



## ORIGINAL RESEARCH

## Paleo-polyploidization in Lycophytes



Jinpeng Wang<sup>1,2,3</sup>, Jigao Yu<sup>1,3</sup>, Pengchuan Sun<sup>1</sup>, Chao Li<sup>1</sup>, Xiaoming Song<sup>1,2</sup>, Tianyu Lei<sup>1</sup>, Yuxian Li<sup>1</sup>, Jiaqing Yuan<sup>1</sup>, Sangrong Sun<sup>1</sup>, Hongling Ding<sup>1</sup>, Xueqian Duan<sup>1</sup>, Shaoqi Shen<sup>1</sup>, Yanshuang Shen<sup>1</sup>, Jing Li<sup>1</sup>, Fanbo Meng<sup>1</sup>, Yangqin Xie<sup>1</sup>, Jianyu Wang<sup>1</sup>, Yue Hou<sup>1</sup>, Jin Zhang<sup>1</sup>, Xianchun Zhang<sup>3</sup>, Xiu-Qing Li<sup>4</sup>, Andrew H. Paterson<sup>5,6,7,8</sup>, Xiyin Wang<sup>1,2,\*</sup>

<sup>1</sup> Center for Computational Biology and Genomics, and School of Life Sciences, North China University of Science and Technology, Tangshan 063200, China

<sup>2</sup> National Key Laboratory for North China Crop Improvement and Regulation, Agriculture University of Hebei, Baoding 071001, China

<sup>3</sup> State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Science, Beijing 100093, China

<sup>4</sup> Fredericton Research and Development Centre, Agriculture and Agri-Food Canada, Fredericton, New Brunswick, E3B 4Z7, Canada

<sup>5</sup> Plant Genome Mapping Laboratory, University of Athens, Athens, GA 30602, USA

<sup>6</sup> Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

<sup>7</sup> Department of Crop and Soil Science, University of Georgia, Athens, GA 30602, USA

<sup>8</sup> Department of Genetics, University of Georgia, Athens, GA 30602, USA

Received 20 July 2018; revised 13 January 2019; accepted 17 April 2019

Available online 4 November 2020

Handled by Xiangfeng Wang

## KEYWORDS

Vascular plant;  
Lycophytes;  
Genome;  
Polyploidy;  
Evolution

**Abstract** Lycophytes and seed plants constitute the typical vascular plants. Lycophytes have been thought to have no paleo-polyploidization although the event is known to be critical for the fast expansion of seed plants. Here, genomic analyses including the homologous gene dot plot analysis detected multiple paleo-polyploidization events, with one occurring approximately 13–15 million years ago (MYA) and another about 125–142 MYA, during the evolution of the genome of *Selaginella moellendorffii*, a model lycophyte. In addition, comparative analysis of reconstructed ancestral genomes of lycophytes and angiosperms suggested that lycophytes were affected by more paleo-polyploidization events than seed plants. Results from the present genomic analyses indicate that paleo-polyploidization has contributed to the successful establishment of both lineages—lycophytes and seed plants—of vascular plants.

\* Corresponding author.

E-mail: [wangxiyin@vip.sina.com](mailto:wangxiyin@vip.sina.com) (Wang X).

Peer review under responsibility of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

<https://doi.org/10.1016/j.gpb.2020.10.002>

1672-0229 © 2020 The Authors. Published by Elsevier B.V. and Science Press on behalf of Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

Extant vascular plants can be divided into two major types, the euphyllophytes (ferns and seed plants) and the lycophytes, which diverged as early as 410 million years ago (MYA) [1]. *Selaginella moellendorffii*, as a model lycophyte, has a dominant and complex sporophyte generation and vascular tissues with lignified cell types. The deciphered genome of *S. moellendorffii* is approximately 212.6 Mb, including two haplotypes of approximately 106 Mb [1]. Recently, the even smaller genome of *S. lepidophylla*, approximately 109 Mb, was deciphered [2].

Recursive polyploidizations or whole-genome duplications have been proposed as a key evolutionary driving force of seed plants, contributing to their divergence and rapid expansion [3,4]. Different from seed plants, genomic analyses with various approaches including a discontinuous MegaBLAST did not find evidence of paleo-polyploidization in both *S. moellendorffii* and *S. lepidophylla* genomes [1,5].

Seed plant genomes can be quite complex, at least partially due to recursive polyploidization and repetitive sequence accumulation. It is often difficult to perform a comprehensive analysis to understand their genome structure and evolution. In several instances, ancestral polyploidization was omitted from genome analysis, resulting in problematic interpretation of the structure, evolution, and/or functional innovation of whole genomes and key gene families [6–9]. Recently, we developed a pipeline [10], including the generation of dot plot chart of homologous genes [11], to facilitate the analysis of complex genomes, especially those of plants. Analysis of the cucurbit genomes using this pipeline revealed an overlooked paleo-tetraploidization event, occurring ~100 MYA, in the common ancestor of Cucurbitaceae plants, which possibly contributed to the establishment and fast divergence of the whole family [10].

