

## Purdue University Purdue e-Pubs

---

Department of Botany and Plant Pathology Faculty  
Publications

Department of Botany and Plant Pathology

---

6-12-2008

# The Highest-Copy Repeats are Methylated in the Small Genome of the Early Divergent Vascular Plant *Selaginella moellendorffii*

Agnes P. Chan  
[achan@jcvi.org](mailto:achan@jcvi.org)

Admasu Melake-Berhan  
[amelake@jcvi.org](mailto:amelake@jcvi.org)

Kimberly O'Brien  
[kobrien@som.umaryland.edu](mailto:kobrien@som.umaryland.edu)

Stephanie Buckley  
*University of Maryland - Baltimore*, [sbuckley@som.umaryland.edu](mailto:sbuckley@som.umaryland.edu)

Hui Quan  
[hui.quan@gmail.com](mailto:hui.quan@gmail.com)

*See next page for additional authors*

Follow this and additional works at: <http://docs.lib.purdue.edu/btnypubs>

---

### Recommended Citation

Chan, Agnes P.; Melake-Berhan, Admasu; O'Brien, Kimberly; Buckley, Stephanie; Quan, Hui; Chen, Dan; Lewis, Matthew; Banks, Jo Ann; and Rabinowicz, Pablo D., "The Highest-Copy Repeats are Methylated in the Small Genome of the Early Divergent Vascular Plant *Selaginella moellendorffii*" (2008). *Department of Botany and Plant Pathology Faculty Publications*. Paper 16.  
<http://dx.doi.org/10.1186/1471-2164-9-282>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.

---

**Authors**

Agnes P. Chan, Admasu Melake-Berhan, Kimberly O'Brien, Stephanie Buckley, Hui Quan, Dan Chen, Matthew Lewis, Jo Ann Banks, and Pablo D. Rabinowicz

Research article

Open Access

## The highest-copy repeats are methylated in the small genome of the early divergent vascular plant *Selaginella moellendorffii*

Agnes P Chan<sup>1</sup>, Admasu Melake-Berhan<sup>1</sup>, Kimberly O'Brien<sup>1,3</sup>,  
Stephanie Buckley<sup>1,4</sup>, Hui Quan<sup>1,5</sup>, Dan Chen<sup>1</sup>, Matthew Lewis<sup>1</sup>,  
Jo Ann Banks<sup>2</sup> and Pablo D Rabinowicz\*<sup>1,3,6</sup>

Address: <sup>1</sup>J. Craig Venter Institute (JCVI), Rockville, MD 20850, USA, <sup>2</sup>Department of Botany and Plant Pathology, Lilly Hall, Purdue University, West Lafayette, IN 47907, USA, <sup>3</sup>Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA, <sup>4</sup>Program in Oncology, University of Maryland School of Medicine, Baltimore, MD 21201, USA, <sup>5</sup>Applied Biosystems, 2130 Woodward St., Austin, TX 78744, USA and <sup>6</sup>Department of Biochemistry and Molecular Biology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

Email: Agnes P Chan - [achan@jvci.org](mailto:achan@jvci.org); Admasu Melake-Berhan - [amelake@jvci.org](mailto:amelake@jvci.org); Kimberly O'Brien - [kobrien@som.umaryland.edu](mailto:kobrien@som.umaryland.edu); Stephanie Buckley - [sbuckley@som.umaryland.edu](mailto:sbuckley@som.umaryland.edu); Hui Quan - [hui.quan@gmail.com](mailto:hui.quan@gmail.com); Dan Chen - [danchen@jvci.org](mailto:danchen@jvci.org); Matthew Lewis - [MLewis@jvci.org](mailto:MLewis@jvci.org); Jo Ann Banks - [banksj@purdue.edu](mailto:banksj@purdue.edu); Pablo D Rabinowicz\* - [prabinowicz@som.umaryland.edu](mailto:prabinowicz@som.umaryland.edu)

\* Corresponding author

Published: 12 June 2008

Received: 14 February 2008

BMC Genomics 2008, 9:282 doi:10.1186/1471-2164-9-282

Accepted: 12 June 2008

This article is available from: <http://www.biomedcentral.com/1471-2164/9/282>

© 2008 Chan et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The lycophyte *Selaginella moellendorffii* is a vascular plant that diverged from the fern/seed plant lineage at least 400 million years ago. Although genomic information for *S. moellendorffii* is starting to be produced, little is known about basic aspects of its molecular biology. In order to provide the first glimpse to the epigenetic landscape of this early divergent vascular plant, we used the methylation filtration technique. Methylation filtration genomic libraries select unmethylated DNA clones due to the presence of the methylation-dependent restriction endonuclease McrBC in the bacterial host.

