Available Online at ESci Journals
Journal of Plant Breeding and Genetics
ISSN: 2305-297X (Online), 2308-121X (Print)
http://www.escijournals.net/JPBG

COMPLETE CHLOROPLAST GENOME SEQUENCE OF ASPARAGUS (*ASPARAGUS OFFICINALIS* L.) AND ITS PHYLOGENETIC POSITION WITHIN ASPARAGALES

Wentao Sheng*, Xuewen Chai, Yousheng Rao, Xutang Tu, Shangguang Du
Department of Biological Technology, Nanchang Normal University, Nanchang 330032, Jiangxi, China.

A B S T R A C T

Asparagus (*Asparagus officinalis* L.) is a horticultural plant with health care, which is meaningful to sequence the entire chloroplast (cp) genome of asparagus with Hiseq4000 platform. The complete cp genome maps a circular molecule of 156,699bp built with a quadripartite organization: two inverted repeats (IRs) of 26,531bp, separated by a large single copy (LSC) sequence of 84,999bp and a small single copy (SSC) sequence of 18,638bp. A total of 112 genes comprising of 78 protein-coding genes, 30 tRNAs and 4 rRNAs were successfully annotated, 17 of which included introns. The identity, number and GC content of asparagus cp genes were similar to those of other asparagus species genome. Analysis revealed 81 simple sequence repeat (SSR) loci, most composed of A or T, contributing to a bias in base composition. A maximum likelihood of phylogenomic evolution analysis showed that asparagus was closely related to *Polygonatum cyrtonema* that belonged to the genus *Asparagales*.The availability of the complete cp genome sequence of asparagus provides valuable information for chloroplast genetic engineering and phylogenetic analyses in *Asparagales*.

Keywords: *Asparagus officinalis* L., Chloroplast genome, Phylogenomic evolution, *Asparagales*.

INTRODUCTION

Compared with the nuclear genome, chloroplast (cp) is a major semi-autonomous organelle, with obvious characteristics of relatively conserved size, multicopies, stable gene content and structure, a maternal inheritance manner, a moderate nucleotide evolution rate with protein-coding genes, and so on (Gao *et al*., 2010). The comparison of complete cp genome sequences is becoming increasingly important for the study of reconstructing the evolutionary relationships, classification and genetic diversity of plants (Li *et al*., 2013). Up to date, the number of 2,113 cp genomes of green plants uploaded to NCBI (https://www.ncbi.nlm.nih.gov /genomes/ Genomes Group.cgi?taxid=2759&opt=plastid) has been available until Aug 16th, 2017.

Asparagus officinalis L. is a perennial herb with functional compound, which is widely cultivated in China (Zhong *et al*., 2015; Chen *et al*., 2016). And *Asparagus officinalis* L. has been used as a model plant for sex determination and differentiation due to the presence of both dioecious and androdioecious plants within a population (Alex and Jim, 2017). However, the taxnomic boundaries are not well defined for *Asparagus officinalis* L. The asparagus plants are divided into the *Liliaceae* family, in the traditional morphological classification method (Stevens, 1986); but in the Angiosperm Phylogeny Group (APG I) classification system, the asparagus is classified into the *Asparagaceae*, which was divided into one individual family (Angiosperm Phylogeny Group,1998); in 2003 revised version of APG II classification, strict definition of *Asparagaceae* only includes asparagus plants, and it was included the following branches of *Asparagaceae*, *Aphyllanthaceae*, *Hyacinthaceae*, *Hesperocallidaceae*, Laxmanniaceae, *Ruscaceae* plants in the wide sense (Angiosperm Phylogeny Group, 2003). Furthermore, the classification of the genus and its subdivisions are still unresolved questions in the taxonomy of *Asparagaceae*. Therefore, availability of a complete *A. officinalis* L. cp genome sequence is crucial

* Corresponding Author:
Email: shengwentao2003@163.com

to understand phylogenetic relationships among major lineages of *Asparagaceae* and facilitates asparagus genetic improvement.

