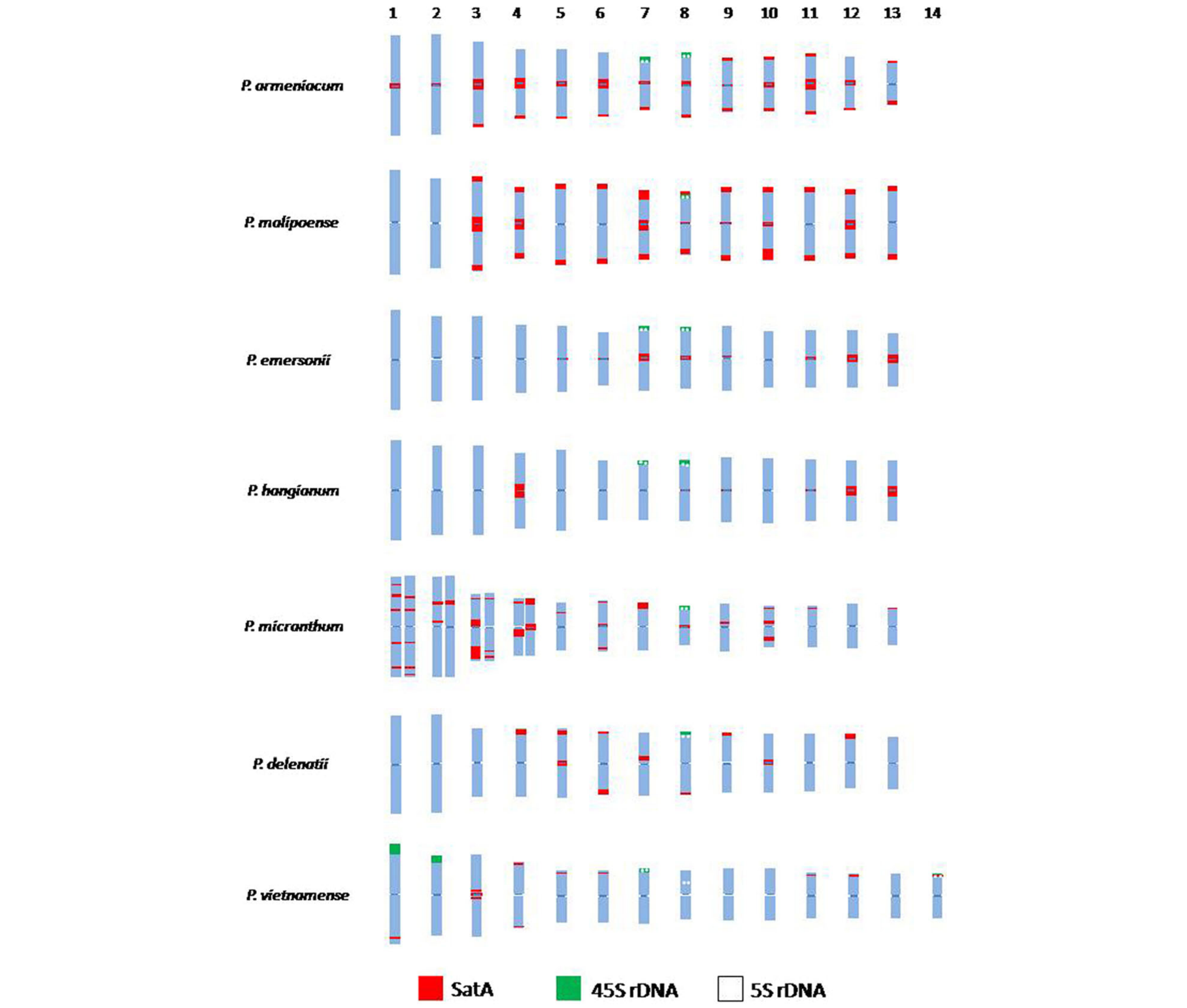

Fig. 4 Ideograms of somatic metaphase chromosomes of species belonging to subgenus *Parvisepalum*.SatA (red), 45S (green) and 5S (white) rDNA signals

Table 2 SatA and rDNA signals in seven species of *Paphiopedilum* subgenus *Parvisepalum*

Taxon	2n	Number of SatA sites				Number of rDNA sites		
		metacentric	sub-telocentric	interstitial	dispersed	45S	5S	45S + 5S co-localization
P. armeniacum	26	24	30			4	2	2
P. delenatii	26	6	14			2	2	2
P. malipoense	26	7	44		present	2	2	2
P. micranthum	26	6	16	14	present	2	2	2
P. emersonii	26	16				4	2	2
P. hangianum	26	14				4	2	2
P. vietnamense	28	2	14	2		8	4	2

distinctive pattern was observed in *P. vietnamense* (Fig. 1g); in addition to the pair of chromosomes with co-localised 45S + 5S rDNA sites, three chromosome pairs had 45S rDNA sites in sub-telomeric regions. Two of these pairs of 45S rDNA signals were of much higher intensity and located on large chromosomes in contrast to the 45S rDNA signals of other species which were all located on smaller chromosomes and were much less intense.

Discussion

Global satellite repeat composition in *Paphiopedilum*

The analysis of satellite DNAs across an ever-increasing diversity of plants has shown that there is considerable diversity in sequence composition, diversity, organization, abundance and evolution [4, 6, 15]. Here we have taken advantage of high-throughput sequencing approaches and the bioinformatic pipeline RepeatExplorer to provide insights into the most abundant satellite DNA repeats (i.e. comprising ≥0.01% of the genome) across the seven subgenera/sections of the slipper orchid genus *Paphiopedilum*. As with some plant genomes analysed to date, e.g. *Silene latifolia* [50], the diversity of satellites identified was low, with only four major satellite types identified. Nevertheless, this contrasts with other plant genomes where a considerably greater diversity has been reported. For example, 51 different satellites were identified in *Vicia peregrina* [15] and up to 54 types, each comprising > 0.01% of the genome, were identified in *Luzula elegans* [45, 51].

