Characterization of Chromosome-Specific Microsatellite Repeats and Telomere Repeats Based on Low Coverage Whole Genome Sequence Reads in *Panax ginseng*

Nomar Espinosa Waminal1,2, Remnyl Joyce Pellerin1, Woojong Jang2, Hyun Hee Kim1*, Tae-Jin Yang2*

1Chromosome Research Institute, Department of Life Science, Sahmyook University, Seoul 01795, Korea
2Department of Plant Science, Plant Genomics and Breeding Institute, and Research Institute of Agriculture and Life Sciences, College of Agriculture and Life Sciences, Seoul National University, Seoul 08826, Korea

**ABSTRACT** Repetitive DNA elements are ubiquitous in plant genomes. Although repeats provide relevant information for cytogenetic, evolutionary, and genomic studies, identifying and characterizing their sequence and chromosomal distribution are not always easily achieved through conventional methods. However, a high-throughput identification of genomic repeats can be obtained with short reads from next-generation sequencing data. Here, we identified the telomeric and two chromosome-specific repeats in *Panax ginseng* using low-coverage whole genome sequence data. The telomeric repeat sequence is same with the canonical angiosperm sequence, (TTTAGGG)n, and localized mostly in every chromosome termini, except for an additional interstitial location in chromosome 10. A dinucleotide (GA) microsatellite, PgGA15, with total genome representation (GR) of more than 33 kb localized in the long arm of chromosome 20. An 11-bp minisatellite, Pgms1, with more than 58 kb of GR localized in the long arm of chromosome 1. This study provides chromosome-specific markers for cytogenetic studies in *P. ginseng*.

**Keywords** FISH, NGS, Cytogenetics, *Panax ginseng*, Repeatomics, Minisatellites

**INTRODUCTION**

Repetitive DNA elements are ubiquitous in plant genomes, and they are amplified or eliminated in a very rapid and dynamic fashion (Michael and Jackson 2013). They could be either dispersed or tandem repeats based on their physical organization in a genome (Kubis et al. 1998). Owing to these characteristics, they are responsible for significant genome size variations even between very closely related species (Tenaillon et al. 2010). Numerous studies have shown the indispensable roles of repeats in a genome which include genotypic and phenotypic variations, heterochromatin formation and environmental adaptation (Kalendar et al. 2000; Lippman et al. 2004; Wei et al. 2014). Because repeat amplification or elimination could happen before or after species divergence, repeats can be an excellent tool in studying evolution of genomes (Choi et al. 2014; Mehrotra and Goyal 2014; Wei et al. 2014). Although different repeats follow different evolutionary pathways, tandem repeats have been evolutionarily associated with other repetitive DNA like transposable elements and the intergenic spacers (IGSs) of the 45S rDNA genes in several plant species (Cheng and Murata 2003; Macas et al. 2003).

Tandem repeats can be classified as microsatellite, minisatellite or satellite according to their repeat unit and array size (2-5 bp in 10-100 units, 6-100 bp in 0.5-30 kb, and 150-400 bp in up to 100 Mb, respectively). Tandem

---

Received February 14, 2018; Revised February 20, 2018; Accepted February 20, 2018; Published March 1, 2018

*Corresponding author* Hyun Hee Kim, kimhh@syu.ac.kr, Tel: +82-2-3399-1715, Fax: +82-2-3399-1729

*Corresponding author* Tae-Jin Yang, tjyang@snu.ac.kr, Tel: +82-2-880-4547, Fax: +82-2-8873-2056

Copyright © 2018 by the Korean Society of Breeding Science

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
repeats are also excellent cytogenetic markers because they are often distributed in unique chromosomal regions, allowing easy discrimination of chromosomes (Mendes et al. 2011; Choi et al. 2014).