Here, we used our pipeline [10], including the homologous gene dot plot analysis [11], to analyze both *Selaginella* genomes and other related plant genomes [12]. We detected multiple polyploidization events in lycophytes.

## Results

### Gene colinearity

Using a gene-colinearity-based approach implemented in ColinearScan, with maximal gap between neighboring colinear genes of 50 genes (consistent with prior studies of many plants), we inferred 302 syntenic blocks in the *S. moellendorffii* genome (Tables S1 and S2). These blocks involved 2632 colinear gene pairs, and covered 87.01% (19,391/22,285) of all genes and 88.24% (19,391/21,975) of assembled genome sequences (Figure 1A). Use of a smaller gap size (25 genes) does not qualitatively change this result (65.92% genome coverage). The assembly had 192 scaffolds (of at least five genes), of which at least 75.52% (145/192) were covered by syntenic blocks with five or more colinear genes (more information about blocks can be found in Table S3).

The syntenic blocks in the *S. moellendorffii* genome were mapped onto the chromosomes and produced coverage as deep as 10. Approximately 19.24% of genome regions were covered with a depth of 4 or more, and 17.55%, 25.36%, 27.06%, and 10.79% of regions were covered with a depth

of 3, 2, 1, and 0, respectively (Figure S1 and Table S4). Correspondence of colinear genes between multiple duplicated regions could often be found, showing probable recursive gene duplications (Figure 1B).

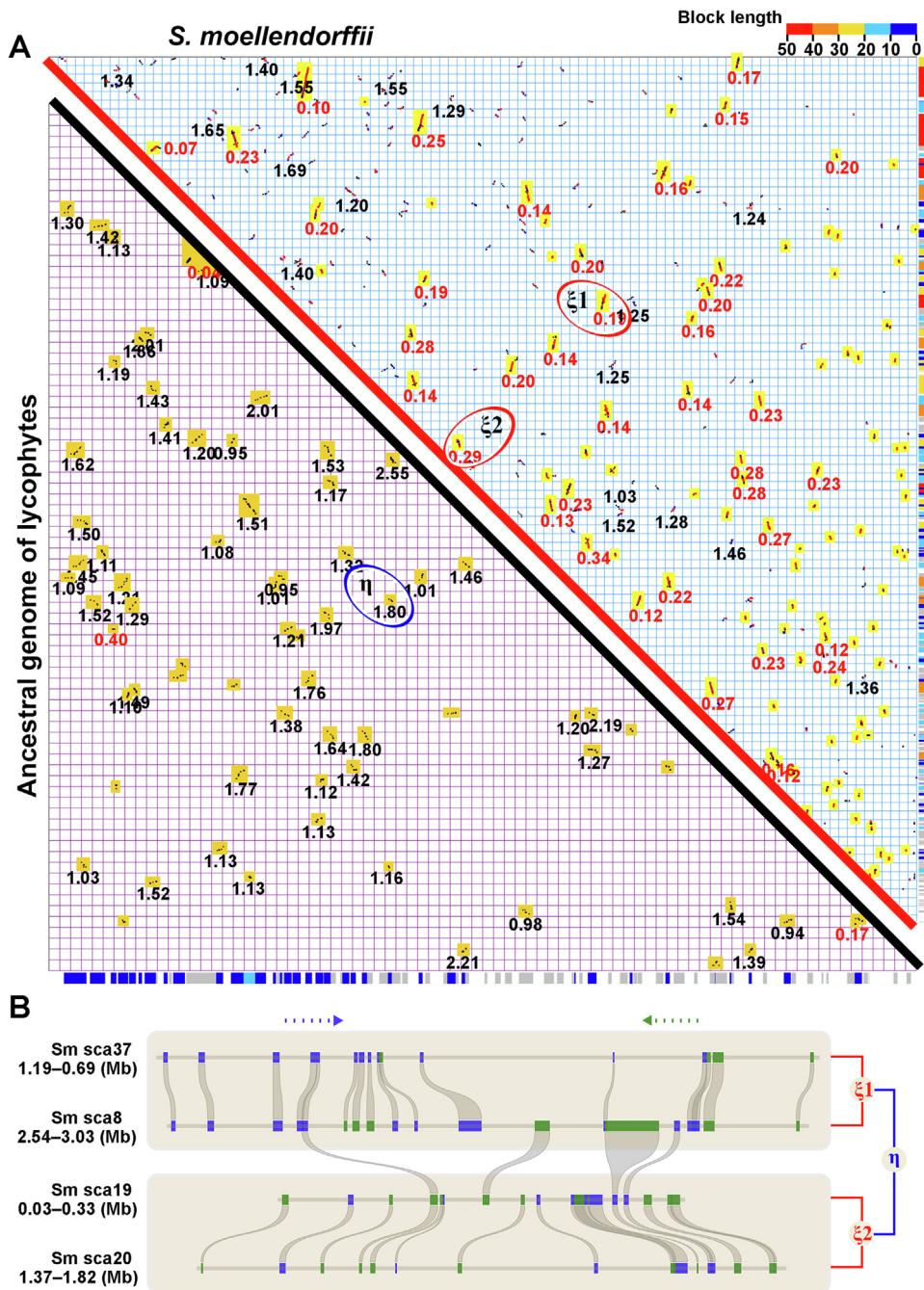
### Evidence of recursive polyploidization events

