

The genome size estimation and telomeric-repeat motifs of thirteen Arctic plants

Jun Kim

Seoul National University

Moonkyo Kim

Korea Polar Research Institute

Yoo Kyung Lee (✉ yklee@kopri.re.kr)

Korea Polar Research Institute

Research Article

Keywords: Arctic plant, chromosome number, genome size, ploidy level, Svalbard, telomeric-repeat motif

Posted Date: July 1st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1783809/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

As a preliminary step of genome analysis, we produced a total of 28–44 Gb DNA sequencing data of whole-genome for thirteen Svalbard plants. We estimated genome sizes using these data with a k-mer-based computation tool and found that genome sizes of eight species ranged from 180–894 Mb: *Cochlearia groenlandica* 180 Mb, *Dryas octopetala* 204 Mb, *Eriophorum scheuchzeri* ssp. *arcticum* 306 Mb, *Oxyria digyna* 352 Mb, *Salix polaris* 383 Mb, *Betula nana* 689 Mb, *Cassiope tetragona* 798 Mb, and *Silene uralensis* ssp. *arctica* 894 Mb. We could not estimate the genome size of the other five species. We also analyzed variations in DNA sequences by identifying putative telomeric-repeat motifs in the thirteen Svalbard plants. Eleven out of thirteen species contained the canonical plant telomeric-repeat motif, TTTAGGG (*Arabidopsis*-type), whereas *S. oppositifolia* contained the *Chlamydomonas*-type telomeric-repeat motif (TTTTAGGG) and *P. dahlianum* had a novel telomeric-repeat motif, TTCAGGG. These findings provide a quantitative guideline for whole-genome sequencing analysis of Arctic plants in the future. The findings in this study provide a quantitative guideline for whole-genome sequencing analysis of Arctic plants, and they also show the potential of Arctic plants to be a new source of telomere diversity.

Introduction

Arctic plants live in a vulnerable environment. They are exposed to a short growing season, temperature fluctuations, strong winds, and oligotrophic soil (Lee 2020). Sometimes their habitats are disturbed by the overflow of glacial melting water (Kim et al. 2022). The formation of thaw ponds by permafrost thawing and the changes in temperature and precipitation result in vegetation shifts and/or community trait change (van der Kolk et al. 2016; Bjorkman et al. 2018). Arctic plants are driven into competition with subarctic plants and influenced by boreal animals such as moose, beaver, red fox, and boreal birds expanding into the Arctic tundra (Speed et al. 2021). Before Arctic tundra plants become a minor population or disappear completely, we should understand the remarkable legacy of their adaptation to the harsh Arctic environment.

Genomic analysis is one of the ways to unveil the life phenomena of these Arctic plants. The genome is the basis for a variety of follow-up studies and a source of insight into biological processes. Genome-based discovery of plant characteristics includes the origins of polyploidy and whole-genome duplication, reproductive diversity such as monoecious and dioecious flowers, self-incompatibility, and evolutionary history (Michael and Jackson 2013; Slotte et al. 2013). Genomic information also reveals the regulation of the development and differentiation of roots, leaves, flowers, and embryos (Du et al. 2018; Radoeva et al. 2019; Beaudry et al. 2020; Deja-Muyllé et al. 2020). The genome can provide essential information on plant survival, such as susceptibility to fungal infections, disease resistance, and resistance to low temperatures or drought (Guo et al. 2015; Mahmood et al. 2019; Goyal et al. 2020; Ke et al. 2020). The genome of Arctic plants may reveal the secrets of these plants being able to survive in barren environments abandoned by other plants.

Estimating the genome size is the basic data for genome analysis. Based on the estimated genome size, the amount of data to be sequenced for whole-genome analysis can be determined. Genome size is used in population studies related to ploidy level, gene flow, and genetic diversity to understand the distribution and adaptation of the populations along with history (Cuba-Díaz et al. 2017). Genome size, chromosome number, and DNA amounts have been reported on Antarctic plants (Bennett et al. 1982; Cuba-Díaz et al. 2017; Pascual-Díaz et al. 2020; Siljak-Yakovlev 2020) and Arctic plants (Nowak et al, 2020), but the genome information is still lacking on most Arctic plants.

Telomere is composed of highly conserved and tandemly repeated sequences, stabilizing chromosomes. Telomeres provide the structural basis for chromosomes in eukaryotes (Peška and Garcia 2020), and take part in chromosome recognition and pairing at the beginning of meiosis (Aguilar and Prieto 2021). Because the changes in telomere sequences affect the stability of the whole genome, telomeric-repeat motifs are highly conserved in animals. However, plants exhibit extreme diversity in telomeric-repeat motifs compared to animals (Fulnecková et al. 2013). The representative telomeric-repeat motif of plants is TTTAGGG, called *Arabidopsis*-type (Richards and Ausubel 1988). Several telomeric-repeat motifs have been reported from plants: TTCAGG and TTTCAGG (*Genlisea*-type; Tran et al., 2015), TTTTTTAGGG (*Cestrum*-type; Peška et al., 2015), CTCGGTATGGG in *Allium* species (Fajkus et al. 2016). Even some plants possess the vertebrate-type telomeric sequence, TTAGGG (Weiss and Scherthan 2002; Sýkorová et al. 2003; Sýkorová et al. 2006).

A telomeric-repeat motif of a specific organism is analyzed based on the genome. But the genomes of Arctic tundra plants have not yet been reported on. By 2020, the genomes of a total of 702 vascular plants have been published (Sun et al. 2022), but few of them represented tundra plants.

Svalbard is an archipelago that extends between 74°–81° north latitudes and 10°–35° east longitudes. Most of Svalbard was ice-covered during the last ice age 11,000 years ago (van der Bilt and Lane 2019), and plants seemed to have migrated and settled down recently. A large proportion of Svalbard is still covered by ice, thus the land where plants can grow is mostly concentrated along the coast and ice-free valleys. More than 200 vascular plants have been reported from Svalbard, and among them, 48 plants are rare species on the Red List (Lee 2020). So far, the genome information of Svalbard plants has not yet been reported.

