DOI:10.13275/j.cnki.lykxyj.2019.02.011

叶绿体全基因组序列确定钻天柳在杨柳科中的系统发育位置

冯楚航[1],何彩云[1],王 莹[2],曾艳飞[1]*,张建国[1]

(1.国家林业和草原局林木培育重点实验室,中国林业科学研究院林业研究所,北京 100091;
2.深圳华大基因科技有限公司,广东 深圳 518083)

摘要:[目的]为解决濒危物种钻天柳在杨柳科中分类学位置的争议。[方法]本研究通过二代测序技术,从头测序、拼接得到钻天柳叶绿体基因组全序列,并与已发表的9种杨属植物和4种柳属植物的叶绿体基因组全序列进行比较,采用最大似然法、最大简约法和贝叶斯推断法分析了这些物种的系统发育关系。[结果]研究发现:钻天柳总基因组为155 661 bp,由长度为84 536 bp的长单拷贝(LSC)区域和16 215 bp的短单拷贝(SSC)区域,以及一对分隔开它们的27 455 bp的反向重复序列(IRS)组成。钻天柳叶绿体基因组总GC含量为36.68%,共有113个不同的基因,包括79个蛋白质编码基因,30个tRNA基因和4个rRNA基因,其中,有20个基因分布于反向重复区;在所有基因中,有14个基因包含1个内含子,3个基因(*rps*12、*clpP*、*ycf*3)内含有2个内含子;系统发育分析以100%的支持率将钻天柳与柳属黄花柳亚属的2个物种聚为一支,杨属的所有物种聚为另一支。[结论]本研究首次组装并注释了钻天柳叶绿体基因组全序列,并明确支持钻天柳并入柳属,而非单独成属,这将为钻天柳甚至杨柳科的系统进化研究提供重要参考。

关键词:二代测序;柳属;杨柳科;叶绿体基因组;钻天柳
中图分类号:S718.46 **文献标识码:**A **文章编号:**1001-1498(2019)02-0073-05

钻天柳(*Chosenia arbutifolia* (Pall.) A. Skv.)是一种主要分布在亚洲东北部的雌雄异株树种,原属杨柳科(Salicaceae)的一个单种属,钻天柳属(*Chosenia*)[1]。由于其形态特征介于杨属(*Populus*)植物和柳属(*Salix*)植物之间,钻天柳分类地位的确定对于研究杨柳科家族的演化有着十分重要的价值。然而,近期的研究对钻天柳在杨柳科中的系统发育位置提出了质疑。例如,Sohma[2]研究了72种柳及杂交种的孢粉形态,其结果认为钻天柳属可合并到柳属中;Azuma等[3]、Chen等[4]基于叶绿体片段分析探讨了一些主要的柳属分类系统,同样支持将钻天柳属归入柳属。然而,这些系统发育研究只是基于个别片段的分析,也并未得到国内分类学家的重视。另外,森林的过度砍伐,使钻天柳生长的林地环境条件发生了变化,供其生长的河流受到污染,极大地影响了钻天柳种子的萌发与幼苗的生长,导致钻天柳的分布区正在日益缩减。钻天柳,因其分布高度受限而濒临灭绝[5],因此,被列为国家Ⅱ级重点保护野生植物。钻天柳系统发育地位的确定,将为钻天柳的保护提供重要的理论基础。

叶绿体起源于与之内共生的蓝藻[6],并且有其自身的遗传物质。叶绿体基因组由120～160 kb的环状双链分子组成,这是一种高度保守的结构[7]。植物叶绿体基因组一般都有其独特的DNA区域,称为长单拷贝区(LSC)和短单拷贝区(SSC),它们由两个反向重复序列(IRS)分隔开[8]。与植物核基因组相比,植物叶绿体基因组的替换率低得多[9],并且由于叶绿体基因组是单亲遗传的,这意味着叶绿体基因组是系统发育分析遗传标记的重要来源[10]。所以,叶绿体基因组可以用来研究植物种群之间的相互关系和生物多样性[11]。

本研究从头测序、组装并注释了钻天柳叶绿体

收稿日期:2018-03-21 修回日期:2019-01-15
基金项目:国家自然科学基金(31670666)
* 通讯作者:曾艳飞.E-mail:zengyf@caf.ac.cn