Sequence divergence of colinear genes in the *S. moellendorffii* genome clearly showed a non-random distribution consistent with the occurrence of polyploidization. Both synonymous nucleotide substitutions per synonymous site ( $K_s$ ) and nucleotide diversity for the fourfold synonymous third-codon transversion position (4Dtv) displayed clear bimodal distributions (Figure 2, Figures S2 and S3). For  $K_s$ , the two peaks were located at 0.12 and 1.2, respectively (Figure 2). Based on the  $K_s$  distribution, we divided the colinear genes into two groups (Figure 1A). Supposing  $6.0\text{--}7.0 \times 10^{-9}$  synonymous substitutions per site per year, borrowed from angiosperms [3,13,14], the two peaks were repositioned at 0.12 and 1.15, respectively, suggesting large-scale genomic duplication events to have occurred ~13–15 MYA (named  $\xi$  or Xi) and ~125–142 MYA (named  $\eta$  or Eta). The use of separate distributions rather than the merged distribution allowed elimination of interactive effects during statistical analysis. The younger group covered 76.65% (16,844/21,975 genes) of the genome and 70.31% (135/192) of total scaffolds (related to all chromosomes). This result indicated that the event  $\xi$  was a polyploidization, with a  $K_s$  peak clearly distinct from any residual heterozygosity of *S. moellendorffii* (DNA similarity ~98.5% [1], corresponding to  $K_s$  ~0.056).

By merging the colinear gene blocks generated from the polyploidization event 13–15 MYA, we inferred that the ancestral genome before this event consisted of at least 120 syntenic blocks, which included 936 colinear genes, and covered 64.79% (8434/13,017) of the ancestral genome. We also discovered the existence of a more ancient whole-genome doubling event (the 125–142  $\eta$  event), because self-comparison of these 120 blocks showed further colinearity that is extremely improbable to occur by chance.

### Comparison between *S. lepidophylla* and angiosperms

Furthermore, we checked whether the two ancient polyploidization events detected in the *S. moellendorffii* genome was shared with *S. lepidophylla* and seed plants. Though two *Selaginella* genomes shared obvious gene colinearity (Tables S1 and S2), an analysis of the *S. lepidophylla* genome did not find evidence of polyploidization. The two lycophytes have 1121 colinear blocks with length  $\geq 4$  colinear genes, including 11,087 colinear genes in total. In *S. lepidophylla*, 110 blocks including 538 colinear genes were found, much fewer than in *S. moellendorffii*. By characterizing  $K_s$  between colinear genes, we found  $K_s$  peaks of putative orthologous genes at 1.35, 1.57, and 1.60 for *S. moellendorffii* vs. *S. lepidophylla*, *Amborella trichopoda* (a model angiosperm), and *Vitis vinifera* (grape), respectively (Figure 2). This result suggested that the two inferred polyploidization events in the *S. moellendorffii* lineage were absent in *S. lepidophylla* and in angiosperms (Figure 3).



**Figure 1** Homologous gene dot plot of *Selaginella moellendorffii* and alignment of genomic regions with colinear genes

**A.** The upper-right diagonal shows dot plots between scaffolds of extant genomes. Both X- and Y-axes are represented by sequentially linked scaffolds. A duplicated gene pair inferred by ColinearScan between X and Y results in a dot. Neighboring dot plots result in blocks showing duplicated segments in the genome. Red, blue, and gray dots indicate best-, second-best, and other matches between genes. The median  $K_s$  value of gene pairs in a block is displayed in red when  $K_s < 0.5$  and in black when  $1 < K_s < 1.7$ . Blocks involved in the recent tetraploidization, highlighted in yellow boxes, are mapped to the Y-axis to the right, with the longest blocks in a neighboring region displayed with a color scheme and shown along the X-axis to the right. The lower-left shows dot plots in the reconstructed ancestral genome before the recent tetraploidization. Dot plots are produced between genes in the ancestral genome. Blocks are highlighted, and  $K_s$  values of blocks are displayed. The blocks are mapped to the X-axis, with the same color scheme defined above to show the longest blocks in a neighboring region. Examples of duplicated blocks produced by  $\eta$  and  $\xi$  are circled, to display gene colinearity in subfigure. **B.** The four regions involved are from *S. moellendorffii* scaffold20, scaffold19, scaffold8, and scaffold37, inferred to have been produced by two recursive polyploidizations,  $\eta$  and  $\xi$ . Genes are shown with rectangles, and the different colors show transcription directions. Colinear genes are linked with curved lines.