These problems would be expected to be solved by phylogenetic analysis of its cp genome. However, there is a little sequence and cp genomic information available for asparagus, which posed a barrier to current studies in these areas (Lee *et al*., 1996). In this study, we reported the complete chloroplast genome sequence of *Asparagus officinalis* L. (deposited into Genebank accession number: NC_034777.1) from the whole-genome high-throughput sequencing data.

MATERIALS AND METHODS

DNA extraction and sequencing: Total genomic DNA was extracted from 1 g of fresh leaves from *Asparagus officinalis* L. cv. Atlas plant (male) cultivated in the greenhouse of Jiangxi Agricultural Science Academy from Nanchang (E115°27′, N28°09′), using mCTAB method (Li *et al*., 2013). The genome libraries were constructed and sequenced on the manufacturer's instructions with multiplexing on HiSeq4000 flow cell lanes (Illumina Inc.).

Chloroplast genome assembly and annotation: For the genome, the raw reads were assembled into non-redundant contigs with SPAdes 3.6.1 and SOAPdenovo2, a de novo sequence assembly software package, with k=30 and scaffolding contigs having a minimum length of 100 bp (Bankevich *et al*., 2012; Luo *et al*., 2012). The contig of the cp genome was screened by Blast program for each of the spliced contig, and the cp genome contig was assembled by using Sequencher 5.4.6 (http://www.genecodes. com) (Altschul *et al*.,1997). Then all the reads were mapped to the spliced cp genome sequence to verify the correctness of the spliced contig used by Geneious 8.1 (Kearse *et al*., 2012). Primer walking was then used to fill the gaps between the seven to twelve large contigs with additional Sanger sequencing, and to verify the junctions between the single-copy and the IRs regions.

Chloroplast genome annotation: Genome annotation was accomplished using the Dual Organellar Genome Annotator (DOGMA) software to annotate the genes encoding proteins, transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs) (Wyman *et al*., 2004). All of the identified tRNA genes were further verified using the corresponding structures predicted by tRNAscan-SE 1.21 (http://lowelab.ucsc.edu/ tRNAscan-SE). The genome map of *Asparagus officinalis* L. was drawn by Organellar Genome DRAW (http://ogdraw.mpimp-golm.mpg.de/ index.shtml) (Lohse *et al*., 2013).

Repeat structure finding: Microsatellite (mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats) detection was performed by using Websat (http://wsmartins. net/websat/) and MISA (http://pgrc.ipk-gatersleben.de/misa/) with thresholds of ten repeat units for mononucleotide, five repeat units for di- and tri-nucleotide, and three repeat units for tetra-, penta- and hexanucleotide.

Comparative chloroplast genomic analysis: With the cp genome of *Asparagus officinalis* L. as a reference, four species (*Asparagus officinalis* L., *Polygonatum sibiricum, Agave americana, Anemarrhena asphodeloides*) wree compared to assess the variability of the cp genome.

Phylogenetic analysis: Forty-two species representing the *Asparagales* lineage of angiosperm were sampled to reconstruct the phylogeny of *Asparagales* with cp coding-protein genes.The cp genome of *Lilium brownii* var. viridulum and *Oryza sativa* L. were used as outgroup. There were altogether seventy-seven genes shared by all forty-four cp genomes. Each gene from each genomic sequence was modified manually and was aligned using ClustalW. The resulting gene alignments were assembled into a data matrix for each genome. Optimal trees were conducted with Maximum likelihood (ML) analyses using in Phylogenetic Analysis Using Parsimony (PAUP) version 4.0b10 (Swofford, USA). The appropriate evolution model was determined implemented in the Model3.7 software.

RESULTS AND DISCUSSION

Genome content and organization in *Asparagus officinalis* L.: The complete cp genome of *Asparagus officinalis* is 156,699 bp in size and exhibits a typical circular structure including a pair of inverted repeats (IRs) of 25,783 bp separated the genome into two single-copy regions (large single copy, LSC 82,212 bp; small single copy, SSC 16,670 bp; Figure.1). Coding regions (91,260 bp), including protein-coding genes (79,413 bp), tRNA genes (2,801 bp), and rRNA genes (9,046 bp), account for 60.66% of the genome, while noncoding regions (59,188 bp), including introns (17,750 bp) and intergenic spacers (41,438 bp), account for the remaining 39.34% of the genome. The overall A+T content of the whole genome is 62.24%.