**Results:** We conducted a characterization of the DNA methylation patterns of the *S. moellendorffii* genome by sequencing a set of *S. moellendorffii* shotgun genomic clones, along with a set of methylation filtered clones. Chloroplast DNA, which is typically unmethylated, was enriched in the filtered library relative to the shotgun library, showing that there is DNA methylation in the extremely small *S. moellendorffii* genome. The filtered library also showed enrichment in expressed and gene-like sequences, while the highest-copy repeats were largely under-represented in this library. These results show that genes and repeats are differentially methylated in the *S. moellendorffii* genome, as occurs in other plants studied.

**Conclusion:** Our results shed light on the genome methylation pattern in a member of a relatively unexplored plant lineage. The DNA methylation data reported here will help understanding the involvement of this epigenetic mark in fundamental biological processes, as well as the evolutionary aspects of epigenetics in land plants.

## Background

DNA methylation has been found throughout the plant kingdom, typically in cytosines, forming part of symmetric (CpNpG and CpG) and asymmetric (CpNpN) sites [1,2]. The proportion of methylated cytosine in plants is variable, ranging from 6% in *Arabidopsis* [3] to 25% in maize [4]. DNA methylation has been associated with the inactivation of transposons and silencing of genes [5-10], and it has also been proposed that the function of DNA methylation is to decrease transcriptional "noise" [11].

In plants, most DNA methylation is found in repetitive elements, while genes and other low copy sequences are generally hypomethylated [12-17].

Because of the large size of many plant genomes, particularly those of important crops [18], gene-enriched sequencing strategies have been designed as an alternative to whole genome sequencing in an attempt to capture the so-called gene-space of such genomes. One of these gene-enrichment techniques, called methylation filtration (MF), takes advantage of the difference in methylation between plant genes and repeats [19]. MF exploits the methylation-dependent restriction endonuclease McrBC (modified cytosine restriction) from *E. coli* [20,21]. This enzyme digests DNA in sequences that contain two sites, each one consisting of a purine and a cytosine methylated in carbon 5, separated by 40–3000 bp [22]. Therefore, using an *mcrBC*<sup>+</sup> *E. coli* strain as a host to construct a genomic shotgun library, heavily methylated repetitive DNA is efficiently counter-selected, while hypomethylated low copy (*i.e.* genic) sequences are substantially over-represented. MF was first tested in maize, where it yielded a 6-fold enrichment for genes relative to a whole genome shotgun (WGS) library used as a control [19]. Subsequently, MF was applied at large scale in maize [23,24] and in sorghum [25], showing that approximately 95% of the genes in each genome were tagged (A. Chan *et al.*, unpublished) and that most genes and regulatory elements are unmethylated in these two species. These results led to the suggestion that a combination of gene-enrichment and traditional genome sequencing techniques could be combined to efficiently sequence large plant genomes [26]. Further analyses of the large-scale MF data in maize and sorghum also provided insights into the biology of transposable element methylation and activity [23-25]. Pilot MF studies of several monocot, dicot, and non-angiosperm plants (such as pine, fern, and moss) were also conducted [27]. These analyses determined that MF enriches for genes in all plants tested, although to different levels, and that it can be an effective approach to selectively clone and sequence genes from some large plant genomes, where the majority of the DNA is composed of methylated repetitive elements.

In this study we performed a MF analysis of the lycophyte *Selaginella moellendorffii* (family Selaginellaceae), representing a clade not included in previous MF studies. The lycophyte clade diverged from the fern/seed-plant lineage about 400 million years ago [28].

The *S. moellendorffii* sporophyte is diploid and consists of dichotomously branching shoot and root systems. The shoot frequently terminates in arrested buds or bulbils that dehisce and allow clonal propagation. The reproductive structures are the strobili, which form toward the tip of the shoot, each one with either one micro- or megasporangium that produce micro- or megaspores, which in turn germinate and divide mitotically to form either the male or female gametophytes, respectively. The gametophyte produces either motile sperm or egg-forming archegonia. After fertilization of the egg, the new sporophyte remains dependent upon the female gametophyte for a short period of time. *S. moellendorffii* is an excellent model system to study some developmental processes, such as sporogenesis and gametophyte development, which are difficult to study in angiosperms because their spores and gametophytes are dependent upon and surrounded by sporophytic tissues. Seedless plants provide an excellent opportunity to study the epigenetics of these processes, but little is known about DNA methylation and other epigenetic marks in early vascular plants, except for the presence of heterochromatic bands identified by cytological staining [29]. Ferns have been used in attempts to address the methylation of the haploid and diploid generations [30] but their genomes are usually large and only specific sequences were analyzed. The extremely small genome of *S. moellendorffii* (90–130 Mbp; [31]) and its available 8× coverage, high-quality draft genome assembly generated by the Joint Genome Institute of the U.S. Department of Energy (JGI-DOE), will facilitate the study of *S. moellendorffii*'s epigenome and its involvement in the alternation of generations. Due to its small genome size, several transposon families, which are common targets of epigenetic modifications, may be low copy in *S. moellendorffii* and their sequence and epigenetics can be studied without the complications of high copy numbers, allowing the unequivocal identification of individual transposon loci.