Many plant satellites have been shown to be species-specific (e.g.in Fabeae, [15]), indicative of relatively rapid sequence divergence rates compared with speciation rate. However, this was not the case in *Paphiopedilum*, in which all four satellites were observed to be more widely distributed, either between more than one species in a subgenus (SatA) or between sections (SatJ), subgenera (SatB) or genera (SatG). The contribution of these four satellites to the whole genome varied considerably between species analysed, from just 0.13% of the genome in *P. concolor* (subgenus *Brachypetalum*) to 14.39% in *P. armeniacum* (subgenus *Parvisepalum*) (Table 1). In addition, there was no evidence of a clear correlation between satellite abundance and genome size. For example, both *P. concolor* (23.1 pg/1C) and *P. armeniacum* (22.8 pg/1C) have similar genome sizes, but the lowest and highest percentage of the genome comprising satellite DNAs of the species investigated (Table 1 and Additional file 2: Table S1). Such a situation has been observed in similar analyses in other groups [14, 15] in which it is generally transposable elements, particularly retrotransposons such as certain Ty1/Copia and Ty3/Gypsy elements, which correlate most closely with genome size [5, 15, 52, 53].

Analysis of SatA: The most abundant satellite in *Paphiopedilum*

Plant genomes are often dominated by just one or a few repeats which have amplified to high copy number. Typically the most abundant repeats are retrotransposons. For example, the Ty3/Gypsy Ogre elements have been shown to comprise 54% of the genome of *Vicia faba* [15], whereas Ty3/Gypsy Gorge3 elements contribute up to 34% of the *Gossypium exiguum* genome [52]. Nevertheless, there are also examples in which satellites have contributed to a substantial proportion of the genome. For example, the VicTR-B satellite repeat represents about 25% of the *Vicia sativa* genome [54], whereas the FriSAT1 repeat accounts for up to 11% of the *Fritillaria affinis* genome [14], and the PaB6 repeat represents about 10% of the *Prospero autumnale* cytotype B^6B^6 genome [55]. In *Paphiopedilum*, the most abundant satellite DNA is SatA (Table 1) which is estimated to comprise c.14.39% of the *P. armeniacum* genome and is specific to subgenus *Parvisepalum* (Figs.1 and 2).

Origin, evolution and organization of SatA in subgenus *Parvisepalum*

Characterization of SatA in *Paphiopedilum* subgenus *Parvisepalum* revealed just four (or conservatively two if CL_940, 393 and 886 are considered to be minor variants of a single sequence) primary monomers which have no homology to any satellite DNA sequences identified in other subgenera or sections of *Paphiopedilum*, the sister genus *Phragmipedium* or other plant genomes published to date. However, given that satellite DNA sequences are known to diverge rapidly and that subgenus *Parvisepalum* may have diverged from the rest of *Paphiopedilum* only c. 20 Mya [28], it is likely that SatA clusters arose independently in the lineage leading to subgenus *Parvisepalum*. The arrangement of the SatA and rDNA sites in *P. vietnamense* is highly distinctive in comparison with the other species. If *P. vietnamense* is sister to the rest of the subgenus, as indicated by ITS sequence divergence, then the distribution patterns in this species maybe either an apomorphic or plesiomorphic character state, or indeed a mixture of both character states, depending on the specific chromosomal signal.

Since the origin of SatA, there have clearly been considerable changes in its abundance and chromosomal distribution between the different species comprising subgenus *Parvisepalum*. However, despite this variability, most signals occur at peri-centromeric or sub-telocentric regions of the chromosome, as is typical of many satellites [22]. When the distribution patterns are considered in a phylogenetic context (Figs. 3 and 4), it is apparent that the most closely related species carry the most similar satellite DNA distributions and that the distribution of satellites carries a strong phylogenetic signal.

The plethora of hemizygous interstitial sites in *P. micranthum* indicates rapid divergence of SatA within a species, perhaps associated with genetic drift in local populations and some gene flow between populations. Further studies of *P. micranthum* at the population level are clearly needed to provide deeper insights into satellite DNA proliferation in this species.

Organization of rDNA sites in subgenus *Parvisepalum*

Previous studies of the organization of rDNA sequences in *Paphiopedilum* have suggested that the ancestral state for the genus is two 45S and two 5S rDNA sites [38]. Although the number of 5S rDNA sites was observed to be maintained throughout subgenus *Parvisepalum*, in our study we found that the number of 45S rDNA sites varied from one to eight (Table 2, Fig. 4), in contrast to the range from two to four previously reported [38]. Indeed, the *P. delenatii*, *P. micranthum* and *P. malipoense* 5S and 45S rDNA FISH profiles in this study (Figs. 1b, e and f, Table 2) differed from those of Lan and Albert [38], which suggests that there may well be intraspecific rDNA site number variation in these species.