There is a total of 136 genes predicted in the genome, of which 112 are unique, including 78 protein-coding genes, 30 tRNA genes, and 4 ribosomal RNA genes

(Figure 1 and Table 1). Ten protein coding genes, 8 tRNA and 4 rRNA genes are duplicated in the IR regions. However, only a part of *ycf1* gene is duplicated in the junction of SSC and IRA regions. According to its function, the predicted genes are divided into three categories. The first category owns a total of 60 genes including RNA polymerase subunits, rRNA, the vast majority of tRNA gene, associated with transcription and translation; the second category has 47 genes related to photosynthesis, including rubisco large subunit genes, the photosynthetic electron transport chain components genes and the presumed NAD(P)H dehydrogenase subunit genes; the third kinds are associated with amino acids, fatty acids and other substances biosynthesis related genes, as well as some unknown function genes (Table 1).

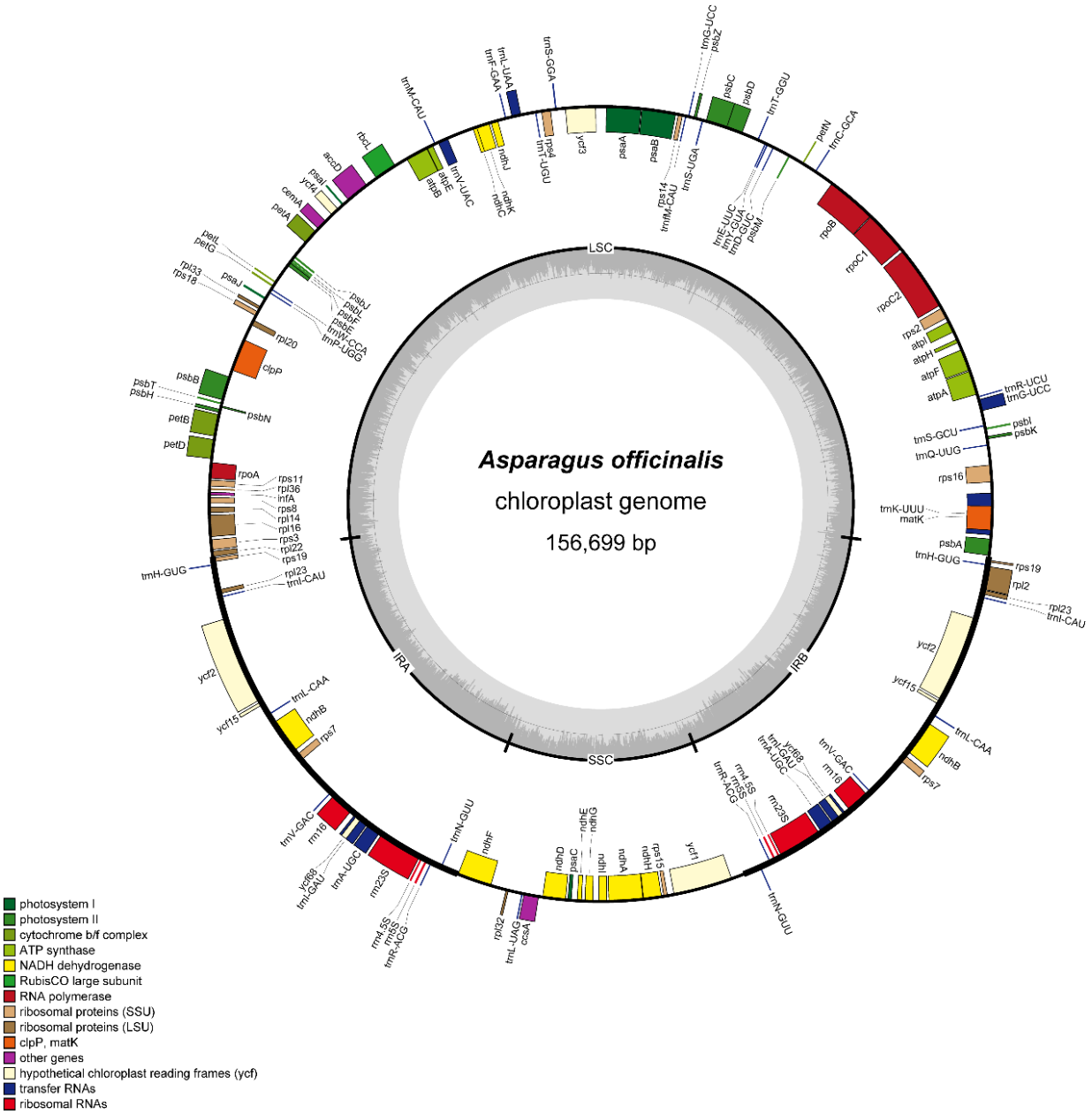

Figure 1: Chloroplast genome map of *Asparagus officinalis* L. The genes inside and outside of the circle are transcribed in the clockwise and counterclockwise directions, respectively. Genes belonging to different functional groups are shown in different colours. The thick lines indicate the extent of the inverted repeats (IRa and IRb) that separate the genomes into small single-copy (SSC) and large single-copy (LSC) regions.

The *infA*, *ycf1*, *ycf68* are pseudogenes judging by the presence of several terminal codons in the coding regions. There are 17 introns containing genes (12 protein-coding and 5 tRNAs), among which 9 protein-coding genes and 5 tRNAs contain an intron, and 3 protein-coding genes (*clpP*, *ycf3*, *rps12*) have two introns (Table 2). And *rps12* is a special trans-splicing gene, with it's 5'-end exon located in the LSC region, and 3'-end exon situated in the IR region.