Sequences from this study have been deposited in NCBI GenBank under the accession numbers [ET218553–ET221769].

## Results and Discussion

### Sequence data and chloroplast content

We constructed MF and WGS libraries from *S. moellendorffii* and produced 1,621 and 1,598 high-quality paired sequence reads, respectively, each set representing approximately 1% of the genome. We did not expect a substantial difference in the proportion of gene-like sequences in

the MF library relative to the WGS library in the small genome of *S. moellendorffii* because previous studies showed that in the ~400 Mbp genomes of rice and *Ceratodon purpureus* (the smallest genomes in which MF has been tested) the gene enrichment factors (GEF, calculated as the ratio between the MF and WGS proportion of non-repetitive, gene-like sequences) were 1.9 and 2.5, respectively [27]. As the chloroplast genome is typically non-methylated, we prepared total *S. moellendorffii* DNA to construct the MF and WGS libraries in order to retain the chloroplast DNA in both libraries and used it to verify that methylated sequences exist in *S. moellendorffii* and are counter-selected by MF. High-stringency alignments against the *Selaginella uncinata* chloroplast genome [32] identified 14.9% and 7.8% chloroplast DNA sequences in the MF and WGS datasets, respectively (Figure 1A), thus demonstrating that the *S. moellendorffii* genome is methylated and that MF selects for non-methylated sequences as expected.

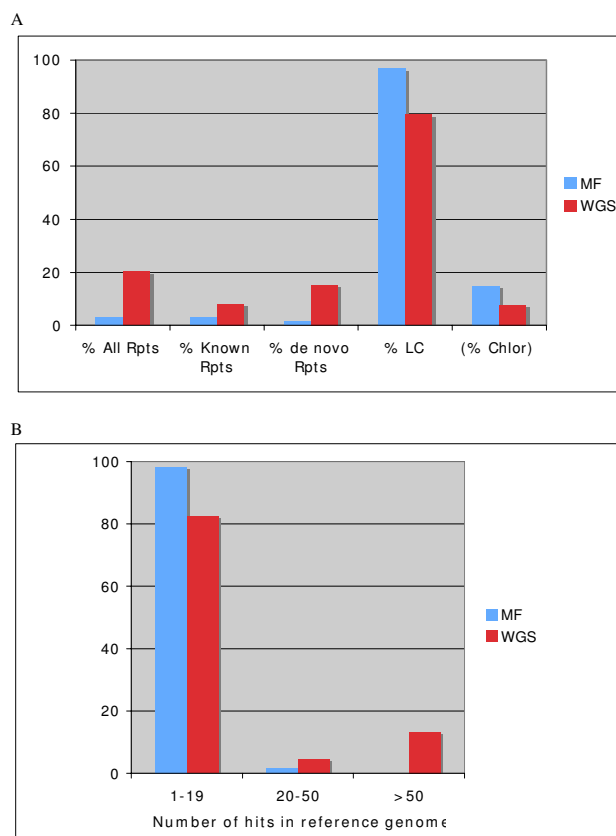
The chloroplast reads identified in this way were not analyzed further and, therefore, a total of 1,379 MF and 1,471 WGS non-chloroplast reads were used in the following analyses.

Overall, the C+G content is slightly higher in MF than in WGS data (47.9% vs. 46.2, respectively), probably due to the higher C+G content of gene sequences, which are predominant in the MF set (Table 1).

Approximately 13% of the sequences could not be aligned to the reference genome assembly at the stringency used in this study. This discrepancy may be due to the exclusion of sequence assemblies shorter than 1 kbp from the reference genome sequence.