As a preliminary step of genome analysis, we have analyzed the genome size of thirteen tundra plants commonly found in Svalbard by using next-generation sequencing (NGS). We also examined whether any novel or noncanonical telomeric-repeat motif has evolved in tundra plants. Among the thirteen plants, *Cassiope tetragona*, *Dryas octopetala*, and *Salix polaris* were shrubs observed at the climax of the succession in Svalbard. *Saxifraga oppositifolia* and *Silene acaulis* are pioneer plants in glacier retreat areas or abandoned coal piles (Těšitel et al. 2014; Oh and Lee 2021). *Bistorta vivipara*, *Cochlearia groenlandica*, *Oxyria digyna*, *Papaver dahlianum* and *Silene uralensis* ssp. *arctica* are widely distributed forbs, and *Eriophorum scheuchzeri* ssp. *arcticum* is a common graminoid in Svalbard. This paper provides useful information for future whole-genome analysis and telomere research of Arctic plants.

Materials And Methods

Sample collection and preprocessing

Arctic plant samples were collected from seven sites in Spitsbergen Island of the Svalbard archipelago: Adventdalen, Blomsterdalen, Endalen, Longyearbyen, Stuphallet, Svalbard Airport, and Svalbard Governor. The sampling period was from July 25 to August 3, 2021. The thirteen species dominant in Svalbard were selected (Fig. 1), and only the amount of leaves required for analysis (0.1–2.0 g) was collected to minimize the impact on the local population. The collected samples were immediately put in a yellow envelope and dried with silica gel in the field. The plant leaves were lyophilized in the envelopes at the Dasan research station and stored at 4°C until they were transferred to Korea in 15 mL conical tubes.

DNA extraction

DNA was extracted using GeneAll® Exgene™ Plant SV mini kit (GeneAll Biotechnology Co. Ltd., Seoul, Korea). The DNA extraction method was carried out by modifying the protocol. As a pretreatment operation, the tissue of the plant sample was pulverized using a mortar to obtain a powder sample. The powder sample of 50–80 mg was placed in a 2.0 mL microcentrifuge tube, and 400 µL of PL buffer and 4 µL of RNase A (100 mg / mL) were added and then vortexed. The tubes were put in a constant temperature water bath at 65°C for 3–4 hours with vortexing 6–8 times. After adding 140 µL of PD buffer, it was placed on ice for 7 minutes and centrifuged at 14,000 rpm for 5 minutes and 30 seconds. The supernatant was placed in an EzSep™ filter (blue) column and centrifuged at 14,000 rpm for 2 minutes. BD buffer solution of 1.5 times the volume was added to the filtrate, and pipetting was performed about 15 times. The mixture was placed on a GeneAll® SV column and centrifuged at 8,000 rpm for 1 minute. And then 700 µL of CW buffer was added to the SV column, centrifuged at 8,000 rpm for 1 minute, and 300 µL of CW buffer was added and centrifuged once more at 14,000 rpm for 2 minutes. The SV column was transferred to a new 1.5 mL microcentrifuge tube, and 50 µL of AE buffer was added to the column membrane, dissolved at room temperature for 1 to 1.5 hours, and centrifuged at 8,000 rpm for 1 minute. This process was performed once more to obtain a total volume of 100 µL. The concentration and quality of the extracted DNA were checked through NanoPhtometer® NP80 (Implen GmbH, Munich, Germany), and DNA quality was checked through electrophoresis.

NGS sequencing

NGS sequencing was performed by DNA Link (Seoul, Korea). DNA library for sequencing was produced using Illumina's DNA sample prep kit, and the completed DNA library was sequenced with a paired-end read of 151 bp using Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA). The prepared library was loaded into the flow cell mounted on the cBot, a cluster generation device. The template was amplified using the bridge amplification method, and the sequencing reaction started in the flow cell. In the sequencing reaction, dNTPs with different fluorescent dyes were synthesized one base at a time in one cycle, and the fluorescence of the used nucleotide was measured to analyze the sequence. The raw data

were submitted to the KoNA BioProject database (https://www.kobic.re.kr/kona/go_browse_bioproject) under accession number PRJKA220343.

Genome size estimation and assembly

KAT (version 2.4.1) was used to estimate the genome sizes of the thirteen Svalbard plants (Mapleson et al. 2017). The command *kat hist -m 27* was applied to all Illumina sequencing data that we prepared. Platanus Assembler (version 1.2.4) was used to assemble Illumina reads of each species (Kajitani et al. 2014). First, we trimmed off low-quality regions of the Illumina reads using *platanus_trim*, and found overlaps among these trimmed reads using *platanus_assemble*. After these steps, we additionally merged the resulting contigs into scaffolds using *platanus_scaffold*.

Analysis of telomeric-repeat motifs

Telomeric-repeat motifs should be found as sequential and continuous repetitive sequences in a single read. To analyze telomeric-repeat motifs, we subsampled 20 million reads from our Illumina reads of the thirteen Svalbard plants. We only used 60-bp regions of each read by trimming off the starting 10 bp and the last 81 bp. We then counted the reads containing the telomeric-repeat motifs, sequentially. The telomeric-repeat units are TTTAGGG (Arabidopsis-type), TTAGGG (vertebrate-type), TTTTAGGG (*Chlamydomonas*-type), TTCAGG (*Genlisea*-type), TTTTCAGG (*Genlisea*-type), TTTTAGG (*Klebsormidium*-type), and TTTTTTAGGG (*Cestrum*-type). After this analysis, we also found that *Papaver dahlianum* has a novel telomeric-repeat motif, TTCAGGG, which was similar to *Genlisea*-types. We also counted TTCAGGG (*Papaver*-type) from all thirteen plant reads.