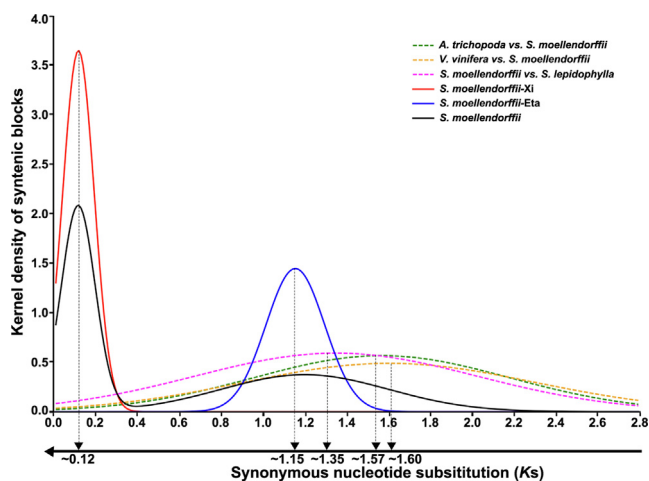


Figure 2 Distribution of  $K_s$  between colinear genes

### Existence of likely more paleo-polyploidization events

An appreciable fraction of *S. moellendorffii* genomic regions were covered at a depth  $> 4$  by colinear blocks, suggesting more ancient polyploidization. Therefore, we performed a deeper search of gene synteny between *S. moellendorffii* and the

two referenced angiosperm genomes. The *Amborella* genome, assembled into scaffolds, has a relatively simple structure and avoided a genome doubling after its split from other angiosperms, despite the existence of evidence of a shared ancient polyploidization [15] (Figure 3). The grape genome, well assembled into pseudochromosomes, was also used to reconstruct the ancestral genome before the major-eudicot-common hexaploidy. We inferred syntenic blocks between *S. moellendorffii* and each of the referenced genomes and mapped the blocks onto each of the genomes involved (Figures S4 and S5). Totals of 14.18% and 9.73% of the *S. moellendorffii* and *Amborella* genomes, respectively, were covered by colinear blocks to depth of 4 or more (Table S5). Likewise, 21.34% and 13.86% of the *S. moellendorffii* and grape genomes, respectively, were covered by colinear blocks to depth of 4 or more (Table S6). These findings suggested additional more ancient polyploidization events during the evolution of vascular plants.

We then explored the coverage depth of homologous regions in the reconstructed pre- $\xi$  ancestral genome of lycophytes (ALG, 11,509 genes) and ancestral genome of angiosperms (AAG, 1686 genes) inferred on the basis of the grape genome. The ALG involved genes from the 34 largest *S. moellendorffii* scaffolds. At a gap size of 50 genes using ColinearScan, significant colinear blocks (*i.e.*, with at least four colinear genes) resulted in homologous coverage depth as deep as 7 and