Table 1. Gene content of complete chloroplast genome of *Asparagus officinalis* L.

Category for genes	Group of gene	Name of gene
Photosynthesis-related genes	the large subunit of Rubisco	*rbcL*
	subunits of photosystem I	*psaA,psaB,psaC,psaI,psaJ*
	assembly/stability of photosystem I	*ycf3*[a]*,ycf4*
	subunits of photosystem II	*psbA,psbB,psbC,psbD,psbE,psbF,psbH,psbI,psbJ,psbK,psbL,psbM,psbN,psbT,psbZ*
	subunits of ATP synthase	*atpA, atpB, atpE, atpF*[a]*, atpH, atpI*
	cytochrome b/f compelx	*petA, petB*[a]*, petD*[a]*, petG, petL, petN*
	c-type cytochrome synthesis	*ccsA*
	subunits of NADPH dehydrogenase	*ndhA*[a]*, ndhB*[a,b]*, ndhC, ndhD, ndhE, ndhF ,ndhG, ndhH, ndhI, ndhJ， ndhK*
Transcription and translation related genes	DNA dependent RNA polymerase	*rpoA, rpoB, rpoC1*[a]*, rpoC2*
	ribosomal proteins	*rps2, rps3, rps4, rps7*[b]*, rps8, rps11, rps12*[a,b]*, rps14,rps15, rps16*[a]*, rps18, rps19,rpl2*[a,b]*, rpl14, rpl16*[a]*, rpl20, rpl22, rpl23*[b]*, rpl32, rpl33,rpl36*
	translation initiation factor	*infA*
RNA genes	ribosomal RNA	*rrn4.5*[b]*, rrn5*[b]*, rrn16*[b]*, rrn23*[b]
	transfer RNA	*trnA-UGC*[a,b]*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-UCC, trnG-GCC*[a]*, trnH-GUG*[b]*, trnI-CAU*[b]*, trnI-GAU*[a,b]*,trnK-UUU*[a]*, trnL-CAA*[b]*, trnL-UAA*[a]*, trnL-UAG, trnfM-CAU,trnM-CAU, trnN-GUU*[b]*, trnP-UGG, trnQ-UUG,trnR-ACG*[b]*, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU,trnT-UGU, trnV-GAC*[b]*, trnV-UAC*[a]*, trnW-CCA, trnY-GUA*
Other genes	muturase	*matK*
	Envelop membrane protein	*cemA*
	ATP-dependent protease subunit p gene	*clpP*[a]
Genes of unknown function	conserved open reading frames	*ycf1, ycf2*[b]

[a]Genes containing introns;[b]duplicated gene (genes present in the IR regions).