Ancestral chromosome number of *Paphiopedilum*

The new karyotype of *P. vietnamense*, with 2n = 28, metacentric chromosomes reported here (Figs. 3 and 4), expands the chromosome number range for subgenus *Parvisepalum*, which was previously considered to be conserved at 2n = 26. Prior to this study, 2n = 28 was only reported for *P. hookerae* and *P. sangii* which both belong to section *Barbata* [36, 56]. Previously it was suggested that *Paphiopedilum* had an ancestral chromosome number of 2n = 26 comprising metacentric chromosomes [57]. However, confidence in this number is dependent on the precise placement of *P. vietnamense* in phylogenetic trees for the genus. If *P. vietnamense* is sister to the rest of subgenus *Parvisepalum,* as suggested from the nrITS sequence data (Additional file 1: Fig. S1), then the ancestral chromosome number of subgenus *Parvisepalum* could be 2n = 26 or 28. However, if *P. armeniacum* is sister to the rest of the subgenus, as suggested by plastid DNA sequences [28], then 2n = 26 is the most likely ancestral chromosome number of the subgenus. Given that 2n = 26 or 2n = 28 are both possible ancestral chromosome numbers for subgenus *Parvisepalum,* and because subgenus *Parvisepalum* is sister to the rest of *Paphiopedilum* [28], the ancestral chromosome number of the genus itself could also be 2n = 26 or 2n = 28. Chromosome numbers from other genera in Cypripedioideae do not support one the alternatives above the other, although perhaps they point most strongly towards an ancestral chromosome number of 2n = 26 for Cypripedioideae. This is because the sister genera *Phragmipedium* and *Mexipedium* are together sister to *Paphiopedilum* and they have chromosome numbers of 2n = 18–30 for *Phragmipedium* and 2n = 26 for *Mexipedium* (a monotypic genus) (Chromosome counts database, http://ccdb.tau.ac.il/search/Phragmipedium/). However, we do not currently know how the range of chromosome numbers of *Phragmipedium* species are distributed across the phylogenetic tree for the genus and without that we cannot readily determine the ancestral number of *Phragmipedium* and *Mexipedium*. Thus, in determining the chromosome ancestry of *Paphiopedilum*, we need greater clarity of the phylogenetic placement of *P. vietnamense* and of the evolution of chromosome numbers in *Phragmipedium*.

Conclusion

We identified and characterized a new satellite repeat, SatA from *P. armeniacum* using the RepeatExplorer pipeline to analyse HiSeq Illumina reads. SatA is specific to subgenus *Parvisepalum* but absent from the other two subgenera in *Paphiopedilum*. Since the distribution pattern of SatA on chromosomes in subgenus *Parvisepalum* is species-specific and hence rapidly evolving, and possesses a strong phylogenetic signal, it is an ideal probe that could be used as a chromosomal marker for characterizing karyotype evolution in species belonging to subgenus *Parvisepalum*. Nevertheless, greater certainty is needed on the phylogenetic placement of *P. vietnamense*, as this will help to shed light on the ancestral karyotype and chromosome number of the genus. Certainly, it is recognized that the currently available phylogenetic data for *Paphiopedilum* is limited and could be greatly enhanced by the application of phylogenomic and bioinformatic approaches, involving multiple genes, repeats or even whole genomes, to generate robust species trees and phylogenetic insights of the genus.

Methods

Plant materials

Details of the origin of plant materials used for Illumina sequencing, fluorescent in situ hybridization (FISH), and genome size estimations and the location of the voucher specimens are given in Additional file 7: Table S3. All other nrITS sequences used in the phylogenetic analysis were downloaded from GenBank.

Genome size estimation using flow cytometry

Flow cytometry (FCM) was used to determine the genome sizes of each studied species and estimate the volume of Illumina sequence DNA required to characterize the repetitive fraction of the genome (see below). Samples for FCM were prepared as in Ebihara et al. [58] with slight modifications. Briefly, c. 1 cm^2 of leaf material of the *Paphiopedilum* sample was co-chopped with the reference standard *Vicia faba* 'Inovec' (2C = 26.9 pg, [59]) in 2.0 mL of Ebihara buffer [1.0% Triton X-100, 140 mM 2-mercaptoethanol, 50 mM Na_2SO_3, 50 mM Tris-HCl (pH 7.5), 40 mg/mL polyvinyl-pyrolidone (PVP-40) and 0.1 mg/mL ribonuclease] on ice in a fume hood. The resulting slurry was incubated for 5 min on ice and filtered through a 30-μm nylon mesh (Partec) into a 2.0-mL tube and the nuclei were subsequently stained with 100 μL propidium iodide (1 mg/mL). The filtrate was incubated for 15 min at 37 °C and then left on ice for c. 30 min. Genome sizes were measured on a Partec-Cyflow SL3 flow cytometer (Partec GmbH, Münster, Germany) fitted with a 100 mW green solid state laser (532 nm, Cobolt Samba, Solna, Sweden). As many orchids, *Paphiopedilum* presents endoreplication [60] and we had to run the samples longer than usual to recover 1000 nuclei in the 2C peak of G1-phase for *Paphiopedilum*. Three measurements were made for each sample. The output histograms were analysed with the FlowMax software v.2.4 (Partec GmbH).

High-throughput sequencing of genomic DNA

Total genomic DNA was isolated from young leaf samples using DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Paired-end sequencing (2× 100 bp or 2 × 125 bp, 300–500 bp insert size) of total genomic DNA was performed using Illumina HiSeq 2000 (Illumina, Inc.) at Genomics BioSci & Tech. Co., Ltd., Taipei, Taiwan and using a HiSeq 2500 at BGI, Hong Kong and Service XS, Leiden, The Netherlands, based on libraries made by the service providers.