基因组完整的 DNA 序列并将其与从美国国家生物技术信息中心(NCBI)上得到的所有其他可用的杨属和柳属物种完整的叶绿体基因组 DNA 序列进行分析比对。这些分析揭示了钻天柳叶绿体基因组的结构和功能信息,并明确了钻天柳在杨柳科的系统发育地位。

1 研究方法

1.1 钻天柳全基因组序列

本研究从中国辽宁省丹东市宽甸白石砬子自然保护区附近的天然林中(40°50′00″ ~ 40°57′12″ N, 124°44′07″ ~ 124°57′30″ E)采集了 1 棵雄性钻天柳的叶片样本,并利用 CTAB 法提取总 DNA;然后采用 Illumina 的 hiseq2000 平台,通过全基因组鸟枪法测序获得 60 G 读长为 100 bp 的 DNA 数据。通过与毛果杨叶绿体基因组全序列(GenBank 登录号为 EF489041)比对,从这些短片段中筛选出 100 Mb 高质量的叶绿体基因组片段;并用 SOAP de novo 执行程序[12]将这些短片段从头组装钻天柳叶绿体基因组序列。所有组装得到的叶绿体重叠群使用 ContigExpress[13]进行全基因组的组装,采用簸箕柳(*S. suchowensis*)叶绿体基因组全序列(KM983390)作为参考。重叠群的间隙在手动设计引物后采用 Sanger 测序来进行人工补洞。得到的完整钻天柳叶绿体基因组序列通过 Dual OrganellarGenoMe Annotator (DOGMA)[14]用默认参数预测其蛋白质编码基因、转运 RNA(tRNA)基因和核糖体 RNA(rRNA)基因。由于 DOGMA 的局限性,一些内含子以及外显子的边界不能很好的识别。因此,Blastn(http://blast.ncbi.nlm.nih.gov/)软件被用来比较钻天柳叶绿体基因组序列与已完成注释的毛果杨和簸箕柳叶绿体基因组序列,从而进一步确定边界位置。注释的钻天柳叶绿体基因组全序列发表在 GenBank(NCBI),登录号为 KX781246。叶绿体的环状基因组图谱采用 OrganellarGenomeDraw 工具[15]绘制得到。

1.2 系统发育分析

从 NCBI 上下载得到银白杨(*P. alba*, AP008956)、大叶钻天杨(*P. balsamifera*, KJ664927)、青杨(*P. cathayana*, KP729175)、胡杨(*P. euphratica*, KJ624919)、弗氏黑杨(*P. fremontii*, KJ664926)、冬青叶杨(*P. ilicifolia*, KX421095)、欧洲山杨(*P. tremula*, KP861984)、毛果杨(*P. trichocarpa*, EF489041)、滇杨(*P. yunnanensis*, KP729176)等 9 种杨树,垂柳(*S. babylonica*, KT449800)、*S. interior*(KJ742926)、红皮柳(*S. purpurea*, KP019639)、簸箕柳(*S. suchowensis*, KM983390)等 4 种柳树,和巴西橡胶树(*Hevea brasiliensis*, HQ285842)的完整叶绿体基因组序列。将这些叶绿体基因组序列与本研究得到的钻天柳的完整叶绿体基因组序列通过 mVISTA[16]进行比对分析。通过进行两两比较获得对齐的序列矩阵,然后利用巴西橡胶树的叶绿体基因组序列作为外类群,采用 MEGA7[17]软件分别用最大似然法(Maximum likelihood, ML)和最大简约法(Maximum parsimony, MP)构建了系统发育树。在 ML 建树过程中,位点采用相同的进化速率(Uniform rates),通过 GTR 模型(General Time Reversible substitution model)确定最优树,并采用 SPR(Subtree-Pruning-Regrafting-Fast)作为 ML 的启发式搜索策略。MP 方法采用 TBR(Tree-Bisection-Reconnection)作为搜索方法,初始树数量为 10,搜索等级为 1,最多保留的树为 100。这 2 种方法中,空白部分都采用“部分缺失”处理,分支的置信度均采用 1 000 次的靴带分析(Bootstrap)。同时利用 MrBayes[18]软件用贝叶斯推断(Bayesian inference, BI)法构建了系统发育树,采用 GTR 替代模型和伽马分布率,后验概率分布获得 200 个样本,每 1 000 代进行 1 次计算。