Table2. Characteristics of genes including introns and exons in the cp genome of *Asparagus officinalis* L.

Gene	Exon I	Intron I	Exon II	Intron II	Exon III
trnA-UGC[c]	38	817	35	-	-
trnG-UCC	24	675	48	-	-
trnI-GAU[c]	36	943	37	-	-

trnK-UUU	35	2586	37	-	-
trnL-UAA	567	529	50	-	-
trnV-UAC	35	598	39	-	-
rps16	209	861	40	-	-
rpl2[c]	431	666	391	-	-
rpl16	399	386	9	-	-
atpF	411	840	144	-	-
petB	6	753	642	-	-
petD	8	736	481	-	-
ndhA	540	1127	558	-	-
ndhB[c]	755	701	777	-	-
ycf3	152	743	229	721	126
clpP	252	667	292	805	71
rpoC1	1625	751	436		

[c]Genes located in the inverted repeats.

Distribution characteristics of SSRs in chloroplast genome: There were three microsatellites including full type microsatellites, incomplete microsatellites and composite microsatellites types in the genome, of which the number was 69, 1 and 11, respectively. In the full type microsatellites, it was mainly composed of forty mono-base and eleven bi-base, accounting for 49.38% and 13.38% of the total number. In the mono-base repeats, the number of A and T were 13 and 25, accounting for 18.8% and 36.2%, while C and G accounted for only 2.8% (Table 3), which showed that A and T repeat were main types in cpSSR of *Asparagus* cp genome. In the bi-base repeats, AT (6) and TA type (2) account for 8.7% and 2.8%, respectively. And the number of tri-base, forth-bases, and fifth-bases types was was 5, 12 and 1, respectively. Therefore, it was concluded that SSR loci mostly composed of low repeat number, contributing to a bias of *Asparagus officinalis* cp genome.

Table 3. Types of complete repeat copy of cpSSR in asparagus cpDNA.

Base repeat type	cpSSR motif	cpSSR number	Ratio of total （%）
Mono-	A	13	18.8
	T	25	36.2
	C	2	2.8
Bi-	AT	6	8.7
	TA	2	2.8
	TC	2	2.8
	GA	1	1.4
Tri-	ATA	1	1.4
	ATT	2	2.9
	CTC	1	1.4
	TTC	1	1.4

Comparison with other cp genomes in genome content and organization in *Asparagales:* The genomic features of *Asparagus officinalis* and those of three species from other families were compared with *Asparagales* order, including *Polygonatum sibiricum* (NC_029485), *Agave americana* (KX519714), *Anemarrhena asphodeloides* (KX931449). The sequence length (156,699 bp), the GC content (37.76%), the gene content and arrangement of *Asparagus officinalis* were similar to those of other species within the *Asparagales* (Table 4).

Table 4. Comparison of general features of the cp genomes in *Asparagales*.

Genome features	*Asparagus officinalis* L.	*Polygonatum sibiricum*	*Agave americana*	*Anemarrhena asphodeloides*
Total length(bp)	156699	152960	157274	156867
LSC length(bp)	84999	81470	86060	85039
SSC length(bp)	18638	18519	19500	19846
IRs(bp)	53062	52971	51714	51982
Total genes	112	115	108	112
Protein genes	78	81	74	78
rRNA genes	4	4	4	4
tRNA genes	30	30	30	30
GC content (%)	37.76	37.95	37.84	37.85

Phylogenetic analysis of *Asparagales* species using cp genome sequences: The phylogenetic relationships of *Asparagus officinalis* in monocotyledon plants was reconstructed, based on 77 protein-coding genes sequences obtained from 42 taxa of *Asparagales*. The total aligned length of these genes was 75,472 characters. Maximum likelihood (ML) analysis was performed with PAUP version 4.0b10. The ML tree strongly supported *Asparagus officinalis* and *Polygonatum cyrtonema* formed as sister taxa with a pp value of 100 % (Figure. 2). And the relationship of *Asparagus officinalis* and *Lilium distichum* evolved from separate monophyletic clade, respectively. Therefore, the phylogenetic position within *Asparagales* from the cp genome data didn't support the traditional classification system of Liliaceae (Gao *et al*., 2013), but it was proved to accord with the new classification APG system (Angiosperm phylogeny group, 2003&2009&2016).