Processing of Illumina HiSeq data

The quality of the raw reads was assessed by Fast QC v0.10.1. Raw reads were trimmed and quality-filtered with a FASTX-Toolkit v 0.013 to give 90-bp length reads with 90% of bases having a minimum Phred score of 20. To remove sequences of organellar origin, custom Perl scripts and the stand-alone version of BLAST v2.2.16 [60] were used to screen the quality-filtered reads against a custom database containing the draft plastid genome of *Phragmipedium longifolium* (unpublished data provided by Dr. W. Mark Whitten, University of Florida, USA) and the published mitochondrial genomes of ten monocots (taken from GenBank). Parameter settings used for BLASTN [61] searches were: -v 1 -G 0 -E 2 -K 0 -b 0 - e 0.000001 -F mL. Sequences with similarity matches to the database (*E*-value $\leq 19 \times 10^{-6}$) were removed. All remaining reads were considered to be of nuclear origin and uploaded to RepeatExplorer (http://www.repeatexplorer.org) within the Galaxy server environment, as described in Dodsworth et al. [62]. Nuclear reads were paired using the FASTQ interlacer tool implemented in RepeatExplorer [63, 64]. Overlapping read pairs, which are caused by the presence of overly short genomic DNA fragments in the sequence library, can adversely affect the subsequent RepeatExplorer clustering analysis. Consequently, they were removed using a RepeatExplorer utility (minimum overlap = 30 nt, maximum mismatch per 100 bp = 1, offset = 5).

Clustering and annotation of satellite DNA with RepeatExplorer (RE)

The RepeatExplorer pipeline runs a graph-based clustering algorithm [63, 64] to assemble the groups of frequently overlapping reads into clusters of reads, each representing a repetitive element, or part of a repetitive element with a higher order genome structure. Similarity- and structure-based repeat identification tools in RepeatExplorer aid the identification of the repeats. RepeatExplorer uses a BLAST threshold of 90% similarity across 55% of the read to identify reads to each clusters (minimum overlap = 55, cluster threshold = 0.01%, minimum overlap for assembly = 40), and the clusters are identified based on a principle of maximum modularity.

To identify major shared and unique repetitive elements between *Paphiopedilum* species, comparative clustering analysis was performed on a combined dataset comprising Illumina sequence reads from seven *Paphiopedilum* species representing each of the seven major *Paphiopedilum* subgroups, and *Phragmipedium* as an outgroup (see Additional file 2: Table S1). For each taxon, the number of HiSeq Illumina reads analysed in RepeatExplorer was scaled according to the genome size of the species, so that the same proportion of the genome (2%) was analysed in each case. To provide greater insights into the characteristics of individual repetitive elements identified in repeat clusters, these eight HiSeq samples were also analysed in RepeatExplorer individually. For that we used the highest possible number of reads that can be readily handled by the software (Additional file 2: Table S1). Preliminary runs of RepeatExplorer indicated that the *P. armeniacum* sample contained a high (> 10%) proportion of satellite DNA. Large repeat clusters containing many sequence reads, and identified as comprising tandem repeats (e.g. satellites) can be problematic in RepeatExplorer analysis as they consume high amounts of computing resources and so reduce the number of sequences that can be handled by the pipeline. Consequently, prior to clustering, the reads were subjected to a custom sequence filter (containing a database of 617 of the most abundant sequence contigs from the satellite SatA, see below) to remove 90% of the satellite sequences, as recommended by the developers of RepeatExplorer. Clusters were identified manually as described in Novák et al. [64] by scanning read similarity hits to the RepeatMasker [65] database, visual examination of graphs and the location of cluster mates. Only clusters with genome proportions (GP) > 0.01% were included in subsequent analyses.

Characterization of the most abundant satellite repeat in *Paphiopedilum*

The structural features of the top four monomers of the most abundant satellite repeat identified by RepeatExplorer (SatA) were characterized using DOTTER [66] and the Vienna RNA fold tool implemented in Geneious v. 9.0.5 [67]. Sequence homology was checked against other *Paphiopedilum* satellites and by BLAST against published satellite DNA sequences in PlantSat (http://w3lamc.umbr.cas.cz/PlantSat/) [68].

PCR amplification, cloning and sequencing of SatA

The consensus sequence of the SatA monomer was used to design oligonucleotide primers (CL3C37-F:CATTTTCAACGTCGAGCC; CL3C37-R:AGACAAATTCTAAGCTATATGGAC; PCR product 147 bp) for amplifying the SatA monomers from genomic DNA of *P. armeniacum*. The amplification was performed in a reaction volume of 20 μl using the KAPA HiFi HotStart PCR kit (KAPA Biosystems) in a Takara thermal cycler (Takara Bio, Shiga, Japan), and the reaction mixture contained 1× KAPA HiFi buffer, 0.3 mM each dNTP, 0.3 μM each primer, and 1unit KAPA HiFi HotStart DNA polymerase. The amplification profile included an initial step at 94 °C for 3 min, followed by 35 cycles of 94 °C for 15 s, 50 °C for 15 s, and 72 °C for 5 s, and finally a 1 min final extension at 72 °C. DNA was cloned using the T&A cloning vector system (Yeastern Biotech Co., Ltd., Taiwan) and DH5α competent cells (Genomics BioSci & Tech. Co., Ltd., Taiwan) following the manufacturer's instructions.

Chromosome preparation

Chromosome preparations for FISH were made according to the methods of Chung et al. [69] with minor modifications. Briefly, young, healthy root tips were harvested and pretreated in 2 mM 8-hydroxyquinoline at 18 °C for 5 h to accumulate metaphase nuclei, rinsed with distilled water and then fixed in freshly prepared Farmer's fluid (3:1 ethanol:glacial acetic acid). Root tips were macerated with 6% cellulose (Onozuka R-10, Yakult Honsha, Japan) and 6% pectinase (Sigma Chemical Co., St. Louis, Mo.) in 75 mM KCl, pH = 4.0 at 37 °C for 90 min, and squashed on a microscope slide in the same fixative. Slides were air-dried and stored at – 80 °C until required.