Results

We produced DNA sequencing data for the thirteen Svalbard plants to estimate their genome size. A total of 28–44 Gb of DNA sequencing data was produced (Table 1), which was suitable for estimating genome sizes of approximately 1 Gb. We estimated genome sizes using these data with a k-mer-based computation tool and found that genome sizes of eight species ranged from 180–894 Mb: *Cochlearia groenlandica* 180 Mb, *Dryas octopetala* 204 Mb, *Eriophorum scheuchzeri* ssp. *arcticum* 306 Mb, *Oxyria digyna* 352 Mb, *Salix polaris* 383 Mb, *Betula nana* 689 Mb, *Cassiope tetragona* 798 Mb, and *Silene uralensis* ssp. *arctica* 894 Mb. The estimated genome sizes of *Bistorta vivipara*, *Papaver dahlianum*, *Polemonium boreale*, and *Silene acaulis* were lower than 10 Mb. The genome size of *Saxifraga oppositifolia* could not be estimated by KAT.

Table 1

The chromosome numbers, ploidy levels, genome size (haploid C-values), genome size range of plants in the Family, estimated genome size, and draft genome assembly size of thirteen Svalbard plants. ND, Not determined; Chromosome numbers and ploidy levels (<https://svalbardflora.no/>); Family genome size (<https://cvalues.science.kew.org/>); The genome size of individual plants on the basis of genome sequencing in the same genus (Sun et al. 2022).

Scientific Name	Chromosome number (Ploidy level)	Genome size of plants in the Family (1C, Mb)	Short-read data size (bp)	Estimated genome (Mb)
Poales		196–9657.9		
Cyperaceae				
<i>Eriophorum scheuchzeri</i> <i>ssp. arcticum</i>	2n = 58 (ND)		34,591,174,828	306
Ranunculales		529.2–8722		
Papaveraceae				
<i>Papaver dahlianum</i>	2n = 70 (10x)	<i>P. somniferum</i> 2870	34,339,113,548	ND
Caryophyllales		392–5782		
Caryophyllaceae				
<i>Silene acaulis</i>	2n = 24 (2x)	<i>S. latifolia</i> 2800	27,861,778,590	ND
<i>Silene uralensis</i> ssp. <i>arctica</i>	2n = 24 (2x)		44,301,561,122	894
Polygonaceae		294–6110.3		
<i>Bistorta vivipara</i>	2n = 77–132 (Polyploidy)		37,899,315,444	ND
<i>Oxyria digyna</i>	2n = 14 (2x)		33,649,545,774	352
Saxifragales		553.7–2450		
Saxifragaceae				
<i>Saxifraga oppositifolia</i>	2n = 26, 39, 52 (2x, 3x, 4x)		32,263,321,682	ND

Scientific Name	Chromosome number (Ploidy level)	Genome size of plants in the Family (1C, Mb)	Short-read data size (bp)	Estimated genome (Mb)
Brassicales		156.8–4639.3		
Brassicaceae				
<i>Cochlearia groenlandica</i>	2n = 14 (2x)		41,139,187,752	180
Malpighiales		347.9–842.8		
Salicaceae				
<i>Salix polaris</i>	2n = 114 (6x)	<i>S. brachista</i> 420 <i>S. dunnii</i> 376 <i>S. matsudana</i> 656 <i>S. suchowensis</i> 425 <i>S. viminalis</i> 360	34,654,183,202	383
Rosales		98–3577		
Rosaceae				
<i>Dryas octopetala</i>	2n = 18 (2x)	<i>D. drummondii</i> 253	37,392,065,070	204
Fagales		396.9–2533.3		
Betulaceae				
<i>Betula nana</i> ssp. <i>nana</i>	2n = 28 (2x)	<i>B. nana</i> 450 <i>B. pendula</i> 440	33,644,296,108	689
Ericales		465.5–29302		
Ericaceae				
<i>Cassiope tetragona</i>	2n = 26 (2x)		30,504,013,132	798
Polemoniaceae		1288.7–6811		
<i>Polemonium boreale</i>	2n = 18 (2x)		29,289,298,766	ND

We also analyzed variations in DNA sequences by identifying putative telomeric-repeat motifs in the thirteen Svalbard plants (Fig. 2). Possible telomeric-repeat motifs of every species were successfully discovered, and eleven out of thirteen species contained the canonical plant telomeric-repeat motif, TTTAGGG (*Arabidopsis*-type). Impressively, the remaining two species, *P. dahlianum* and *S. oppositifolia*, were found to have distinct, noncanonical plant telomeric-repeat motifs, TTCAGGG and TTTTAGGG, respectively. The telomeric-repeat motif of *S. oppositifolia* was known to be the *Chlamydomonas*-type telomeric-repeat motif (Fulnečková et al. 2012). However, the TTCAGGG telomeric-repeat motif of *P. dahlianum* was novel.

Discussion

We produced DNA data of thirteen Svalbard plant species and estimated the genome sizes of eight species out of the thirteen. The genome sizes estimated by flow cytometry and by k-mer analysis of genome sequences were similar in *Salix brachista*: 400 Mb and 421 Mb, respectively (Chen et al. 2019). Because there is no information on the genome data of the eight Svalbard species, to check the accuracy of our genome size, we listed the range of genome sizes using flow cytometry for the Family the plant belonged to (Table 1; Pellicer and Leitch 2019). Draft genome assembly has never been reported for these Svalbard plants, and *Betula nana* of the United Kingdom is the only plant with genome sequencing data (Wang et al. 2013). As the other species had no genome data, we listed the genome sizes of individual plants in the same genus on the basis of genome sequencing data (Table 1; Sun et al. 2022).