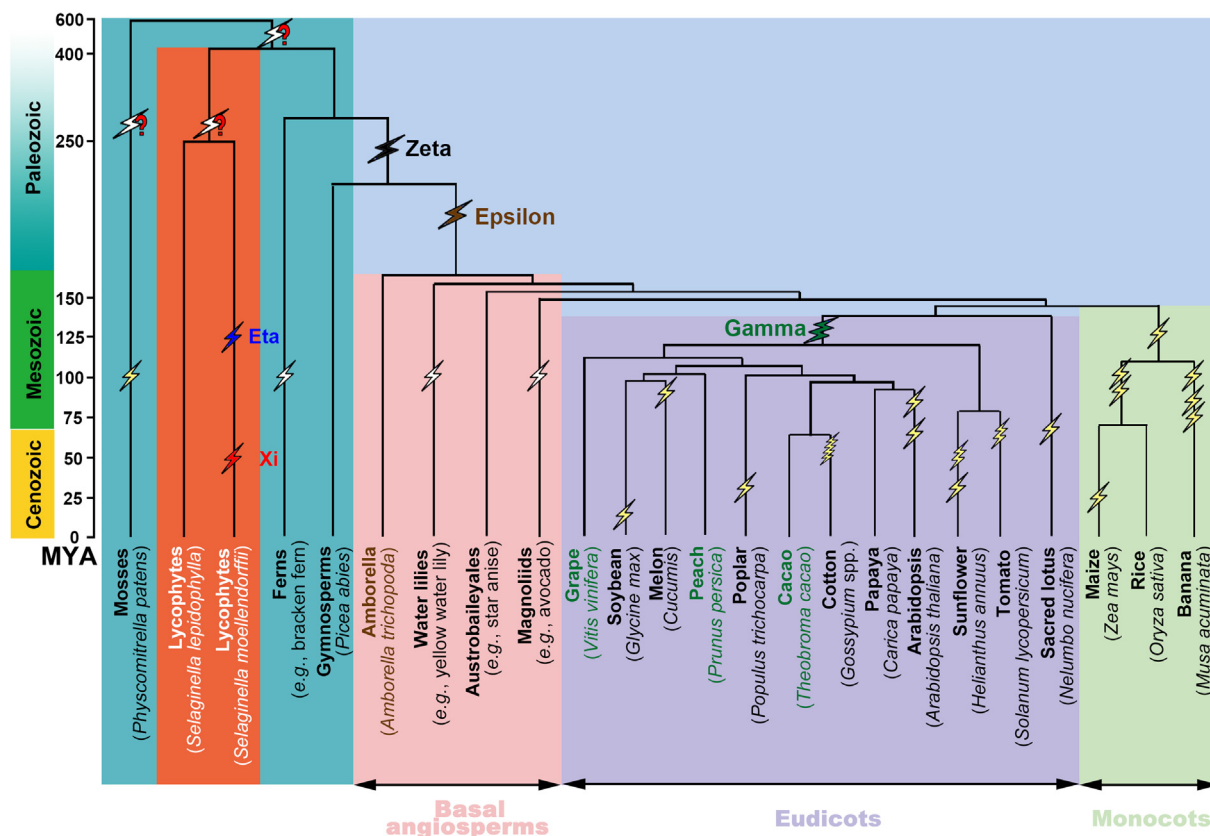


Figure 3 Plant phylogeny and inferred ancient polyploidies

Flash marks are used to show polyploidizations, with two zigzags showing tetraploidy, three showing hexaploidy, five showing decaploidy, and a question mark indicating an undetermined ploidy level. The events  $\eta$  and  $\xi$ , inferred in the present work, are shown in blue and red, respectively. More ancient events that were inferred to have likely occurred are also shown with question marks.

18 in the AAG and ALG, respectively (Figure 4). Indeed, 78.28% and 7.49% of the AAG and ALG, respectively, were covered to depth of 4 or more (Table S7). To be careful, we tried different maximal gap sizes in each compared lineage (Table S7) and obtained similar conclusions (Figure 3).

In that *S. lepidophylla* was not affected by polyploidization after its divergence with *S. moellendorffii*, it could provide insight into polyploidization(s) older than their split from angiosperms. Therefore, we managed to use *S. lepidophylla* genomic DNA, though not as well assembled as that of *S. moellendorffii*, to infer gene colinearity. The *S. lepidophylla* contigs were sorted as to their best matched *S. moellendorffii* scaffolds. We found appreciable amount of colinearity within the genome itself, and between *S. lepidophylla* and AAG (Figure S6 and Table S1). Regarding coverage depth, 68.55% and 20.74% of the AAG and *S. lepidophylla*, respectively, were covered to depth of 4 or more (Table S8). At least 40% of the inferred colinear blocks terminate at the ends of assembled scaffolds, presumably broken by incomplete assembly rather than by lack of colinearity (Figure S7).

Since lycophytes and seed plants are thought to have diverged 410 MYA [1], and the two *Selaginella* plants (*S. moellendorffii* and *S. lepidophylla*) diverged approximately 250 MYA [2], the present findings of at least two paleo-polyploidization events in less than 150 MYA suggest that there may have been additional paleo-polyploidizations in each lineage of lycophytes and seed plants, separately. Moreover, greater coverage mapped onto the AAG means that lycophytes might have experienced more polyploidizations than seed plants.

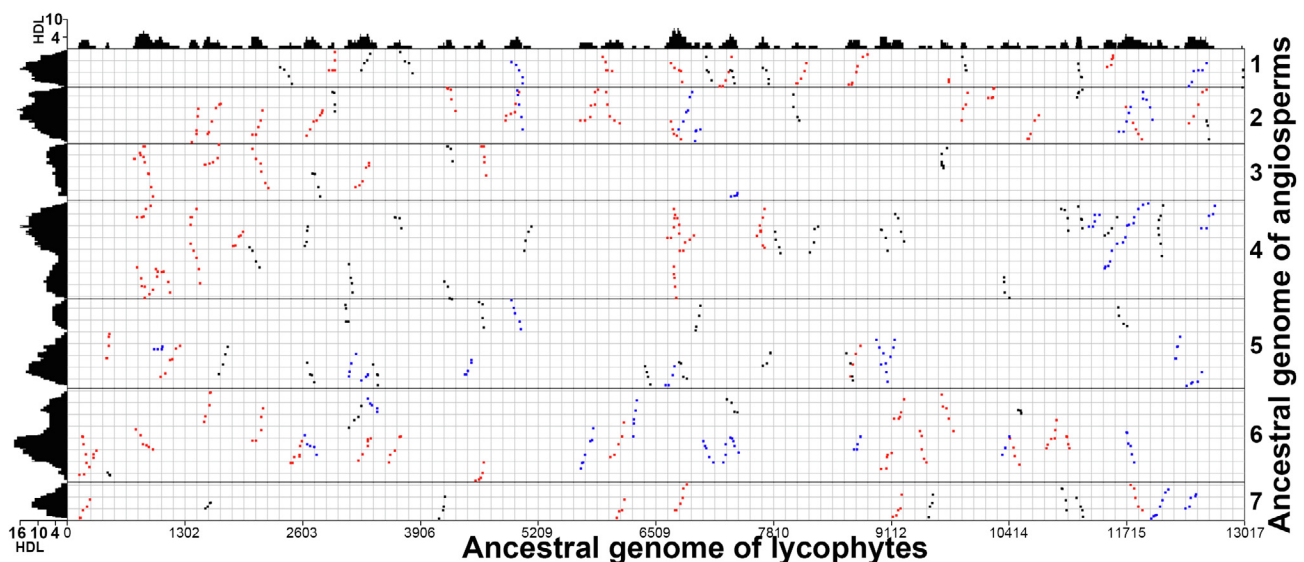
We have to note that a more precise determination of the numbers and dates of these ancient polyploidization events is beyond the resolution of *Ks*. Other sequence divergence analysis approaches using likely more taxa of basal vascular plants will be required.