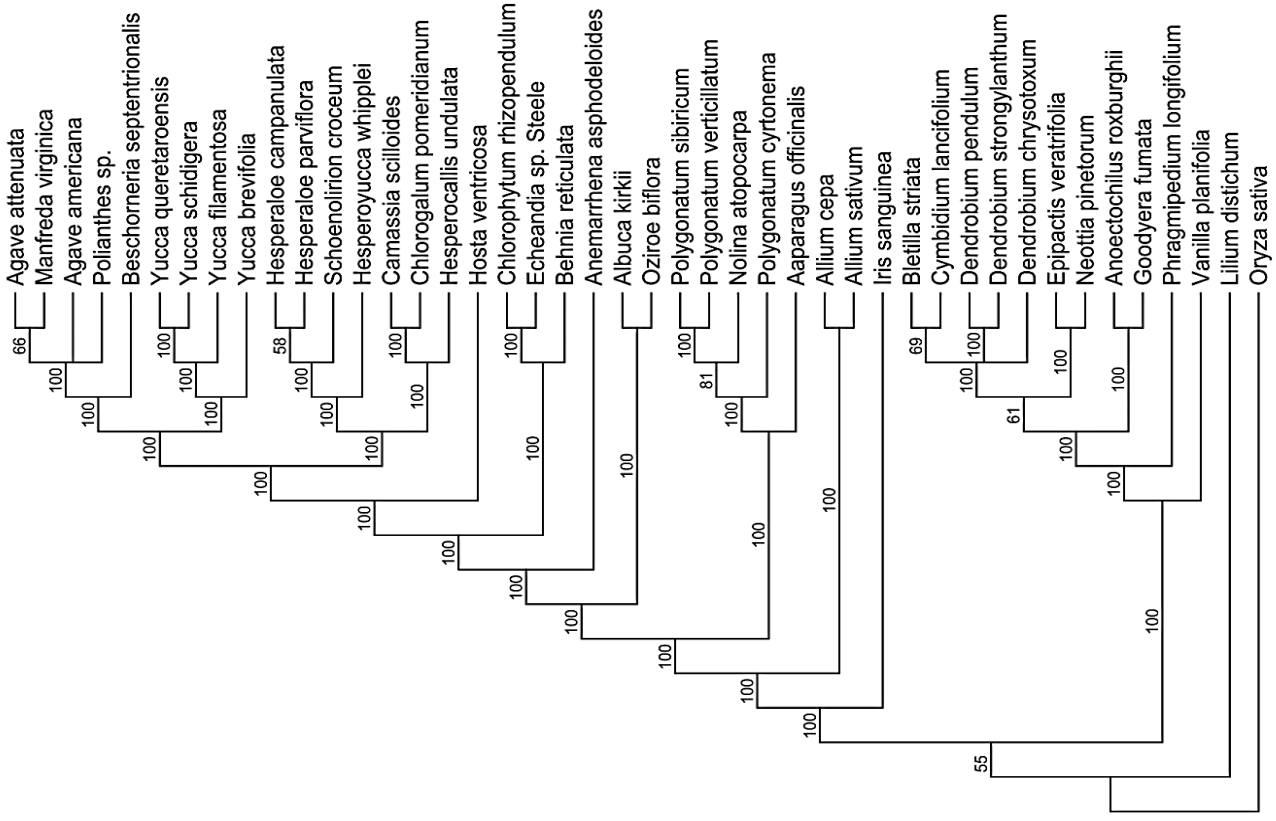

Figure 2 Maximum likelihood (ML) analysis inferred from 77 genes protein coding genes from 42 taxa.

CONCLUSION

It was reported that the whole cp genome sequence of *Asparagus officinalis* built with a typical quadripartite organization of 156,699 bp in this study. And the analysis of genome structure and sequence showed that it was nearly identical to those species in *Asparagales.* The phylogenetic evolution of 77 coding genes in cp genome indicated that *Asparagus officinalis* were much more closely related to *Polygonatum cyrtonema*, but not to *Lilium distichum* in *Liliaceae*.