Cochlearia groenlandica is a diploid plant with 14 chromosomes, common in Svalbard and Greenland (Luka et al., 2022). The genome size of *C. groenlandica* estimated in this study was 180 Mb, which was smaller than those of *Cochlearia* species ranged 196–735 Mb estimated by flow cytometry (Peer et al. 2003; Kochjarova et al. 2006; Lysak et al. 2009). *Cochlearia* plants with a large DNA content have more chromosome numbers: *C. tatrae* (2n = 42; 1C = 491 Mb), *C. borzaeana* (2n = 48; 1C = 683 Mb), *C. danica* (2n = 42; 1C = 686 Mb), and *C. officinalis* (2n = 32; 1C = 735 Mb) (Kochjarova et al. 2006; Lysak et al. 2009). The genome size of *C. groenlandica* was similar to 196 Mb of *C. pyrenaica* with 12 chromosomes.

Dryas octopetala is a diploid plant with 18 chromosomes, a common prostrate shrub forming a dense mat in the Arctic. The *Dryas* genus comprises of three species: *D. drummondii*, *D. integrifolia* and *D. octopetala*. Among them, *D. drummondii* contains root nodules while the others don't (Billault-Penneteau et al. 2019). The genome size of *D. drummondii* was reported to be 253 Mb, which was obtained through whole genome sequencing (Griesmann et al. 2018). The genome size of *D. octopetala*, 204 Mb, was smaller than that of *D. drummondii*. The genome size of *D. octopetala* was estimated as 567 Mb from the fluoresce of nuclei (Dickson et al. 1992), which seemed to be collected from North America. It is not clear whether this difference in genome size is due to differences in experimental methods or whether the plants are taxonomically different. Further study such as a comparative analysis of the genomes of Svalbard plants and North American plants is therefore needed for clarification.

Eriophorum scheuchzeri ssp. *arcticum* is a diploid plant with 58 chromosomes, a circumpolar species living in the High Arctic. Its genome size, 306 Mb, is relatively small in the Cyperaceae (Table 1). The genome sizes of *E. angustifolium* and *E. vaginatum* were 587 Mb and 489 Mb, respectively (Grime and Mowforth 1982), and their chromosome number was both 58 (Love and Love 1981). The genome size of *E. scheuchzeri* ssp. *arcticum* was smaller than that of *Eriophorum* plants with the same chromosome number.

Oxyria digyna is a diploid plant with 14 chromosomes, a common forb in the Arctic and high mountains. The genus *Oxyria* has only three species, and the genome size has not been reported on any of them. The genome sizes of the closest genus, *Rumex* species, ranged from 470–6,110 Mb, which is larger than 352 Mb of *O. digyna*.

Betula nana ssp. *nana* is a diploid plant with 28 chromosomes, a cold-adapting circumpolar species. The triploid *Betula* plant with 42 chromosomes was reported, which was a hybrid of *B. nana* and *B. pubescens* (Anamthawat-Jónsson et al. 2010). The genome size of *B. nana* ssp. *nana* analyzed in this study was 689 Mb, which is similar to the DNA content of the triploid, 670.8 Mb (Anamthawat-Jónsson et al. 2010). The plastid DNA of *B. nana* living in Svalbard shared its characteristics with *B. pubescens* (Eidesen et al. 2015). Therefore, the genome size estimated in this study seemed to be that of a triploid hybrid plant. Further study is required to clarify the ploidy level, morphology, and chromosome number of Svalbard *B. nana* ssp. *Nana*.

Silene is a giant genus with nearly 900 species. The DNA content of six species out of 900 was analyzed using flow cytometry and they were more than 900 Mb: *S. ciliate* 924 Mb; *S. vulgaris* 1,100 Mb; *S. pendula* 1,149 Mb; *S. rupestris* 1,663 Mb; *S. latifolia* female 2,802 Mb; *S. latifolia* male 2,861 Mb; *S. chalconica* 3,223 Mb (Siroký et al. 2001). The genome size of *S. uralensis* ssp. *arctica*, 894 Mb, estimated in this study was smaller than those of *Silene* species. We could not estimate the genome size of *S. acaulis*, and it is probably because the genome size is over 1 Gb.

Salix species are creeping shrubs that live in the Arctic and the alpine. The estimated genome size of *S. polaris* was 383 Mb, which is similar to that of two *Salix* species based on the whole genome sequence data: *S. brachista* 420 Mb, *S. dunnii* 376 Mb, *S. matsudana* 656 Mb, *S. suchowensis* 425 Mb, and *S. viminalis* 360 Mb (Chen et al. 2019; Almeida et al. 2020; Wei et al. 2020; Zhang et al. 2020; He et al. 2021). These *Salix* species were all diploid plants with 38 chromosomes, on the contrary, *S. polaris* is a hexaploid species with 114 chromosomes (Table 1). Due to the difference in the ploidy level, the C-value of hexaploid *S. polaris* was 1,148 Mb.

The four Svalbard plants — *C. groenlandica*, *D. octopetala*, *E. scheuchzeri* ssp. *arcticum*, and *O. digyna* — had small genome sizes (< 500 Mb), which could be good candidates to assemble high-quality, reference-level genomes using current long-read sequencing technologies.

Unfortunately, we failed to estimate the genome size for five out of the thirteen species, which suggests that these five Svalbard plants may have a much larger genome size than 1.0 Gb. Indeed, of these five

species, *S. oppositifolia* was known to have a genome size of over 1.4 Gb (Loureiro et al. 2013). The sequencing data in this study was just 32 Gb, as such it may not be enough to estimate such a large genome. The genome sizes of relative plants of *Polemonium boreale* and *Silene acaulis* are much larger than 1 Gb (Table 1). *Bistorta vivipara* and *Papaver dahlianum* are polyploidy plants. To analyze the genomes of these plants, sufficiently large sequence data will be required.