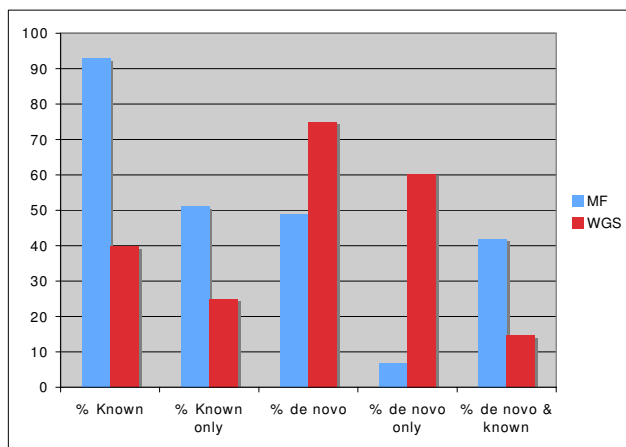
### Repetitive Sequences

In order to identify repetitive sequences we used nucleotide and amino acid databases of plant repetitive sequences [27]. Consistent with the notion that repetitive elements are methylated in plants, only 2.9% of the MF sequences had a match in either of the repeat databases, while matches in the WGS set reached 8.1% (Figure 1A). As sequences from early vascular plants are underrepresented in available databases, it is possible that many *S. moellendorffii* repetitive elements will not be identified by comparison with previously annotated plant repeats. To better estimate the repeat content of each dataset, we attempted to identify repeats *de novo* by aligning our MF and WGS reads to the draft genome assembly produced by JGI-DOE. Any sequence that had 20 or more high-stringency matches in the reference genome was considered repetitive (Figure 1B and Additional files 1 and 2). A 10-fold reduction in the percentage of these *de novo* repeats was observed in the MF vs. the WGS data set, showing that



**Figure 1**  
**Proportion of repetitive and low-copy sequences in the MF and WGS libraries.** **A:** Proportion of all low-copy sequences (transcribed, gene-like and anonymous) are shown together (LC). Proportion of all repeats (All Rpts) and their break down into known (Known Rpts) and *de novo* repeats (de novo Rpts) are shown separately. The percentage of chloroplast (Chlor) sequences is calculated relative to the total number of sequences in each library. All the other percentages are calculated relative to the total of non-chloroplast reads in each library. **B:** Percentages of MF and WGS reads matching the reference genome, classified by the number of hits. Any read showing 20 or more hits in the reference genome is considered a *de novo* identified repeat.

the highest-copy elements (largest number of hits in the reference genome) are methylated in *S. moellendorffii* (Figure 1A). Taken together, the known and *de novo* repeats account for 3.1% and 20.4% of the MF and WGS reads, respectively (Figure 1A). Among the repetitive MF reads, most were identified as known repeats by database searches, and nearly half were also identified *de novo* (Figure 2), although no MF repeats shows more than 42 copies in the *S. moellendorffii* reference genome sequence, and many are ribosomal RNA sequences (see Additional file 1). Interestingly, all MF sequences matching known transposable elements are low-copy in *S. moellendorffii* (*i.e.*



**Figure 2**  
**Proportions of repetitive sequences.** Repeats are classified as matching the repeat databases (Known), identified *de novo*, matching the repeat databases but not identified *de novo* (Known only), identified *de novo* (de novo) but with no match in the repeat databases (de novo only), or that were identified *de novo* and also have a match in the repeat databases (de novo & known). Percentages are calculated relative to the total number of repetitive sequences.

have less than 20 hits in the genome; Additional file 1). In contrast, over 60% of the WGS repeats were identified *de novo*, and do not have a database match, while among those that have similarity to known repeats, nearly 1/3 are high-copy (Figure 2). Furthermore, known WGS repeats show a maximum of 234 hits in the genome, but 1/3 of the WGS repeats have more than 234 copies, the highest having over 500 (see Additional file 2). The prevalence of low-copy repeats detected by MF suggests that low-copy transposons are unmethylated and, therefore, potentially active [6,33,34]. The observed substantial number of WGS unknown high-copy elements highlights the diversity of transposable elements throughout the plant kingdom.

Sequence composition analysis of the repetitive sequences showed that MF repeats are richer in C+G than those in the WGS set, probably due to the abundance of conserved, non-methylated ribosomal RNA sequences among the MF repeats (Table 1).