ACKNOWLEDGMENT

We thank Dr Guangyu Chen and the Innovative Team of *Asparagus officinalis* L, Jiangxi Academy of Agricultural Sciences, Nanchang 330000, China for providing seeds of asparagus. We also acknowledge financial supports by Natural science youth foundation project of Jiangxi (20171BAB214024), Science and technology project of Jiangxi Provincial Department of Education (GJJ161243), a Doctoral research project of Nanchang Normal University (NSBSJJ2015036), and the program of the molecular breeding innovation team of Nanchang Normal University (090150000114).

REFERENCES

Alex, H., L. M. Jim. 2017. A century of sex determination in flowering plants. J. Hered., 108(1): 69-77.

Alekseyev, M.A., P.A. Pevzner. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol., 19:455-477.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants.Ann. Mo. Bot. Gard., 85(4):531-553.

Angiosperm phylogeny group. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc., 141: 399-436.

Angiosperm phylogeny group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Bot. J. Linn. Soc., 161: 105-121.

Angiosperm phylogeny group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Bot. J. Linn. Soc., 181(1):1-20.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25:3389-3402.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, Sirotkin, A. V. Vyahhi. N.G. Tesler, R. B. Luo, B. H. Liu, Y. L. Xie, Z. Y. Li, W. H. Huang, J. Y. Yuan, G. Z. He, Y. X. Chen, Q. Pan, Y.J. Liu, J. B. Tang, G. X. Wu, H. Zhang, Y. J. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. L. Han, D. W. Cheung, S. M. Yiu, S. L. Peng, X.Q. Zhu, G. M. Liu, X. K. Liao, Y. R. Li, H. M. Yang, J. Wang, T. W. Lam, J. Wang. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 1(1):1-6.

Chen, J. L., Y. Zheng, L. Qin, Y. Wang, L. F. Chen, Y. J. He, Z. J. Fei, G. Lu, 2016.Identification of miRNAs and their targets through high-throughput sequencing and degradome analysis in male and female *Asparagus officinalis*. BMC Plant Biol., 16:80-98.

Gao, L., Y. J. Su, T. Wang. 2010. Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. J. Syst. Evol., 48(2):77-93.

Gao, W. J., S. F. Li, G. J. Zhang, N. N. Wang, C. L. Deng, L. D. Lu. 2013. Comparative analysis of gene expression by microarray analysis of male and female flowers in *Asparagus officinalis*. Biosic. Biotechnol. Biochem., 77(6):1193-1199.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, A. Drummond. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647-1649.

Lee, Y. O., A. Kanno, T. Kameya. 1996. The physical map of the chloroplast DNA from *Asparagus officinalis* L. Theor. Appl. Genet., 92(1):10-14.

Lohse, M., O. Drechsel, S. Kahlau, R. Bock. 2013. OrganellarGenomeDRAW--a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res., 41: W575-581.

Li, J. L., S. Wang, J. Yu, L. Wang, S. L. Zhou. 2013. A modified CTAB protocol for plant DNA extraction.Chinese Bulletin of Botany. 48 (1): 72-78.

Li, X. W., H. H. Gao, Y. T. Wang, J. Y. Song, R. Henry, H. Z.

Wu, Z. G. Hu, H. Yao, H. M. Luo, K. Luo, H. L. Pan, S. L. Chen. 2013. Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. Sci. China Life Sci., 56: 189-198.

Stevens, P.F..1986. Evolutionary classification in botany, 1960-1985. Journal of the Arnold Arboretum. 67(3):313-339

Wyman, S. K., R. K. Jansen, J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 20:3252-3255.

Zhong, C. G., C.Y. Jiang, X.C. Xia, T. Mu, L. G. Wei, Y. T. Lou, X. S. Zhang, Y. Q. Zhao, X. L. Bi. 2015. Antihepatic fibrosis effect of active components isolated from green Asparagus (*Asparagus officinalis* L.) involves the inactivation of hepatic stellate cells. J. Agric. Food Chem., 63 (26):6027-6034.