Fluorescent in situ hybridization

The FISH procedure followed that described in Lee et al. [70]. The SatA probe was designed from a conserved region of the SatA monomer alignment and labeled with digoxigenin using oligonucleotide-5′-end-labeling (AAATCTGACCTAATTTGGACCCAATCTTTGAACCTTC TAATTGAAGGTCAATTGGTGT). Probes for 45S rDNA (pTA71 containing a repetitive unit of 45S rDNA from *Triticum aestivum*) [71] and 5S rDNA (pTA794 containing the 5S rDNA repeat unit from *T. aestivum*) [72] were also used. The rDNA sequences were labeled by nick translation with digoxigenin-11-dUTP orbiotin-16-dUTP (Roche Diagnostics GmbH, Penzberg, Germany). Digoxigenin-labeled probes were detected by anti-digo xigenin-rhodamine (Roche Diagnostics GmbH), whereas biotin-labeled probes were detected using fluorescein isothiocyanate (FITC)-conjugated avidin (Vector Laboratories, Burlingame, CA, USA). Chromosomes were counterstained with 4′, 6-diamidino-2-phenylindole (DAPI) in an anti-fade solution (Vector Laboratories, CA, USA). All images were captured digitally using a CCD camera attached to an epifluorescence microscope (Axioskop 2, Carl Zeiss AG, Germany). The CCD camera was controlled by Image-Pro Plus software (version 4.5.1, Media Cybernetics, Yorktown, VA, USA), and final image adjustments were made with Adobe Photoshop CS2 (version 9.0.2, Adobe Systems Inc., San Jose, CA, USA).

Analysis of phylogenetic relationships in *Paphiopedilum*

DNA was extracted from each sample using a DNeasy Plant Mini Kit (Qiagen, Hilden, Germany). The nrITS regions, which include ITS1, ITS2 and the 5.8S nuclear rRNA gene, were amplified by PCR with the primer combinations of Sun et al. [73] and White et al. [74] and Sanger sequenced as described in Chochai et al. [27]. PCR products of *P. emersonii* were difficult to sequence directly and thus were cloned using the pGEM-T Vector System II (Promega, Madison, WI, USA). This yielded five distinct nrITS sequences which were verified using a BLAST search against the NCBI sequence database (National Center for Biotechnology Information, GenBank). GenBank accession numbers (http://www.ncbi.nlm.nih.gov) of nrITS sequences were listed in Additional file 8: Table S4.

Phylogenetic relationships between the seven *Paphiopedilum* species comprising subgenus *Parvisepalum* were analysed using the newly generated nrITS sequences together with those downloaded from GenBank for representative species of the other subgenera/sections of *Paphiopedilum* and with *Phragmipedium besseae* and *P. longifolium* as outgroup species. Sequences were aligned using CLUSTALW [75] implemented in Geneious v.9.0.5 and checked by eye. Phylogenetic relationships were analysed using a model-based Bayesian approach with MrBayes 3.2.1 [76]. The 'best-fit' model of evolution was selected under the Akaike information criterion test [77] as implemented in MrModel test 2.2 [78]. The general time reversal plus invariant rates and a gamma distribution (GTR + I + Γ) model was selected for the analyses. Two separate runs of four Monte Carlo Markov chains (MCMC; Yang and Rannala [79]) were performed for 10,000,000 generations until the mean deviation of split frequency dropped below 0.01, and a tree was sampled every 1000th generation. Trees from the first 25% of generations were discarded using the "burn-in" command, and the remaining trees were used to calculate an all-compatible consensus topology and posterior probability (PP) values for individual branches. The alignment datasets were further analysed using maximum parsimony (MP) in PAUP* version 4.0b10 [80]. Support for groups was evaluated using the bootstrap method [81] with 1000 replicates. The trees obtained in these analyses were drawn with the TreeGraph 2 software [82].

Additional files

Additional file 1: Figure S1. A MrBayes ITS subtree showing relationships in *Paphiopedilum* subgenus *Parvisepalum* is presented. Numbers above branches indicate bootstrap and posterior probability support values. Length of branches indicate number of changes. Numbers on tips indicate ITS clones. (JPG 186 kb)

Additional file 2: Table S1. Reads clustered and genome proportion for RepeatExplorer (RE) analysis. (XLSX 12 kb)

Additional file 3: Table S2. Characteristics of the top four-most SatA abundant monomers. (DOCX 12 kb)

Additional file 4: Figure S2. Hypothetical folding of the four most abundant SatA monomers: (A) CL1_965, (B) CL1_940, (C) CL1_393 and (D) CL1_886, when viewed as continuous molecules following the DNA energy model (Mathews 2004) implemented in Geneious v9.0.5. The repeat/inverted repeats in the monomers pair and fold to form hairpin-loop structures. (JPG 209 kb)

Additional file 5: Figure S3. Dot plots for the four most abundant SatA monomers: (A) CL1_965, (B) CL1_940, (C) CL1_393 and (D) CL1_886, by DOTTER2 (Sonnhammer and Durbin 1995) implemented in Geneious v 9.0.5. (JPG 539 kb)

Additional file 6: Figure S4. Sequence of the most abundant SatA monomers: (1) CL1_965, (2) CL_940, (3) CL1_393 and (4) CL1_886. Annotations show positions of the major subunits and major (> 10 bp long) repeat/inverted regions. (JPG 704 kb)