#### Gene sequences, expressed sequences and gene enrichment

Using BLASTX, the MF and WGS non-repetitive sequences were compared to a partially curated, non-identical amino

acid sequence database (NIAA) maintained at JCVI, containing most proteins available from GenBank. The percentages of BLASTX matches against this database were 35% and 22% in MF and WGS sequences, respectively (Figure 3), representing a 1.6-fold enrichment in MF relative to WGS sequences, indicating that protein-encoding genes are frequently hypomethylated. Therefore, MF enriches for genes even in the minute genome of *S. moellendorffii*. We also performed high stringency alignments of our sequences to the *S. moellendorffii* assembled ESTs [35], which showed that MF enriches for transcribed sequences to a similar level as it does for protein sequences, suggesting hypomethylation of expressed sequences. Combining the protein and transcribed sequences alignments, 49% of the MF and 31% of the WGS sequences matched either database, while sequences with no database match represented 48% of each dataset (Figure 3).

In order to estimate the level of gene enrichment achieved with MF in *S. moellendorffii* in comparison with previous studies done in other plants [27], all sequences that were not identified as repeats or chloroplast were compared to the same curated database of known gene sequences used in those studies. In this way, 12.8% of the WGS sequences and 21.5% of the MF sequences had a match in the known gene database, resulting in a GEF of 1.7 (Figure 3).

We attempted to confirm that the exclusion of high-copy sequences in the MF library was due to DNA methylation using an independent assay. To do this, we digested *Selaginella* genomic DNA with the methylation-sensitive restriction enzyme *HpaII*, whose restriction target site (CCGG) includes the frequently methylated CpG motif. We then selected 5 of the highest-copy sequences in the WGS library with no match in the databases (*de novo* highest-copy repeats), as well as 5 low-copy MF sequences showing similarity to ESTs and known genes. We designed polymerase chain reaction (PCR) primer pairs so that the expected amplification product would include at least one *HpaII* site, and carried out PCR reactions with each primer pair using *HpaII*-digested and undigested DNA as template. The results in Figure 4A show that the amount of amplification product obtained with *HpaII*-digested template was substantially reduced relative to the undigested control in the 5 low-copy MF sequences, indicating that these sequences are not methylated in the genome, allowing digestion by *HpaII* and cleavage of the target template sequence. On the other hand, no difference in amplifica-

**Table 1: C+G content in different sequence classes**

	%C+G total	%C+G in repeats	% C+G in low-copy DNA	% C+G in genes and EST hits
MF	47.9	50.9	47.8	49.7
WGS	46.2	47.5	45.8	49.6

tion was observed between the *HpaII*-digested and undigested high-copy templates. As each of these PCR products correspond to a mixture of sequences from multiple repeated loci, it is possible that some copies of these repeats show variation with respect to the presence of *HpaII* recognition sites, due to polymorphisms relative to the sequence used for PCR primer design. Thus, the absence of digestion may reflect a lack of *HpaII* sites or the presence of methylated *HpaII* sites. To test if *HpaII* sites were present in the high-copy WGS PCR products amplified from *HpaII*-digested DNA, we treated these 5 PCR products with *HpaII*. We observed digestion in all PCR products, indicating that *HpaII* sites were present and thus, methylated (figure 4B). Nevertheless, we also observed the presence a low proportion of undigested PCR product in addition to the expected digestion products, as well as additional bands, suggesting that multiple polymorphic copies of each repeat were amplified in all cases.

## Conclusion

Our results show that even in the small genome of *S. moellendorffii*, MF sequences display much lower repeat content than WGS sequences, and that each of the identified MF repeats has less than 42 copies in the genome. If the MF repeat sequences are aligned to the reference genome at higher stringency, the number of hits for each repeat decreases, indicating that polymorphisms can be found inside families of repetitive elements (data not shown). Therefore, by sequencing the hypomethylated fraction of the *S. moellendorffii* genome using MF it would be possible to identify which copies of these repetitive elements are methylated. MF of the *S. moellendorffii* genome can be used to obtain information on gene methylation as well, as it has been shown in *Arabidopsis*, where a fraction of the genes do contain cytosine methylation (although at a lower level than repeats and pseudogenes) and this methylation is predominant in particular regions of the genes [14-17]. In consequence, a genome-wide DNA methylation profile can be generated by comprehensive MF sequencing of this genome. Furthermore, combining MF with ultra-high throughput next-generation sequencing techniques will facilitate this kind of analyses using the sequenced genome as a reference. As the variety of *S. moellendorffii* whose genome was sequenced by JGI-DOE has two distinct haplotypes that differ in nucleotide sequence by ~2-5%, (J. Banks, unpublished), it will be possible to determine if there is haplotype-specific DNA methylation using MF sequencing. Genome-wide epigenetic studies of early-diverging land plants will provide the foundation to broaden our understanding of the evolution of epigenetic regulation of developmental processes in plant biology.