The genome size of plants is proportional to the amount of noncoding DNA (Barakat et al. 1997), and the noncoding DNA is mainly composed of repetitive DNA, such as transposable elements and telomeric repeats (Kubis et al. 1998). We have analyzed telomeric-repeat motifs from the genome sequencing data of the thirteen Svalbard plants. We discovered a novel telomeric-repeat motif TTCAGGG in *Papaver dahlianum*, which again emphasizes the importance of investigating poorly studied species in the polar regions.

The DNA amount in a nucleus has been considered to be closely related to nuclear and cell sizes, and negatively correlated with cell division rate (Bennett and Leitch 2005). Bennett (1972) suggested that DNA amount is positively correlated with minimal generation time (MGT), which is the minimum time required to produce the first mature seed since germination. The seed weight of legume species and the seed dry mass of *Allium* species showed a positive correlation with the DNA amount (Bennett and Leitch 2005). As plant growth is affected by environmental factors such as temperature and moisture, MGT changes according to the environment; in the Arctic tundra, with low temperature and short growing season, the MGT becomes long. If the MGT of a plant is longer than one year, the plant should be perennial. In fact, most Arctic tundra plants are perennial.

The genome size of Svalbard plants analyzed in this study was smaller than that of plants belonging to the same genus or Family. Small amounts of DNA in polar plants have also been observed in Antarctic and sub-Antarctic plants (Bennett et al. 1982). It is suggested that there is a maximum DNA amount per diploid genome that allows plants to survive in extreme environments with low temperatures and a short growing season (Bennett et al. 1982; Knight et al. 2005). Plants with higher DNA amounts than the maximum DNA amount are supposed to be gradually excluded, and the DNA amount may determine the latitudinal limits where the plants can be distributed (Bennett et al. 1982). The plants under global and local conservation concerns in United Nations Environment Programme World Conservation and Monitoring Centre (UNEP-WCMC) Species Database (<http://quin.unep-wcmc.org/isdb/taxonomy/>) posed larger genome sizes than no concerned plants (Vinogradov 2003). In the Arctic where temperature increases faster than anywhere else on Earth, and moisture condition changes dramatically in permafrost, genome size could be one of the prioritized criteria to be considered to conserve endangered tundra plants.

The findings in this study provide a quantitative guideline for whole-genome sequencing analysis of Arctic plants. They also show the potential of Arctic plants to be a new source of telomere diversity. As we used only thirteen species among the multitude of Arctic plants, it is anticipated that we could have

elucidated much more unprecedented genetic and genomic variation if we study more arctic plants. It will deepen our understanding of Arctic plants and their genome evolution.

Declarations

ACKNOWLEDGEMENT This research was supported by Korea Polar Research Institute (PE22450). JK was supported by a grant from the National Research Foundation of Korea funded by the Korean government (MEST) [2019R1A6A1A10073437]. We thank professor Sangkyu Park, and Mr. Youngil Ryu for their help to collect the leaf samples.

Author Contributions JK and YKL conceived and designed the research. YKL collected the plant samples. JK and MK conducted experiments and analyzed data. JK, MK, and YKL wrote the manuscript. All authors read and approved the manuscript.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Aguilar M, Prieto P (2021) Telomeres and subtelomeres dynamics in the context of early chromosome interactions during meiosis and their implications in plant breeding. *Front Plant Sci* 12:672489. <https://doi.org/10.3389/fpls.2021.672489>
2. Almeida P, Proux-Wera E, Churcher A, Soler L, Dainat J, Pucholt P, Nordlund J, Martin T, Rönnerberg-Wästljung AC, Nystedt B, Berlin S, Mank JE (2020) Genome assembly of the basket willow, *Salix viminalis*, reveals earliest stages of sex chromosome expansion. *BMC Biol* 18:78. <https://doi.org/10.1186/s12915-020-00808-1>
3. Anamthawat-Jónsson K, Thórsson ÆT, Temsch EM, Greilhuber J (2010) Icelandic birch polyploids - the case of a perfect fit in genome size. *J Bot* 2010:347254. <https://doi.org/10.1155/2010/347254>
4. Barakat A, Carels N, Bernardi G (1997) The distribution of genes in the genomes of Gramineae. *Proc Natl Acad Sci USA* 94:6857–6861
5. Beaudry FEG, Rifkin JL, Barrett SCH, Wright SI (2020) Evolutionary genomics of plant gametophytic selection. *Plant Commun* 1:100115. <https://doi.org/10.1016/j.xplc.2020.100115>
6. Bennett MD (1972) Nuclear DNA content and minimum generation time in herbaceous plants. *Proc R Soc Lond B Biol Sci* 181:109–135. <https://doi.org/10.1098/rspb.1972.0042>
7. Bennett MD, Leitch IJ (2005) Genome size evolution in plants. In: Gregory T (ed) *The evolution of the genome*. Elsevier, San Diego, California, pp 89–162. <https://doi.org/10.1016/B978-012301463-4/50004-8>
8. Bennett MD, Smith JB, Smith RIL (1982) DNA amounts of angiosperms from the Antarctic and South Georgia. *Environ Exp Bot* 22:307–318. [https://doi.org/10.1016/0098-8472\(82\)90023-5](https://doi.org/10.1016/0098-8472(82)90023-5)