Additional file 7: Table S3. The vouchers and sources used for Illumina HiSeq, RepeatExplorer (RE) clustering, FISH and genome size estimation in this study. (DOCX 18 kb)

Additional file 8: Table S4. GenBank accession number of nrITS sequences used in phylogenetic analysis. (XLSX 14 kb)

Abbreviations

CCD: Charge-coupled device; DAPI: 4′, 6-diamidino-2-phenylindole; FCM: Flow cytometry; FISH: Fluorescent in situ hybridization; FITC: Fluorescein isothiocyanate; GP: Genome proportions; MCMC: Monte Carlo Markov chains; nrITS: Nuclear ribosomal internal transcribed spacer region; PCR: Polymerase chain reaction; PP: Posterior probability; PVP: Polyvinyl-pyrolidone; rDNA: Ribosomal DNA; RE: RepeatExplorer; SAT: Satellite DNA

Acknowledgements

JWY was supported by a UK Commonwealth Scholarship and an Emily Holmes Memorial Scholarship. The Hortus botanicus of Leiden University is thanked for contributing plant material.

Funding

This work was supported by the Ministry of Science and Technology, Taiwan, ROC [NSC 102–2313-B-178-001-MY3] and National Museum of Natural Science, Taiwan, ROC to YIL in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

YIL, JWY, IJL, MFF, ARL conceived the study; YIL, JWY, IJL, ARL designed the study; YIL, JWY, SD performed RepeatExplorer analyses; JWY performed molecular analyses; YIL, YCL performed cytological analyses; OH performed flow cytometry analyses; SI, MJMS, BG provided some sequence data; YIL, JWY, IJL, MFF,ARL wrote the paper; All authors read and approved the final manuscript.

Authors' information

[1]Biology Department, National Museum of Natural Science, No 1, Kuan-Chien Rd., 40,453 Taichung, Taiwan, ROC, [2]Department of Life Sciences, National Chung Hsing University, Taichung 40,227, Taiwan, ROC, [3]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK, [4]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK, [5]Forest Research Institute Malaysia (FRIM), 52,109 Kepong, Selangor Darul Ehsan, Malaysia, [6]Plant Breeding, Wageningen University & Research, P.O. Box 386, NL-6700 AJ Wageningen, The Netherlands, [7]Department of Crop Science, Faculty of Agriculture, University Putra Malaysia (UPM) Serdang, Selangor, Malaysia, [8]Schoolof Plant Biology, University of Western Australia, Crawley, WA 6009, Australia, [9]Endless Forms group, Naturalis Biodiversity Center, Vondellaan 55, 2332 AALeiden, The Netherlands, [10]Faculty of Science and Technology, University of Applied Sciences Leiden, Zernikedreef 11, 2333 CK Leiden, The Netherlands, [11]Institute Biology Leiden, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands.

Ethics approval and consent to participate

All the *Paphiopedilum* species seedlings used in this study are artificially propagated that approved by the Council of Agriculture of Executive Yuan in Taiwan with the regulations of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Biology Department, National Museum of Natural Science, No 1, Kuan-Chien Rd, 40453 Taichung, Taiwan, Republic of China. [2]Department of Life Sciences, National Chung Hsing University, 40227 Taichung, Taiwan, Republic of China. [3]School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK. [4]Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK. [5]Forest Research Institute Malaysia (FRIM), 52109 Kepong, Selangor Darul Ehsan, Malaysia. [6]Plant Breeding, Wageningen University & Research, P.O. Box 386, NL-6700, AJ, Wageningen, The Netherlands. [7]Department of Crop Science, Faculty of Agriculture, University Putra Malaysia (UPM) Serdang, Serdang, Selangor, Malaysia. [8]Department of Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, UK. [9]School of Plant Biology, University of Western Australia, Crawley, WA 6009, Australia. [10]Endless Forms Group, Naturalis Biodiversity Center, Vondellaan 55, 2332, AA, Leiden, The Netherlands. [11]Faculty of Science and Technology, University of Applied Sciences Leiden, Zernikedreef 11, 2333, CK, Leiden, The Netherlands. [12]Institute Biology Leiden, Leiden University, Sylviusweg 72, 2333, BE, Leiden, The Netherlands.

Received: 24 November 2017 Accepted: 23 July 2018
Published online: 02 August 2018

References

1. Heslop-Harrison JS, Schwarzacher T. Organisation of the plant genome in chromosomes. Plant J. 2011;66:18–33.
2. El Baidouri M, Panaud O. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. Genome BiolEvol.2013;5:954–65.