## Methods

Total DNA was purified using DNeasy kits (Qiagen, CA) from green tissues of *S. moellendorffii* plants kept in growth chamber. The DNA was mechanically sheared using a Hydroshear device (Genomic Solutions, MI) and fragments ranging from 3 to 4 kb were eluted from an agarose gel after electrophoresis, end-repaired, and ligated into a cloning vector. DNA ligation reactions were transformed into *E. coli* DH5 $\alpha$  (*mcrBC*<sup>+</sup>) to construct the MF library. The WGS library was constructed by introducing the same ligation reaction into *E. coli* GC10 (*mcrBC*<sup>-</sup>). Recombinant clones were sequenced using Big Dye Terminator chemistry and ABI 3730xl sequencers (Applied Biosystems, CA), and vector and low-quality sequences were electronically trimmed.

Chloroplast sequences were identified by BLASTN alignment to the *S. uncinata* chloroplast genome (GenBank accession [AB197035](#)) at high stringency (E value smaller than 10<sup>-56</sup>). The chloroplast sequences were excluded from any further sequence analyses. Protein sequence alignments against the NIAA database were done using BLAT. Alignments with at least 70% similarity and 40 amino acids long were recorded as matches.

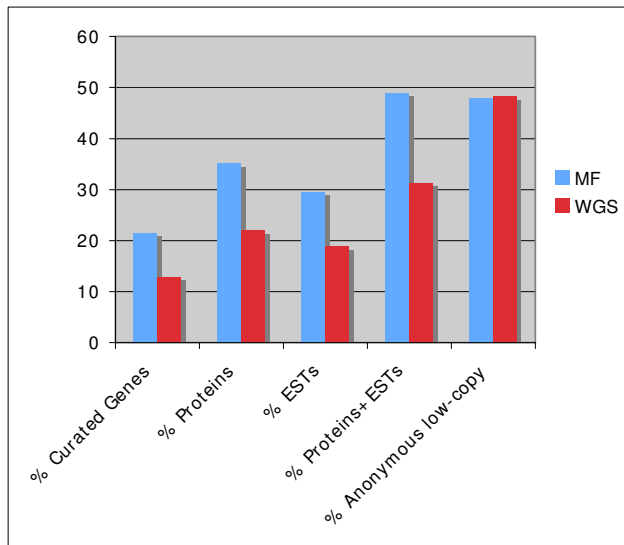
Alignments to assembled EST sequences were done using BLASTN at high stringency. Matches showing an E value smaller than 10<sup>-56</sup> were recorded.

*De novo* repeats were identified by aligning MF and WGS reads to the JGI-DOE *S. moellendorffii* genome assembly using BLASTN and matches covering 50% of the read with 95% identity were recorded.

Alignments to the curated database of known genes were done as previously reported [27], using BLASTX and recording matches with an E value better than 10<sup>-7</sup>.

Known repeats were identified using a nucleotide database and a protein database of known repetitive elements described earlier [27]. These databases do not contain simple sequence repeats. Repetitive element proteins were identified using the protein database of repeats. The same criteria were used to identify known genes, while repetitive nucleotide sequences were identified using BLASTN with an E value smaller than 10<sup>-10</sup>.

DNA digestion with *HpaII* was performed following manufacturer recommendations. PCR assays were carried out using 50 ng of *HpaII*-digested or undigested genomic DNA as template, and denaturing 3 minutes at 94°C followed by 25 amplification cycles using the following program: 30 seconds at 94°C, 30 seconds at 59°C, and 60 seconds at 72°C. Elongation was allowed for 10 minutes



**Figure 3**  
**Gene and transcribed sequence content.** Percentages of matches to the database of curated genes, matches to the NIAA protein database, matches to the *S. moellendorffii* EST assemblies (ESTs), the combination of matches to NIAA protein and EST assemblies databases, and the anonymous low copy-sequences are shown.

at 72°C after amplification. Target and primer sequences are shown in Additional file 3.

### Authors' contributions

APC and HQ performed the sequence analysis; AM-B, DC and SB worked on DNA preparations and library constructions; KO'B performed PCR assays; ML was in charge of sequencing; JAB contributed to the intellectual design of the study; PDR conceived the study, participated in the analysis, and drafted the manuscript. All authors have read and approved the final manuscript.

### Additional material

#### Additional file 1

*Complete MF analysis results.* An excel file listing BLAST hits of all MF sequences against all databases used, classified as "repeats", "low-copy sequences", and "chloroplast sequences".