Conventional methods in identifying and characterizing repeats may be costly and tedious, often involving extensive molecular procedures while producing low throughput repeat identification and characterization (Dong et al. 2000). With the advancement of next-generation sequencing (NGS) technologies and bioinformatics pipelines to handle NGS data such as TAREAN (Novák et al. 2017), high-throughput and more efficient genome-wide repeat identification can be obtained even with low-coverage sequencing. This approach has been successfully demonstrated in some species like pea, banana and cabbage (Macas et al. 2007; Hribova et al. 2010; Novák et al. 2013; Novák et al. 2017; Perumal et al. 2017; Waminal et al. 2017).

*Panax ginseng* (Korean ginseng or Asian ginseng) is an important allotetraploid (2n = 4x = 48) perennial crop (Choi et al. 2014; Jang et al. 2017), highly priced for its ginsenoside, which is known to have a wide range of therapeutic effects (Court 2000; Park et al. 2012). In addition, its sensitivity to light, slow growth (∼four years/generation) and low seed yield (∼40 seeds/plant) have contributed to its high price and a multi-billion USD market (Baeg and So 2013). Although there has been a relatively long history of pharmacological studies in *P. ginseng* (Leung and Wong 2010), genomics and molecular cytogenetics have just gained pace in recent years, and several questions still remain to be answered like the origin of the extant *P. ginseng* genome and the pathways of specific repeat evolution in light of the *P. ginseng* genome function. In a previous study (Waminal et al. 2017), we reported about Pg167TR, a 167-bp satellite DNA that localized in all *P. ginseng* chromosomes with a distinct pattern in each chromosome. However, we have not identified any chromosome-specific repeat so far. Here, using low-pass NGS data, we validated telomere repeats and its distribution and further identified novel chromosome-specific repeats that will be useful for *P. ginseng* cytogenetic and evolutionary studies.

**MATERIALS AND METHODS**

Repeat identification and probe design

Approximately 0.03× of *P. ginseng* WGS reads were sampled from bulk NGS data deposited at the National Agricultural Biotechnology Information Center (NABIC, acc. code NN-0076-000001, https://nabic.rda.go.kr:2360/) was used to mine for microsatellite, minisatellite, and satellite repeats using the RepeatExplorer2 pipeline (Novák et al. 2013) after quality trimming as described in RepeatExplorer2 guidelines. Estimation of genome representation (GR) was carried out by multiplying the total number of repeats with the repeat period size followed by normalization of the size of raw reads used with the 3.6 Gbp genome size of *P. ginseng* (Waminal et al. 2016; Waminal et al. 2017). Read mapping and sequence characterization were carried out using CLC Assembly Cell ver. 4.21 and CLC Main Workbench ver. 7.9.1 using default parameters. A 30-bp oligo probe each was designed each for PgGA15, Pgms1, and PgTel and synthesized by Bioneer Corp. (http://www.bioneer.com/, Korea).

**FISH analysis**

Chromosome spreads were prepared according to Waminal et al. (2017). Thirty-two microliters of FISH hybridization master mix (50% formamide, 10% dextran sulfate, and 2× SSC), 25 ng of each FAM- and Cy3-modified oligoprobes, and deionized water to a total volume of 40 μL were added to slides. Chromosomal DNA on a glass slide was denatured at 80°C for 5 minutes after the addition of hybridization mix. Hybridization was carried out for 30 minutes at room temperature. Stringency washes were performed using 2× SSC at room temperature for 5 minutes, 0.1× SSC at 42°C for 20 minutes, and 2× SSC for 5 minutes at room temperature. The slides were dehydrated in an ethanol series of 70%, 90%, and 100%, air-dried, and counterstained with premixed 4',6-diamidino-2-phenylindole (DAPI) solution (1 μg/mL; DAPI in Vectashield, Vector Laboratories, Burlingame, CA, USA). Images were captured under a model BX53 fluorescence microscope (Olympus, Tokyo, Japan) equipped with a DFC365 FS CCD camera (Leica Microsystems, Wetzlar, Germany) and processed using Cytovision ver. 7.2 (Leica
RESULTS