2 研究结果

2.1 钻天柳叶绿体基因组的结构

组装注释得到钻天柳叶绿体基因组结构见图 1。该叶绿体基因组包括 84 536 bp 的长单拷贝区(LSC),16 215 bp 的短单拷贝区(SSC),2 个单拷贝区被一对 27 455 bp 的反向重复序列(IRA, IRB)分隔开;共注释了 113 个基因,包括 79 个编码蛋白,30 个 tRNA 和 4 个 rRNA,并且有一个分布在反向重复 B 区的由 *ycf*1 基因的部分序列组成的假基因。在蛋白编码基因中,有 14 个基因包含 1 个内含子,3 个基因(*clpP*, *ycf*3 和 *rps*12)包含 2 个内含子;*rps*12 基因被反式剪接分为 3 部分,其中一部分位于长单拷贝区,另外 2 部分分别位于 2 个反向重复区。叶绿体基因组的整体 GC 含量是 36.68 %,长单拷贝区为 34.40%,短单拷贝区为 30.96%,反向重复区则为 41.88%。

2.2 钻天柳在杨柳科中的系统发育地位

采用叶绿体全基因组序列,基于 ML、MP 和 BI 方法构建的钻天柳与 9 种杨属植物和 4 种柳属植物

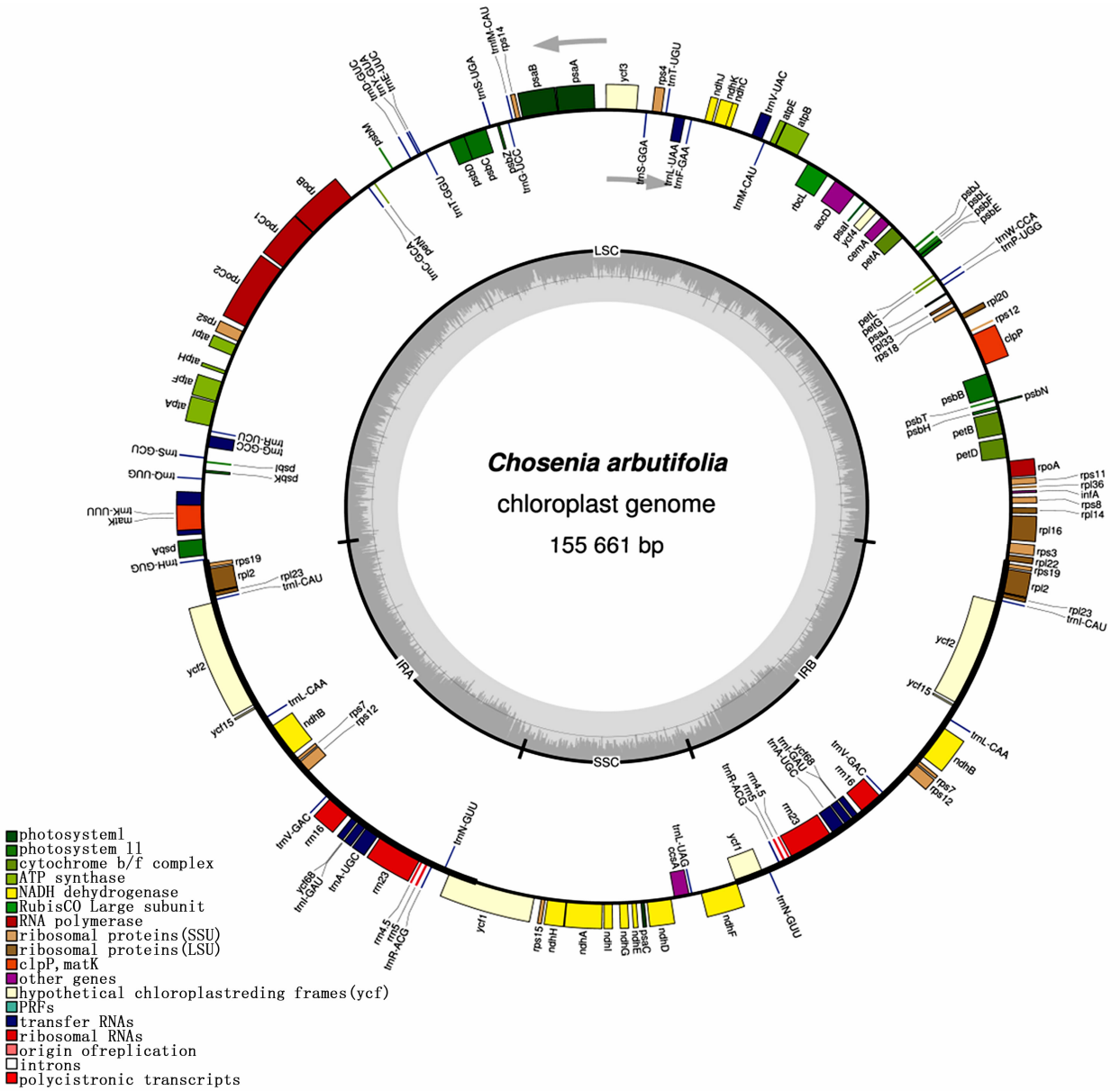


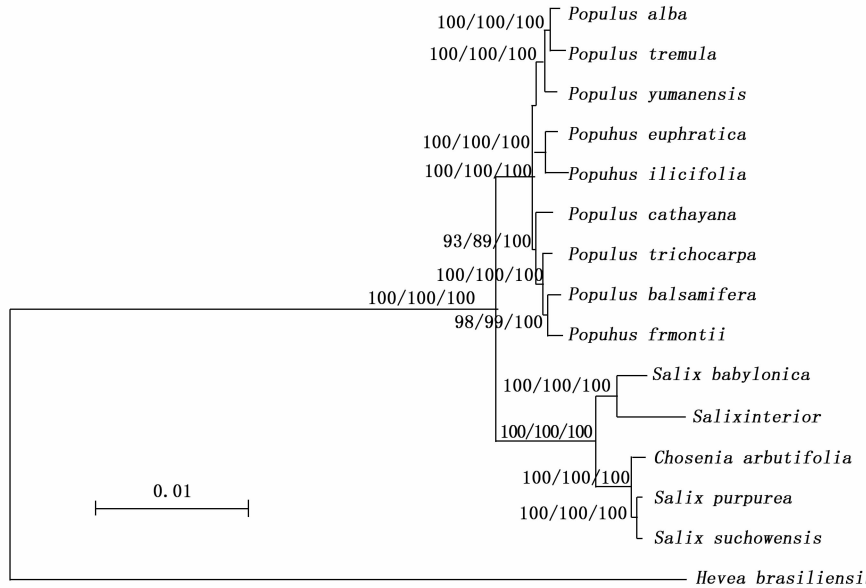
图 1　钻天柳叶绿体全基因组结构（GenBank 登陆号为 KX781246）.