Click here for file

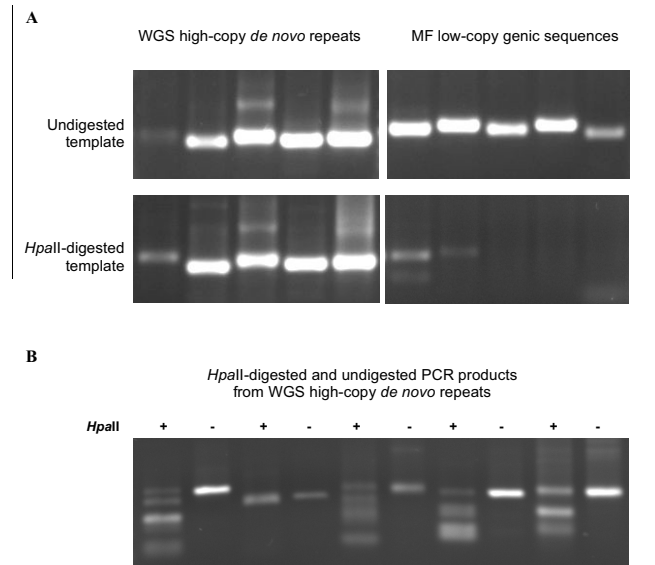
[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-282-S1.xls>]

#### Additional file 2

*Complete WGS analysis results.* An excel file listing BLAST hits of all WGS sequences against all databases used, classified as "repeats", "low-copy sequences", and "chloroplast sequences".

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-9-282-S2.xls>]



**Figure 4**  
**HpaII-digestion and PCR amplification of low- and high-copy sequences.** **A:** PCR products ran on agarose gels are shown. Five highly repeated WGS sequences (left panels) and 5 MF sequences (right panels) were amplified from HpaII-digested (bottom panels) or undigested (top panels) *S. moellendorffii* genomic DNA. **B:** HpaII digestion of the 5 PCR products obtained with HpaII-digested genomic DNA from panel A. From left to right, digested and undigested PCR products (in the same order as panel A, bottom left).

### Acknowledgements

This work was funded by The Institute for Genomic Research (TIGR, now called J. Craig Venter Institute or JCVI, Rockville, MD).

### References

- Gruenbaum Y, Naveh-Many T, Cedar H, Razin A: **Sequence specificity of methylation in higher plant DNA.** *Nature* 1981, **292**:860-862.
- Meyer P, Niedenhof I, ten Lohuis M: **Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*.** *Embo J* 1994, **13**:2084-2088.
- Kakutani T, Munakata K, Richards EJ, Hirochika H: **Meiotically and mitotically stable inheritance of DNA hypomethylation induced by *ddm1* mutation of *Arabidopsis thaliana*.** *Genetics* 1999, **151**:831-838.
- Papa CM, Springer NM, Muszynski MG, Meeley R, Kaeppler SM: **Maize chromomethylase *Zea methyltransferase2* is required for CpNpG methylation.** *Plant Cell* 2001, **13**:1919-1928.
- Bird A: **DNA methylation patterns and epigenetic memory.** *Genes Dev* 2002, **16**:6-21.
- Chandler VL, Walbot V: **DNA modification of a maize transposable element correlates with loss of activity.** *Proc Natl Acad Sci USA* 1986, **83**:1767-1771.
- Colot V, Rossignol JL: **Eukaryotic DNA methylation as an evolutionary device.** *Bioessays* 1999, **21**:402-411.
- Martienssen RA, Colot V: **DNA methylation and epigenetic inheritance in plants and filamentous fungi.** *Science* 2001, **293**:1070-1074.
- Flavell RB: **Inactivation of gene expression in plants as a consequence of specific sequence duplication.** *Proc Natl Acad Sci USA* 1994, **91**:3490-3496.