Identification of novel tandem repeats using Illumina NGS data

NGS reads have numerous uses depending on one’s research objectives (Fig. 1A). Our interest is in using NGS reads for repeatomics, characterizing the totality of repeats in a genome for structural characterization (Waminal et al. 2016). In this study, we focused on micro- and minisatellites that make a considerable physical array that can be observable using molecular cytogenetic techniques such as FISH. To do that, individual Illumina reads were concatenated into a single contig by adding 50-bp Ns in between reads using a Perl script from RepeatExplorer2 (Novák et al. 2013). The single contig was scanned for tandem repeats using Tandem Repeats Finder (Benson 1999) and sorted and quantified with TRAP (Sobreira et al. 2006) (Fig. 1B). Five microsatellites (2 or 3-bp) and two minisatellites (7 and 11-bp) with >25 kb total bases were identified (Table 1). Out of these seven tandem repeats, three showed reproducible distinct chromosome-specific signals. We identified one microsatellite (PgGA15), and two minisatellites including the telomeric repeat of P. ginseng (Pgms1 and PgTel) (Table 1). While all have

Table 1. List of tandem repeats >25 kb identified in P. ginseng concatenated Illumina reads.

<table>
<thead>
<tr>
<th>No.</th>
<th>Sequence</th>
<th>Period size (base)</th>
<th>%AT</th>
<th>Total repeat units</th>
<th>Ave. repeat/locus</th>
<th>Total repeat loci</th>
<th>Total bases</th>
<th>Given name in this study</th>
<th>Chromosomal distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(TA)_n</td>
<td>2</td>
<td>100</td>
<td>44,836</td>
<td>0.032</td>
<td>11.8</td>
<td>3,790</td>
<td>89,187</td>
<td>None detected</td>
</tr>
<tr>
<td>2</td>
<td>(TAT)_n</td>
<td>3</td>
<td>100</td>
<td>28,869</td>
<td>0.031</td>
<td>9.3</td>
<td>3,091</td>
<td>86,167</td>
<td>None detected</td>
</tr>
<tr>
<td>3</td>
<td>(GA)_n</td>
<td>2</td>
<td>50</td>
<td>16,347</td>
<td>0.012</td>
<td>9.3</td>
<td>1,762</td>
<td>32,591</td>
<td>PgGA15</td>
</tr>
<tr>
<td>4</td>
<td>(TG)_n</td>
<td>2</td>
<td>50</td>
<td>13,464</td>
<td>0.010</td>
<td>9.3</td>
<td>1,446</td>
<td>26,890</td>
<td>None detected</td>
</tr>
<tr>
<td>5</td>
<td>(TCT)_n</td>
<td>3</td>
<td>67</td>
<td>8,370</td>
<td>0.009</td>
<td>6.7</td>
<td>1,252</td>
<td>25,204</td>
<td>None detected</td>
</tr>
<tr>
<td>6</td>
<td>(TTAGGG)_n</td>
<td>7</td>
<td>57</td>
<td>6,527</td>
<td>0.016</td>
<td>11.9</td>
<td>550</td>
<td>45,675</td>
<td>PgTel</td>
</tr>
<tr>
<td>7</td>
<td>(ACATTCTTGAT)_n</td>
<td>11</td>
<td>73</td>
<td>5,427</td>
<td>0.021</td>
<td>8.7</td>
<td>627</td>
<td>58,266</td>
<td>Pgms1</td>
</tr>
</tbody>
</table>

Fig. 1. Diagram in identifying repeats in P. ginseng. (A) A general non-comprehensive scheme of different applications for NGS data. Following this scheme, we used 0.3× of P. ginseng Illumina reads for tandem repeat analysis in this study. (B) A simple workflow for quantifying micro- and minisatellites.

\(^{a)}\)Genome representation.
potential of becoming cytogenetic markers, no discernable FISH signals were observed using oligonucleotide probes for other repeats.