9. Billault-Penneteau B, Sandré A, Folgmann J, Parniske M, Pawlowski K (2019) *Dryas* as a model for studying the root symbioses of the Rosaceae. *Front Plant Sci* 10:661. <https://doi.org/10.3389/fpls.2019.00661>
10. Bjorkman AD, Myers-Smith IH, Elmendorf SC et al. (2018) Plant functional trait change across a warming tundra biome. *Nature* 562:57–62. <https://doi.org/10.1038/s41586-018-0563-7>
11. Chen JH, Huang Y, Brachi B, Yun QZ, Zhang W, Lu W, Li HN, Li WQ, Sun XD, Wang GY, He J, Zhou Z, Chen KY, Ji YH, Shi MM, Sun WG, Yang YP, Zhang RG, Abbott RJ, Sun H (2019) Genome-wide analysis of Cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nat Commun* 10:5230. <https://doi.org/10.1038/s41467-019-13128-y>
12. Cuba-Díaz M, Cerda G, Rivera C (2017) Genome size comparison in *Colobanthus quitensis* populations show differences in species ploidy. *Polar Biol* 40:1475–1480. <https://doi.org/10.1007/s00300-016-2058-z>
13. Deja-Muyllé A, Parizot B, Motte H, Beeckman T (2020) Exploiting natural variation in root system architecture via genome-wide association studies. *J Exp Bot*. 71:2379–2389. <https://doi.org/10.1093/jxb/eraa029>
14. Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ (1992) Nuclear DNA content variation within the Rosaceae. *Am J Bot* 79:1081–1086. <https://doi.org/10.1002/j.1537-2197.1992.tb13697.x>
15. Du F, Guan C, Jiao Y (2018) Molecular mechanisms of leaf morphogenesis. *Mol Plant* 11:1117–1134. <https://doi.org/10.1016/j.molp.2018.06.006>
16. Eidesen PB, Alsos IG, Brochmann C (2015) Comparative analyses of plastid and AFLP data suggest different colonization history and asymmetric hybridization between *Betula pubescens* and *B. nana*. *Mol Ecol* 24:3993–4009. <https://doi.org/10.1111/mec.13289>
17. Fajkus P, Peška V, Sitová Z, Fulnecková J, Dvořáčková M, Gogela R, Sýkorová E, Hapala J, Fajkus J (2016) Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)_n is synthesized by telomerase. *Plant J* 85:337–347. <https://doi.org/10.1111/tpj.13115>
18. Fulnecková J, Sevcíková T, Fajkus J, Lukesová A, Lukes M, Vlcek C, Lang BF, Kim E, Eliás M, Sykorová E (2013) A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol* 5:468–483. <https://doi.org/10.1093/gbe/evt019>
19. Goyal N, Bhatia G, Sharma S, Garewal N, Upadhyay A, Upadhyay SK, Singh K (2020) Genome-wide characterization revealed role of NBS-LRR genes during powdery mildew infection in *Vitis vinifera*. *Genomics* 112(1):312–322. <https://doi.org/10.1016/j.ygeno.2019.02.011>
20. Griesmann M, Chang Y, Liu X, et al. (2018) Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science*. 361:eaat1743. <https://doi.org/10.1126/science.aat1743>
21. Grime JP, Mowforth MA (1982) Variation in genome size—an ecological interpretation. *Nature* 299:151–153. <https://doi.org/10.1038/299151a0>
22. Guo L, Qiu J, Han Z, Ye Z, Chen C et al (2015) A host plant genome (*Zizania latifolia*) after a century-long endophyte infection. *Plant J* 83:600–609. <https://doi.org/10.1111/tpj.12912>

23. He L, Jia KH, Zhang RG, Wang Y, Shi TL, Li ZC, Zeng SW, Cai XJ, Wagner ND, Hörandl E, Muyle A, Yang K, Charlesworth D, Mao JF (2021) Chromosome-scale assembly of the genome of *Salix dunnii* reveals a male-heterogametic sex determination system on chromosome 7. *Mol Ecol Resour* 21:1966–1982. <https://doi.org/10.1111/1755-0998.13362>. Epub 2021 Mar 16
24. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, and Maruyama H (2014) Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research* 24:1384–1395. <https://doi.org/10.1101/gr.170720.113>
25. Ke L, Lei W, Yang W, Wang J, Gao J, Cheng J, Sun Y, Fan Z, Yu D (2020) Genome-wide identification of cold responsive transcription factors in *Brassica napus* L. *BMC Plant Biol* 20:62. <https://doi.org/10.1186/s12870-020-2253-5>
26. Knight CA, Molinari N, Petrov DA (2005) The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann Bot* 177–190. <https://doi.org/10.1093/aob/mci011>
27. Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. *Ann Bot* 82:45–55. <https://doi.org/10.1006/anbo.1998.0779>
28. Kim YJ, Laffly D, Kim S, Nilsen L, Chi J, Nam S, Lee YB, Jeong S, Mishra U, Lee YK, Jung JY (2022) Chronological changes in soil biogeochemical properties of the glacier foreland of Midtre Lovénbreen, Svalbard, attributed to soil-forming factors. *Geoderma* 415:115777. <https://doi.org/10.1016/j.geoderma.2022.115777>.
29. Lee YK (2020) *Arctic Plants of Svalbard: What We Learn from the Green in the Treeless White World*. Springer, Switzerland
30. Loureiro J, Castro M, Cerca de Oliveira J, Mota L, Torices R (2013) Genome size variation and polyploidy incidence in the alpine flora from Spain. *Annales del Jardín Botánico de Madrid* 70:39–47. <https://doi.org/10.3989/ajbm.2350>
31. Love A, Love D (1981) In *Chromosome number reports LXXIII*. *Taxon* 30:845–851. <https://www.jstor.org/stable/1220093>
32. Luka NO, Brandrud MK, Mandáková T, Lysak MA, Bjorå CS, Cires E, Nordal I, Brysting AK (2022) Genetic structure and evolution of diploid *Cochlearia* in Iceland. *Botanical Journal of the Linnean Society* boac018, <https://doi.org/10.1093/botlinnean/boac018>
33. Lysak MA, Koch MA, Beaulieu JM, Meister A, Leitch IJ (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol Biol Evol* 26:85–98. <https://doi.org/10.1093/molbev/msn223>
34. Mahmood T, Khalid S, Abdullah M, Ahmed Z, Shah MKN, Ghafoor A, Du X (2019) Insights into Drought Stress Signaling in Plants and the Molecular Genetic Basis of Cotton Drought Tolerance. *Cells* 9:105. <https://doi.org/10.3390/cells9010105>
35. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J and Clavijo BJ (2017) KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* 33:574–576. <https://doi.org/10.1093/bioinformatics/btw663>