10. Martienssen R: **Transposons, DNA methylation and gene control.** *Trends Genet* 1998, **14**:263-264.
11. Bird AP: **Gene number, noise reduction and biological complexity.** *Trends Genet* 1995, **11**:94-100.
12. Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P: **Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA.** *Genome* 1994, **37**:565-576.
13. Rabinowicz PD, Palmer LE, May BP, Hemann MT, Lowe SW, McCombie WR, Martienssen RA: **Genes and transposons are differentially methylated in plants, but not in mammals.** *Genome Res* 2003, **13**:2658-2664.
14. Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R: **Role of transposable elements in heterochromatin and epigenetic control.** *Nature* 2004, **430**:471-476.
15. Vaughn MW, Tanurdic M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA: **Epigenetic Natural Variation in Arabidopsis thaliana.** *PLoS Biol* 2007, **5**:e174.
16. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S: **Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription.** *Nat Genet* 2007, **39**:61-69.
17. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, Ecker JR: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis.** *Cell* 2006, **126**:1189-1201.
18. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species.** *Plant Mol Biol Rep* 1991, **9**:208-218.
19. Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Parnell LD, Stein L, McCombie WR, Martienssen RA: **Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome.** *Nat Genet* 1999, **23**:305-308.
20. Dila D, Sutherland E, Moran L, Slatko B, Raleigh EA: **Genetic and sequence organization of the mcrBC locus of Escherichia coli K-12.** *J Bacteriol* 1990, **172**:4888-4900.
21. Raleigh EA, Wilson G: **Escherichia coli K-12 restricts DNA containing 5-methylcytosine.** *Proc Natl Acad Sci USA* 1986, **83**:9070-9074.
22. Sutherland E, Coe L, Raleigh EA: **McrBC: a multisubunit GTP-dependent restriction endonuclease.** *J Mol Biol* 1992, **225**:327-348.
23. Palmer LE, Rabinowicz PD, O'Shaughnessy AL, Balija VS, Nascimento LU, Dike S, de la Bastide M, Martienssen RA, McCombie WR: **Maize genome sequencing by methylation filtration.** *Science* 2003, **302**:2115-2117.
24. Whitelaw CA, Barbazuk WB, Perteua G, Chan AP, Cheung F, Lee Y, Zheng L, van Heeringen S, Karamycheva S, Bennetzen JL, SanMiguel P, Lakey N, Bedell J, Yuan Y, Budiman MA, Resnick A, Van Aken S, Utterback T, Riedmuller S, Williams M, Feldblyum T, Schubert K, Beachy R, Fraser CM, Quackenbush J: **Enrichment of gene-coding sequences in maize by genome filtration.** *Science* 2003, **302**:2118-2120.
25. Bedell JA, Budiman MA, Nunberg A, Citek RW, Robbins D, Jones J, Flick E, Rholving T, Fries J, Bradford K, McMenamy J, Smith M, Holuman H, Roe BA, Wiley G, Korf IF, Rabinowicz PD, Lakey N, McCombie WR, Jeddelloh JA, Martienssen RA: **Sorghum genome sequencing by methylation filtration.** *PLoS Biol* 2005, **3**:e13.
26. Rabinowicz PD, Bennetzen JL: **The maize genome as a model for efficient sequence analysis of large plant genomes.** *Curr Opin Plant Biol* 2006, **9**:149-156.
27. Rabinowicz PD, Citek R, Budiman MA, Nunberg A, Bedell JA, Lakey N, O'Shaughnessy AL, Nascimento LU, McCombie WR, Martienssen RA: **Differential methylation of genes and repeats in land plants.** *Genome Res* 2005, **15**:1431-1440.
28. Kenrick P, Crane PR: **The origin and early evolution of plants on land.** *Nature* 1997, **389**:33-39.
29. Marcon AB, Barros IC, Guerra M: **Variation in chromosome numbers, CMA bands and 45S rDNA sites in species of Selaginella (Pteridophyta).** *Ann Bot (Lond)* 2005, **95**:271-276.
30. McGrath JM, Pichersky E: **Methylation of somatic and sperm DNA in the homosporous fern Ceratopteris richardii.** *Plant Mol Biol* 1997, **35**:1023-1027.
31. Wang W, Tanurdic M, Luo M, Sisneros N, Kim HR, Weng JK, Kudrna D, Mueller C, Arumuganathan K, Carlson J, Chapple C, de Pamphilis C, Mandoli D, Tomkins J, Wing RA, Banks JA: **Construction of a bacterial artificial chromosome library from the spikemoss Selaginella moellendorffii: a new resource for plant comparative genomics.** *BMC Plant Biol* 2005, **5**:10.
32. Tsuji S, Ueda K, Nishiyama T, Hasebe M, Yoshikawa S, Konagaya A, Nishiuchi T, Yamaguchi K: **The chloroplast genome from a lycophyte (microphylophyte), Selaginella uncinata, has a unique inversion, transpositions and many gene losses.** *Journal of plant research* 2007, **120**:281-290.
33. Singer T, Yordan C, Martienssen RA: **Robertson's Mutator transposons in A. thaliana are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDMI).** *Genes Dev* 2001, **15**:591-602.
34. Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T: **Mobilization of transposons by a mutation abolishing full DNA methylation in Arabidopsis.** *Nature* 2001, **411**:212-214.
35. **Purdue Selaginella Genomics** [<http://selaginella.genomics.purdue.edu/data.html>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