## Discussion

Polyploidization or whole-genome duplication was inferred to have contributed to the early divergence of seed plants according to phylogenetic tree reconstruction [4]. By dissecting features of *Ks* distribution with transcriptome data, horsetail, a basal land plant, was found to be an ancient polyploid [16]. A large-scale cytogenetic and phylogenetic analysis indicated that ferns might have been more frequently affected by polyploidization [17]. Together with many other publications proposing ancient polyploidization [18–21], these works each shed light on the distant past of plant evolution. However, although the first *Selaginella* genome has been available for years, genome-scale evidence of ancient whole-genome duplication in the lycophytes remains elusive. Here, using the sophisticated pipeline that we developed for deciphering complex genomes, we provided clear and solid evidence that two polyploidization events have contributed to the evolution of the lycophyte *S. moellendorffii* after its split from *S. lepidophylla*, and several older polyploidization events likely shared also by *Selaginella* plants have occurred in the early days of vascular plants.

Polyploidizations were proposed to be the likely answer to the mystery about the abrupt origin and fast divergence of seed and flowering plants on Earth raised by Charles Darwin [20]. The present finding that lycophytes were also recursively affected by polyploidizations may benefit further evaluation how these events contributed to the expansion of different lineages of vascular plants.

The availability of whole-genome sequences provides evidence of ancient polyploidy. However, it is technically challenging to deconvolute these ancient large-scale events [6–9]. The MegaBLAST analysis of the *Selaginella* genome overlooked these large-scale evolutionary events [1], possibly due to the complicated nature of plant genomes and/or draw-



**Figure 4** Inferred colinear genes between ancestral genome of lycophytes (ALG) and ancestral genome of angiosperms (AAG)

Inferred ancestral regions represented by genes are arranged along X- and Y-axes, respectively. Statistically significant colinear blocks are displayed and mapped onto each axis to produce homologous coverage depth. Scale bars are displayed to the homologous depth level (HDL) in the genome. Blocks with median *Ks* < 1.8 are colored red, blocks with *Ks* ≥ 1.8 are colored blue, and others are colored gray.

backs in the methodology adopted [10]. To decipher the complex genomes, genome dot plot often makes a good illustration of gene synteny/colinearity. Divergence of inferred colinear genes reveals non-random patterns that distinguish subsets with different ages. The exploration of colinear genes can provide insight into more ancient events, as proposed, and can be successfully used to decipher recursive polyploidizations during the evolution of eudicots based on the small genome of *Arabidopsis* [11]. Methodology that lacks inference of gene colinearity or use of homologous gene dot plots, e.g., based largely on inferences of *Ks* distribution or phylogenetic tree topology, could easily overlook evidence of paleo-polyploidization. Even *Ks* distributions that give the appearance of polyploidization could be caused by genes duplicated by other events, such as waves of retrotransposition. The combination of temporal (*Ks*) and positional (colinearity) evidence can provide for falsification of alternative hypotheses, to uncover ancient, especially recursive polyploidizations in extant complex genomes. Future research may help to understand how these polyploidization events have contributed to biological and genetic innovations during evolution.

## Conclusion

A major lineage of land plants, lycophytes, was thought to have avoided paleo-polyploidization. Here, using a sophisticated comparative genomics approach, we inferred that the genome of the first sequenced lycophyte *S. moellendorffii* was affected by at least two rounds of paleo-polyploidization events; and more analyses with another lycophyte and seed plants suggested more large-scale genome duplications during the evolution of land plants.

## Materials and methods

### Materials

Genome sequences and annotations were downloaded from the Joint Genome Institute (JGI, <https://phytozome.jgi.doe.gov/pz/portal.html>); *S. moellendorffii*, *S. lepidophylla*, *A. trichopoda*, and *V. vinifera* L.).