Identification of GA satellite blocks and chromosomal localization

PgGA15 is a GA dinucleotide microsatellite that is estimated to exist in more than 16,000 copies or about 33 kb in total base-pairs or 0.012× of the *P. ginseng* genome (Table 1). It has an average repeat number of 9.3 repeats (or 18.6 bp) per locus in more than 1,700 loci. The average repeat length per locus is less than 20% of the size of one read length, indicating shorter arrays of PgGA15 in the genome, characteristics of microsatellites. A 30-bp oligonucleotide probe, PgGA15_OP, was designed and modified at the 5’ with 6-carboxyfluorescein (FAM) fluorescent dye (Table 2). FISH analysis with PgGA15_OP revealed an intense signal on the DAPI band on one long chromosome. Based on distribution pattern of Pg167TR on each chromosome, we confirmed the chromosome number which show the PgGA15 signal. The PgGA15 hybridized on the long arm of chromosome 20, indicating a concentration of this microsatellite at this genomic region (Fig. 2).

Identification of telomere repeats and chromosomal localization

PgTel is a 7-bp minisatellite which corresponded with the canonical angiosperm telomere repeat TTTAGGG and was identified and estimated at more than 6,000 copies or

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Sequence</th>
<th>Length</th>
<th>Used label</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PgGA15_OP</td>
<td>GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA</td>
<td>30 bases</td>
<td>5’-FAM</td>
</tr>
<tr>
<td>2</td>
<td>Pgms1_OP</td>
<td>ACATTCTTGATACATTCTTGATACATTCTT</td>
<td>30 bases</td>
<td>5’-Cy3</td>
</tr>
<tr>
<td>3</td>
<td>PgTel_OP</td>
<td>TTTAGGGTTTAGGTAGGTAGGTAGGTAGGTAGG</td>
<td>34 bases</td>
<td>5’-Cy3</td>
</tr>
</tbody>
</table>

**Table 2.** List of oligonucleotide probes used for FISH.

![Fig. 2](image-url) **Fig. 2.** FISH analysis of PgGA15 on *P. ginseng* chromosomes. (A-C) Raw images of DAPI stained chromosomes, PgGA15, and Pg167TR signals, respectively. (D and E) Pseudo-colored images of PgGA15 signal and Pg167TR. The PgGA15 signal was localized on the long arm of chromosome 20. (F) Merged images of PgGA15 (arrowheads) and Pg167TR signals. Bar = 10 μm.
about 46 kb or 0.016× of the genome (Table 1). It has an average repeat number of 11.9 repeats (or 83.3 bp) per locus in more than 550 loci. The average repeat length per locus is about 80% of the size of one read length, suggesting that PgTel is from much longer arrays in the actual genome. A 34-bp oligonucleotide probe, PgTel_OP, which is about 5 copies of the unit, was designed and modified at the 5’ with Cy3 fluorescent dye (Table 2). FISH analysis with PgTel_OP revealed a distribution at the chromosome termini of all P. ginseng chromosomes. However, an additional interstitial PgTel signal at the most intense DAPI band in the long arm of chromosome 10 was also observed (Fig. 3).

**Identification of an 11-bp minisatellite and chromosomal localization**

Pgms1 is an 11-bp minisatellite that is estimated to about 5,000 copies or 58 kb, representing about 0.021× of the genome (Table 1). It has an average repeat number of 8.7 repeats (or 95.7 bp) per locus in more than 600 copies. The average repeat length per locus is almost as long as one read length, suggesting arrays of longer lengths for Pgms1. A 30-bp oligonucleotide probe, Pgms1_OP, which is about 2.5 copies of the repeat unit, was designed and modified at the 5’ end with Cy3 fluorescent dye (Table 2). FISH analysis with Pgms1_OP showed a distinct signal at the interstitial region of one long chromosome. Using Pg167TR as a reference marker of P. ginseng chromosomes, we confirmed that Pgms1 located at chromosome 1 (Fig. 4).

**DISCUSSION**

The novel micro- and minisatellites identified in this
Fig. 4. FISH analysis of Pgms1 and Pg167TR on *P. ginseng* chromosomes. (A-C) Raw images of Pgms1, Pg167TR, and DAPI stained chromosomes, respectively. (D-F) Pseudo-colored signals from A-C, respectively, indicating the localization of Pgms1 on the long arm of chromosome 1. (F) Merged images of Pgms1 (arrowheads) and Pg167TR. Bar = 10 μm.