3. Dodsworth S, Leitch AR, Leitch IJ. Genome size diversity in angiosperms and its influence on gene space. Curr Opin Genet Dev. 2015;35:73–8.
4. Garrido-Ramos MA. Satellite DNA in plants: more than just rubbish. Cytogenet Genome Res2015;146:153–170.
5. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 2014;65:505–30.
6. Mehrotra S, Goyal V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genomics Proteomics Bioinformatics. 2014;12:164–71.
7. Weiss-Schneeweiss H, Leitch AR, McCann J, Jang J, Macas J. Exploring the role of next generation sequencing to understand the repeat landscape of the plant genome. In: Hörandl E., Appelhans MS.(eds.) Next-generation sequencing in plant systematics. Regnum Vegetabile, vol. 158: International Association for Plant Taxonomy (IAPT); 2015. p. 294.
8. Kubis S, Schmidt T. Heslop-Harrison JS (pat). Repetitive DNA elements as a major component of plant genomes. Ann Bot. 1998;82A:45–55.
9. Pich U, Schubert I. Terminal heterochromatin and alternative telomeric sequences in *Allium cepa*. Chromosom Res. 1998;6:315–21.
10. Schmidt T, Heslop-Harrison JS. Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci. 1998;3:195–9.
11. Ohmido N, Kijima K, Akiyama Y, de Jong HJ, Fukui K. Quantification of total genomic DNA and selected repetitive sequences reveals concurrent changes in different DNA families in indica and japonica rice. Mol Gen Genet. 2000;63:388–94.
12. Weiss-Schneeweiss H, Riha K, Jang CG, Puizina J, Scherthan H, Schweizer D. Chromosome termini of the monocot plant *Othocallis siberica* are maintained by telomerase, which specifically synthesises vertebrate-type telomere sequences. Plant J. 2004;37:484–93.
13. Lim YK, Kovarik A, Matyasek R, et al. Comparative genomics and repetitive sequence divergence in the species of diploid *Nicotiana* section *Alatae*. Plant J. 2006;48:907–19.
14. Kelly LJ, Renny-Byfield S, Pellicer J, et al. Analysis of the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 2015;208:596–607.
15. Macas J, Novák P, Pellicer J, et al. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS ONE.2015;10:e0143424.
16. Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst Evol. 2016; https://doi.org/10.1007/s00606-016-1356-9.
17. Lim YK, Matyášek R, Lichtenstein PC, Leitch RA. Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section *Tomentosae*. Chromosoma. 2000;109:245–58.
18. Taketa S, Ando H, Takeda K, Harrison EG, Heslop-Harrison JS. The distribution, organization and evolution of two abundant and widespread repetitive DNA sequences in the genus *Hordeum*. Theor Appl Genet. 2000;100:169–76.
19. Skalická K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovarik A. Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. New Phytol. 2005;166:291–303.
20. Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR. Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). New Phytol. 2005;168:241–52.
21. Begum R, Zakrzewski F, Menzel G, Weber B, Alam SS, Schmidt T. Comparative molecular cytogenetic analyses of a major tandemly repeated DNA family and retrotransposon sequences in cultivated jute *Corchorus* species (Malvaceae). Ann Bot. 2013;112:123–34.
22. Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. New Phytol. 2010;186:148–60.
23. Fay M, Rankou H. Slipper orchids on the IUCN Red List. In: 2015 Annual report to the environment agency - Abu Dhabi. Framework Support for Implementing the Strategic Plan of the IUCN Species Survival Commission. 2016:106–11. http://cmsdata.iucn.org/downloads/2015_ead_report_final_web.pdf
24. Cox AV, Eon AMPR, Albert VA, Chase MW. 1997. Phylogenetics of the slipper orchids (Cypripedioideae, Orchidaceae): nuclear rDNA ITS sequences. Plant Syst Evol. 1997;208:197–223.
25. Cribb P. The genus*Paphiopedilum*. Malaysia: Natural History Publications (Borneo); 1998.
26. Averyanov LV, Cribb P, Ke Loc P, Tien HN. Slipper orchids of Vietnam. Kew: Royal Botanic Gardens. 2003;
27. Chochai A, Leitch IJ, Ingrouille MJ, Fay MF. Molecular phylogenetics of *Paphiopedilum* (Cypripedioideae; Orchidaceae) based on nuclear ribosomal ITS and plastid sequences. Bot J Linn Soc. 2012;170:176–96.
28. Guo YY, Luo YB, Liu ZJ, Wang XQ. Reticulate evolution and sea-level fluctuations together drove species diversification of slipper orchids (*Paphiopedilum*) in Southeast Asia. Mol Ecol. 2015;24:2838–55.
29. Yap JW. Molecular and genome evolution in the Malesian slipper orchids (*Paphiopedilum* section *Barbata*). PhD Thesis. Queen Mary University of London, London, UK; 2016.
30. Karasawa K. Karyomorphological studies in *Paphiopedilum*. Orchidaceae Bull Hiroshima Bot Gard. 1979;2:1–149.
31. Karasawa K. Karyomorphological studies in *Phragmipedium*. Orchidaceae. Bull Hiroshima Bot Gard. 1980;3:1–49.
32. Karasawa K. Karyomorphological studies on four species of *Paphiopedilum*. Bull Hiroshima Bot Gard. 1982;5:70–9.
33. Karasawa K. Karyomorphological studies on nine taxa of *Paphiopedilum*. Orchidaceae. Bull Hiroshima Bot Gard. 1986;8:23–42.
34. Karasawa K, Tanaka R. C-banding study on centric fission in the chromosome of *Paphiopedilum*. Cytologia. 1980;45:97–102.
35. Karasawa K, Tanaka R. A revision of chromosome number in some hybrids of *Paphiopedilum*. Bull Hiroshima Bot Gard. 1981;4:1–8.
36. Karasawa K, Saito K. A revision of the genus *Paphiopedilum* (Orchidaceae). Bull Hiroshima Bot Gard. 1982;5:1–69.
37. Karasawa K, Aoyama M. Karyomorphological studies on two species of *Paphiopedilum*. Bull Hiroshima Bot Gard. 1988;10:1–6.
38. Lan T, Albert VA. Dynamic distribution patterns of ribosomal DNA and chromosomal evolution in *Paphiopedilum*, a lady's slipper orchid. BMC Plant Biol. 2011;11:126.
39. Wicker T, Taudien S, Houben A, et al. A wholegenomesnapshot of 454 sequences exposes the composition of the barleygenome and provides evidence for parallel evolution of genome size in wheatand barley. Plant J. 2009;59:712–22.
40. Deschamps S, Campbell MA. Utilization of next-generation sequencingplatforms in plant genomics and genetic variant discovery. MolBreeding.2010;25:553–70.
41. Macas J, Neumann P, Navratilova A. Repetitive DNA In the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics. 2007;8:427.
42. Kelly LJ, Leitch IJ. Exploring giant plant genomes withnext-generation sequencing technology. Chromosom Res. 2011;19:939–53.
43. Torres GA, Gong Z, Iovene M, et al. Organization and evolution of subtelomericsatellite repeats in the potato genome. G3-GenesGenom. Genet. 2011;1:85–92.
44. KellyLJ, Leitch AR, Fay MF, et al.Why size really matters when sequencing plantgenomes. Plant Ecol Divers.2012;5:415–25.
45. Heckmann S, Macas J, Kumke K, et al. The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. Plant J. 2013;73:555–65.
46. Harkess A, Mercati F, Abbate L, et al. Retrotransposon proliferation coincident with the evolution of dioecy in *Asparagus*. G3-GenesGenomGenet.2016;6:2679–85.
47. Choudhury RR, Neuhaus JM, Parisod C. Resolving fine-grained dynamics of retrotransposons: comparative analysis of inferential methods and genomic resources. Plant J. 2017;90:979–93.
48. de Jong JH, Fransz P, Zabel P. High resolution FISH in plants-techniques and applications. Trends Plant Sci. 1999;4:258–62.
49. Lim KY, Matyášek R, Lichtenstein CP, Leitch AR. Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section Tomentosae. Chromosoma. 2000;109:245–58.
50. Macas J, Kejnovský E, Neumann P, Novák P, Koblížková A, Vyskot B. Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. PLoS ONE.2011;6:e27335.
51. Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci Rep. 2016; 6:28333.
52. Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF. Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res. 2006;16:1252–61.