### Gene colinearity

The pipeline to decipher complex genomes was followed as described previously [10]. Homologous gene dot plots within a genome or between genomes were produced by using MCSCAN toolkits [22], of which the corresponding author directly contributed to the development. Colinear genes were inferred by using ColinearScan [23], a statistically well-supported software in pairwise homology. ColinearScan implements a modified dynamic programming algorithm and recursively searches the remaining longest path with colinear gene pairs as nodes. The maximal gap between genes involved in colinearity along a chromosome was evaluated computationally, and each colinear block inferred was then statistically assessed of their significance. Here, the maximal gap was set as 50 genes apart, and repetitive genes with > 30 copies were

removed from the present analysis, as adopted in many previous studies [6,24–27]. Evolutionary divergence between homologous genes was estimated as described previously [24,26].

### Analysis of *Ks*

The synonymous nucleotide substitutions (*Ks*) between duplicated genes in colinearity resulted from each of considered polyploidization and divergence events were analyzed. *Ks* values were estimated using Nei-Gojobori approach by implementing the program codeml in the “phylogenetic analysis by maximum likelihood” software PAML [28,29]. Gene CDS alignment was performed using ClustalW with default parameters [30].

### Construction of ancestral genomes

Construction of ancestral genomes was also based on our previously detailed methods [10,11]. For clarification, here, we reconstructed the ancestral genome of lycophytes (ALG) before the polyploidization  $\xi$ , and the ancestral genome of angiosperms (AAG). The ALG was reconstructed by utilization of one copy of the colinear genes produced by the recent polyploidy, in that the colinear genes probably preserved their ancestral gene locations (Figure S8A). In a similar approach, the AAG genome was reconstructed by inferring ancestral genes using grape colinear genes produced by the major-eudicot-common hexaploidy (previously named  $\Upsilon$ ) (Figure S8B). We did not infer AAG using a grape-*Amborella* or grape-rice comparison due to incomplete assembly of the *Amborella* genome and severe genomic fractionation after multiple polyploidies in monocot lineages [31].

## CRedit author statement

**Jinpeng Wang:** Writing - original draft, Supervision, Methodology, Software. **Jigao Yu:** Software, Formal analysis, Data curation, Visualization. **Pengchuan Sun:** Software, Formal analysis, Visualization. **Chao Li:** Software, Formal analysis. **Xiaoming Song:** Software, Formal analysis, Data curation. **Tianyu Lei:** Software, Formal analysis. **Yuxian Li:** Software, Formal analysis, Data curation. **Jiaqing Yuan:** Software, Formal analysis, Data curation. **Sangrong Sun:** Formal analysis, Data curation. **Hongling Ding:** Formal analysis, Data curation. **Xueqian Duan:** Formal analysis, Data curation. **Shaoqi Shen:** Formal analysis, Data curation. **Yanshuang Shen:** Formal analysis, Data curation. **Jing Li:** Formal analysis, Data curation. **Fanbo Meng:** Investigation, Data curation. **Yangqin Xie:** Investigation, Data curation. **Jianyu Wang:** Investigation, Data curation. **Yue Hou:** Data curation. **Jin Zhang:** Data curation. **Xianchun Zhang:** Validation. **Xiu-Qing Li:** Writing - review & editing. **Andrew H. Paterson:** Writing - review & editing. **Xiyin Wang:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision. All authors read and approved the final manuscript.

## Competing interests

The authors have declared no competing interests.

## Acknowledgments

We appreciate financial support from the Ministry of Science and Technology of the People's Republic of China (Grant No. 2016YFD0101001), the China National Science Foundation (Grant Nos. 31371282 to XW, 31510333 to JW, and 31661143009 to XW), the Natural Science Foundation of Hebei Province (Grant No. C2015209069 to JW), and Tangshan Key Laboratory Project to XW. We thank the helpful discussion with researchers at iGeno Co. Ltd, China.

## Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.gpb.2020.10.002>.

## ORCID

0000-0002-2253-7752 (Jinpeng Wang)  
 0000-0001-9156-2532 (Jigao Yu)  
 0000-0002-8999-4894 (Pengchuan Sun)  
 0000-0002-9607-6899 (Chao Li)  
 0000-0003-0084-3668 (Xiaoming Song)  
 0000-0001-9151-213X (Tianyu Lei)  
 0000-0002-2075-6724 (Yuxian Li)  
 0000-0002-0192-9351 (Jiaqing Yuan)  
 0000-0003-2971-5213 (Sangrong Sun)  
 0000-0001-5345-2594 (Hongling Ding)  
 0000-0002-1045-3818 (Xueqian Duan)  
 0000-0002-5769-7430 (Shaoqi Shen)  
 0000-0003-3430-9367 (Yanshuang Shen)  
 0000-0002-7772-7676 (Jing Li)  
 0000-0002-6042-090X (Fanbo Meng)  
 0000-0002-8166-7071 (Yangqin Xie)  
 0000-0002-0442-9205 (Jianyu Wang)  
 0000-0001-6786-2236 (Yue Hou)  
 0000-0002-3954-5038 (Jin Zhang)  
 0000-0003-3425-1011 (Xianchun Zhang)  
 0000-0001-7740-0932 (Xiu-Qing Li)  
 0000-0003-2159-0487 (Andrew H. Paterson)  
 0000-0003-3454-0374 (Xiyin Wang)