36. Michael TP, Jackson S (2013) The first 50 plant genomes. *Plant Genome* 6:1–7. <https://doi.org/10.3835/plantgenome2013.03.0001in>
37. Nowak MD, Birkeland S, Mandáková T, et al. (2021) The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. *Mol Ecol Resour* 21:661–676. doi: 10.1111/1755-0998.13280
38. Oh M, Lee EJ (2021) Cushion plant *Silene acaulis* is a pioneer species at abandoned coal piles in the High Arctic, Svalbard. *J Ecology Environ* 45:1–9. <https://doi.org/10.1186/s41610-020-00177-4>
39. Pascual-Díaz JP, Serçe S, Hradecká I, Vanek M, Özdemir BS, Sultana N, Vural M, Vitales D, Garcia S (2020) Genome size constancy in Antarctic populations of *Colobanthus quitensis* and *Deschampsia antarctica*. *Polar Biol* 43:1407–1413. <https://doi.org/10.1007/s00300-020-02699-y>
40. Peer WA, Mamoudian M, Lahner B, Reeves RD, Murphy AS, Salt DE (2003) Identifying model metal hyperaccumulating plants: germplasm analysis of 20 Brassicaceae accessions from a wide geographical area. *New Phytol* 159:421–430. <https://doi.org/10.1046/j.1469-8137.2003.00822.x>
41. Pellicer J, and Leitch IJ (2019) The Plant DNA C-values database (release 7.1): an updated online repository of plant genome size data for comparative studies. *New Phytologist* 226:301–305. <https://doi.org/10.1111/nph.16261>
42. Peška V, and Garcia S (2020) Origin, Diversity, and Evolution of Telomere Sequences in Plants. *Frontiers in Plant Science* 11:117. <https://doi.org/10.3389/fpls.2020.00117>
43. Peška V, Fajkus P, Fojtová M, Dvořáčková M, Hapala J, Dvořáček V, Polanská P, Leitch AR, Sýkorová E, Fajkus J (2015) Characterisation of an unusual telomere motif (TTTTTTAGGG)_n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant J* 82:644–654. <https://doi.org/10.1111/tpj.12839>
44. Radoeva T, Vaddepalli P, Zhang Z, Weijers D (2019) Evolution, initiation, and diversity in early plant embryogenesis. *Dev Cell* 50:533–543. <https://doi.org/10.1016/j.devcel.2019.07.01>
45. Richards EJ, and Ausubel FM (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* 53:127–136. [https://doi.org/10.1016/0092-8674\(88\)90494-1](https://doi.org/10.1016/0092-8674(88)90494-1)
46. Siljak-Yakovlev, S., Lamy, F., Takvorian, N. et al. (2020) Genome size and chromosome number of ten plant species from Kerguelen Islands. *Polar Biol* 43:1985–1999. <https://doi.org/10.1007/s00300-020-02755-7>
47. Siroký J, Lysák MA, Dolezel J, Kejnovský E, Vyskot B (2001) Heterogeneity of rDNA distribution and genome size in *Silene* spp. *Chromosome Res* 9:387–93. <https://doi.org/10.1023/a:1016783501674>
48. Slotte T, Hazzouri KM, Agren JA et al (2013) The *Capsella rubella* genome and the genomic consequences of rapid mating system evolution. *Nat Genet* 45:831–835. <https://doi.org/10.1038/ng.2669>
49. Speed JDM, Chimal-Ballesteros JA, Martin MD, Barrio IC, Vuorinen KEM, Soininen EM (2021) Will borealization of Arctic tundra herbivore communities be driven by climate warming or vegetation change? *Glob Chang Biol* 27:6568–6577. <https://doi.org/10.1111/gcb.15910>

50. Sun Y, Shang L, Zhu QH, Fan L, Guo L (2022) Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* 27:391–401. <https://doi.org/10.1016/j.tplants.2021.10.006>
51. Sýkorová E, Fajkus J, Mezníková M, Lim KY, Neplechová K, Blattner FR, Chase MW and Leitch AR (2006) Minisatellite telomeres occur in the family Alliaceae but are lost in *Allium*. *Am J Bot* 93:814–823. <https://doi.org/10.3732/ajb.93.6.814>
52. Sýkorová E, Lim KY, Kunická Z, Chase MW, Bennett MD, Fajkus J and Leitch AR (2003) Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* 270:1893–1904. <https://doi.org/10.1098/rspb.2003.2446>
53. Těšitel J, Těšitelová T, Bernardova A, Drdova EJ, Lučanová M, Klimešová J (2014) Demographic population structure and fungal associations of plants colonizing High Arctic glacier forelands, Petuniabukta, Svalbard. *Polar Research* 33:20797. <https://doi.org/10.3402/polar.v33.20797>.
54. Tran TD, Cao HX, Jovtchev G, Neumann P, Novák P, Fojtová M, Vu GT, Macas J, Fajkus J, Schubert I et al (2015) Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *Plant J* 84:1087–1099. <https://doi.org/10.1111/tpj.13058>
55. van der Bilt WGM, Lane CS (2019) Lake sediments with Azorean tephra reveal ice-free conditions on coastal northwest Spitsbergen during the Last Glacial Maximum. *Sci Adv* 5:eaaw5980. <https://doi.org/10.1126/sciadv.aaw5980>
56. van der Kolk H-J, Heijmans MMPD, van Huissteden J, Pullens JWM, Berendse F (2016) Potential Arctic tundra vegetation shifts in response to changing temperature, precipitation and permafrost thaw. *Biogeosciences* 13:6229–6245. <https://doi.org/10.5194/bg-13-6229-2016>
57. Vinogradov AE (2003) Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet* 19:609–614. <https://doi.org/10.1016/j.tig.2003.09.010>
58. Wang N, Thomson M, Bodles WJ, Crawford RM, Hunt HV, Featherstone AW, Pellicer J, Buggs RJ (2013) Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol Ecol* 22:3098–111. <https://doi.org/10.1111/mec.12131>
59. Wei S, Yang Y, Yin T (2020) The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution. *Hortic Res* 7:45. <https://doi.org/10.1038/s41438-020-0268-6>
60. Weiss H, Scherthan H (2002) *Aloe* spp. - plants with vertebrate-like telomeric sequences. *Chromosome Res* 10:155–164. <https://doi.org/10.1023/A:1014905319557>
61. Zhang J, Yuan H, Li Y, Chen Y, Liu G, Ye M, Yu C, Lian B, Zhong F, Jiang Y, Xu J (2020) Genome sequencing and phylogenetic analysis of allotetraploid *Salix matsudana* Koidz. *Hortic Res* 7:201. <https://doi.org/10.1038/s41438-020-00424-8>