53. Ištvánek J, Jaroš M, Křenek A, Řepková J. Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). Am J Bot. 2014;101:327–37.
54. Macas J, Pozarkova D, Navratilova A, Nouzova M, Neumann P. 2000. Two new families of tandem repeats isolated from genus *Vicia* using genomicself-priming PCR. Mol Gen Genet. 2000;263:741–51.
55. Emadzade K, Jang TS, Macas J. Differential amplification of satellite PaB6 in chromosomally hypervariable*Prospero autumnale* complex (Hyacinthaceae). Ann Bot. 2014;114:1597–608.
56. Karasawa K, Aoyama M, Kamimura T. Karyomorphological studies on five rare species of *Paphiopedilum*. Orchidaceae AnnTsukuba Bot Gard. 1997;16:29–39.
57. Cox AV, Abdelnour GJ, Bennett MD, Leitch IJ. Genome size and karyotype evolution in the slipper orchids (Cypripedioideae: Orchidaceae). Am J Bot. 1998;85:681–7.
58. Ebihara A, Ishikawa H, Matsumoto S, et al. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. Am J Bot. 2005;92:1535–47.
59. Doležel J, Greilhuber J, Lucretti S, et al. Plant genome size estimation by flow cytometry: inter-laboratory comparison. Ann Bot. 1998;82:17–26.
60. Trávníček P, Ponert J, Urfus T, Jersakova J, Vrana J, Hribova E, Dolezel J, Suda J. Challenges of flow-cytometric estimation of nuclear genome size in orchids, a plant group with both whole-genome and progressively partial endoreplication. Cytometry A. 2015;87:958–66.
61. Altschul S, Madden T, Schaffer A, et al. Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402.
62. Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst Evol.2016;DOI https://doi.org/10.1007/s00606-016-1356-9.
63. Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010;11:378.
64. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013;29:792–3.
65. Smit AFA, Hubley R, Green P. RepeatMaskerOpen-4.0; 2013-2015 <http://www.repeatmasker.org>.
66. Sonnhammer ELL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 1995; 167:GC1–GC10.
67. Kearse M, Moir R, Wilson A, et al. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28:1647–9.
68. Macas J, Es Aros T, Nouzo M. PlantSat: a specialized database for plant satellite repeats. Bioinformatics. 2002;18:28–35.
69. Chung MC, Lee YI, Cheng YY, Chou YJ, Lu CF. Chromosomal polymorphism of ribosomal genes in the genus *Oryza*. Theor Appl Genet. 2008;116:745–53.
70. Lee YI, Chang FC, Chung MC. Chromosome pairing affinities in interspecific hybrids reflect phylogenetic distances among lady's slipper orchids (*Paphiopedilum*). Ann Bot. 2011;108:113–21.
71. Gerlach WL, Bedbrook JR. Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. 1979;7:1869–85.
72. Gerlach WL, Dyer TA. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res. 1980;8: 4851–65.
73. Sun Y, Skinner DZ, Liang GH, Hulbert SH. Phylogenetic analysis of *Sorghum* and related taxa using internal transcribed spacers of nuclear ribosomal DNA. Theor Appl Genet. 1994;89:26–32.
74. White TJ, Brus T, Lee S, Taylor J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ, editors. PCR protocols: a guide to methods and applications. San Diego: Academic Press; 1990. p. 315–22.
75. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673–80.
76. Ronquist F, Huelsenbeck JP. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 2001;19:1572–4.
77. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19:716–23.
78. Nylander JA. MrModeltest v2. Program distributed by the author: Evolutionary Biology Centre, Uppsala University; 2004.
79. Yang Z, Rannala B. Bayesian phylogenetic Monte Carlo method inference using DNA sequences: a Markov chain Monte Carlo method. Mol Biol Evol. 1997;14:717–24.
80. Swofford DL. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland; 2002.
81. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783–91.
82. Stöver BC, Müller KF. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. BMC Bioinformatics. 2010;11:1–9.