## References

- [1] Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, et al. The *Selaginella* genome identifies genetic changes associated with the evolution of vascular plants. *Science* 2011;332:960–3.
- [2] VanBuren R, Wai CM, Ou S, Pardo J, Bryant D, Jiang N, et al. Extreme haplotype variation in the desiccation-tolerant clubmoss *Selaginella lepidophylla*. *Nat Commun* 2018;9:13.
- [3] Paterson AH, Bowers JE, Chapman BA. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. *Proc Natl Acad Sci U S A* 2004;101:9903–8.
- [4] Jiao Y, Wickett NJ, Ayyampalayam S, Chandrabali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* 2011;473:97–100.
- [5] Baniaga AE, Arrigo N, Barker MS. The small nuclear genomes of *Selaginella* are associated with a low rate of genome size evolution. *Genome Biol Evol* 2016;8:1516–25.
- [6] Wang X, Guo H, Wang J, Lei T, Liu T, Wang Z, et al. Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionation. *New Phytol* 2016;209:1252.
- [7] Li Z, Baniaga AE, Sessa EB, Scascitelli M, Graham SW, Rieseberg LH, et al. Early genome duplications in conifers and other seed plants. *Sci Adv* 2015;1:e1501084.
- [8] Wang X, Shi X, Hao B, Ge S, Luo J. Duplication and DNA segmental loss in the rice genome: implications for diploidization. *New Phytol* 2005;165:937–46.
- [9] Paterson AH, Bowers JE, Van dPY, Vandepoele K. Ancient duplication of cereal genomes. *New Phytol* 2005;165: 658–61.
- [10] Wang J, Sun P, Li Y, Liu Y, Yang N, Yu J, et al. An overlooked paleo-tetraploidization in Cucurbitaceae. *Mol Biol Evol* 2018;35:16–26.
- [11] Bowers JE, Chapman BA, Rong J, Paterson AH. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 2003;422:433–8.
- [12] Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007;449:463–7.
- [13] Moniz dSM, Drouin G. Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol* 1996;13:1198–212.
- [14] Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science* 2000;290:1151–5.
- [15] Albert VA, Barbazuk WB, dePamphilis CW, Der JP, Leebens-Mack J, Ma H, et al. The *Amborella* genome and the evolution of flowering plants. *Science* 2013;342:1241089.
- [16] Vanneste K, Sterck L, Myburg AA, Van dPY, Mizrahi E. Horsetails are ancient polyploids: evidence from *Equisetum giganteum*. *Plant Cell* 2015;27:1567–78.
- [17] Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A* 2009;106:13875.
- [18] Soltis DE, Buggs RJ, Barbazuk WB, Schnable PS, Soltis PS. On the origins of species: does evolution repeat itself in polyploid populations of independent origin?. *Cold Spring Harb Symp Quant Biol* 2009;74:215–23.
- [19] Chalhoub B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, et al. Plant genetics. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 2014;345:950–3.
- [20] Van de Peer Y. A mystery unveiled. *Genome Biol* 2011;12:1–2.
- [21] D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, Garsmeur O, et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* 2012;488:213–7.
- [22] Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 2012;40:e49.
- [23] Wang X, Shi X, Zhe L, Zhu Q, Lei K, Wen T, et al. Statistical inference of chromosomal homology based on gene colinearity and applications to *Arabidopsis* and rice. *BMC Bioinformatics* 2006;7:447.
- [24] Wang J, Sun P, Li Y, Liu Y, Yu J, Ma X, et al. Hierarchically aligning 10 legume genomes establishes a family-level genomics platform. *Plant Physiol* 2017;174:284.
- [25] Wang J, Yu J, Sun P, Li Y, Xia R, Liu Y, et al. Comparative genomics analysis of rice and pineapple contributes to understand the chromosome number reduction and genomic changes in grasses. *Front Genet* 2016;7:174.
- [26] Wang X, Wang J, Jin D, Guo H, Lee TH, Liu T, et al. Genome alignment spanning major poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol Plant* 2015;8:885–98.

- [27] Liu Y, Wang J, Ge W, Wang Z, Li Y, Yang N, et al. Two highly similar poplar paleo-subgenomes suggest an autotetraploid ancestor of salicaceae plants. *Front Plant Sci* 2017;8:571.
- [28] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 2007;24:1586–91.
- [29] Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418–26.
- [30] Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* 2002, Chapter 2: Unit 2.3.
- [31] Tang H, Bowers JE, Wang X, Paterson AH. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A* 2010;107:472–7.