Figures



Figure 1

Thirteen Arctic plants analyzed in this paper. **a** *Eriophorum scheuchzeri* ssp. *arcticum*, **b** *Papaver dahlianum*, **c** *Silene acaulis*, **d** *Silene uralensis* ssp. *arctica*, **e** *Bistorta vivipara*, **f** *Oxyria digyna*, **g** *Saxifraga oppositifolia*, **h** *Cochlearia groenlandica*, **i** *Salix polaris* male, **j** *Salix polaris* female, **k** *Dryas octopetala*, **l** *Betula nana* ssp. *nana*, **m** *Cassiope tetragona*, **n** *Polemonium boreale*

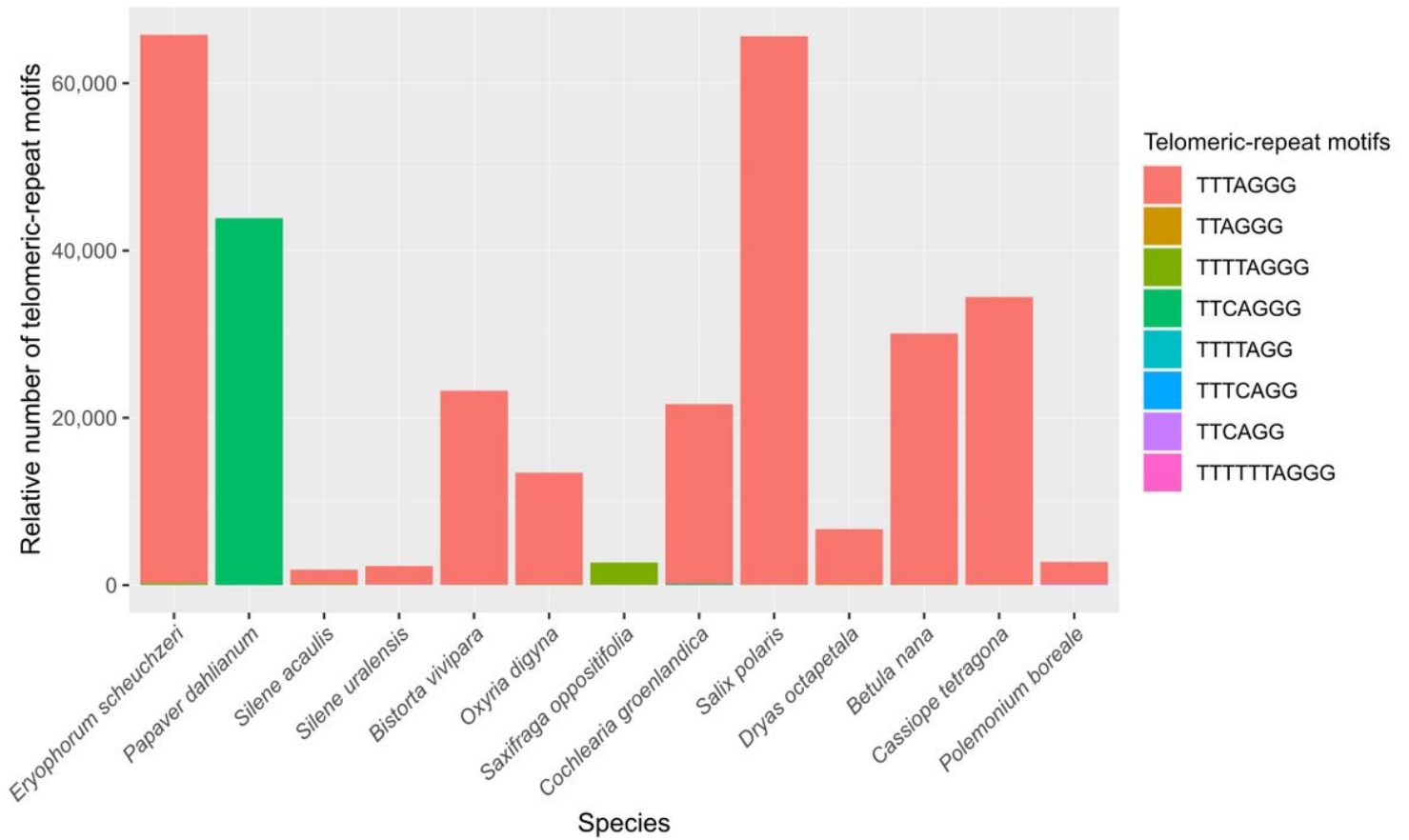


Figure 2

Relative number of telomeric-repeat motifs that were previously known or newly identified in this study (TTCAGGG). A total eight types of telomeric-repeat motifs were counted in raw DNA sequencing reads of the thirteen Svalbard plants. Different colors represent different telomeric-repeat motifs.