Fig. 5. Karyotype idiogram of *P. ginseng* showing different cytogenetic markers. Navy blue, orange, red, and green bands represent DAPI, Pg167TR, 45S rDNA, and 5S rDNA as described in Waminal et al. (2017). The DAPI bands in the long arm of chromosomes 1 and 20 are the sites for Pgms1 (patterned arrowhead) and PgGA15 (grey arrowhead) tandem repeats, respectively. PgTel repeats are depicted by magenta circles. An interstitial PgTel locus localized in the intense DAPI band in the long arm of chromosome 10 (green arrowhead).

Study were generated from concatenated 101-bp Illumina reads with 50-bp ‘N’ spacer, so array identification is limited to individual 101-bp reads. Although this limitation may not identify full-length repeat arrays, it is enough to identify short micro- and minisatellite repeat units. Consequently, the GR values may be an underestimate of the actual coverage in the genome. PgGA15 with short average repeat length per locus, compared with PgTel and Pgms1, may be physically organized as dense clusters in chromosome 20 with short arrays separated by non PgGA15 sequences. Whereas Pgms1 may be either of similar pattern with PgGA15 but of longer units or perhaps just a few long arrays but were separated by the short reads, as each read can be considered as a single locus after separation with the 50-bp N spacer between reads.

All identified repeats have more than 50% AT content and all hybridized to DAPI bands in their respective locus. The high AT content of these repeats may have contributed to the DAPI bands, as DAPI is known to preferentially intercalate at AT-rich DNA (Schweizer 1976). However, not all DAPI bands may be solely composed of one type of AT-rich repeats, but could be of several types, as in the case
of the Pgms1 locus, which localized other minisatellites.

In addition, telomeric repeats are usually found at most of chromosome termini. However, several studies have reported interstitial localization of telomeric repeats, like that observed in the interstitial region of chromosome 10. Interstitial telomeric sites have been demonstrated to be recombination hot-spots, genome rearrangement breakpoints, and even shown to be functional in several species (Day et al. 1998; Hastie and Allshire 1989; Park et al. 1992; Tremousaygue et al. 1999; Uchida et al. 2002). P. ginseng have undergone two major rounds of genome duplications (Choi et al. 2013). This interstitial telomeric site on P. ginseng chromosome 10 may be just a relic of genomic rearrangements that occurred in P. ginseng, perhaps prior to the divergence between P. ginseng and P. quinquefolius, as both species share a similar chromosome bearing interstitial telomeric sequence (Fig. 3) or may still serve some unknown functions.

Recently, our studies revealed the evolution of major repeats in Panax genus (Jang et al. 2017; Lee et al. 2017). We also provided distribution pattern of these major repeats and established chromosome identification system using these repeats (Waminal et al. 2017). However, ginseng has 24 chromosome pairs (48 chromosomes) with similar size and shape that ask easy and reliable chromosome discrimination method. Here, we have provided chromosome-specific cytogenetic markers (Fig. 5 and Table 1) that will be useful tools for cytogenetic studies in P. ginseng and related species. It has improved the karyotype of P. ginseng. Future investigation on other chromosome-specific repeats and their impact on the evolution and function of the P. ginseng genome will be necessary for the molecular cytogenetic and genomics in ginseng.

ACKNOWLEDGEMENTS

This work was carried out with the support of the Cooperative Research Program for Agriculture Science & Technology Development (PJ013119), Rural Development Administration, Korea.

REFERENCES

The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol. Biol. 36: 821-831.


Mendes MM, Da Rosa R, Giuliano-Caetano L, Dias AL. 2011. Karyotype diversity of four species of the incertae sedis group (Characeae) from different hydrographic basins: analysis of AgNORs, CMA3 and 18S rDNA. Genet. Mol. Res. 10: 3596-3608.


Bioinformatics 29: 792-793.