Fig. 1　Gene map of the complete Chosenia arbutifolia chloroplast genome (GenBank accession number KX781246)

的系统发育树（图 2）得到的拓扑结构基本一致，只在个别节点支持率略有差别。以巴西橡胶树作为外类群，杨柳科植物在 3 种方法中都以 100% 的支持率聚为一大支。然后，杨柳科分成两支：所有杨属植物聚于一支，钻天柳和 4 种柳树植物聚为一支，3 种方法在这两支上的支持率均为 100%。杨属分支又分为了 3 个分支，分支 1 与分支 2 分别以 56%、56% 和 0%（ML、MP 和 BI）的支持率构成姐妹群关系，然后再与分支 3 构成姐妹群关系，其支持率为 100%。分支 1 中青杨处于基部，并分别以 93%、89% 和 100%（ML、MP 和 BI）的支持率与弗氏黑杨、大叶钻天杨及毛果杨聚在一起；分支 2 中银白杨和欧洲山杨构成姐妹群，然后再与滇杨构成姐妹群，其支持率均为 100%；分支 3 中胡杨和冬青叶杨以 100% 的支持率构成姐妹群。柳属植物中则分为 2 个分支，垂柳和 *S. interior* 在 3 种方法中都以 100% 的支持率聚为一支，钻天柳则与构成姐妹群的红皮柳和簸箕柳均以 100% 的支持率聚于另一分支。

3　讨论

本研究通过二代测序技术首次得到了钻天柳的叶绿体基因组全序列，该序列对于后期利用叶绿体



注:系统发育树分支上的数值分别为最大似然法(ML)、最大简约法(MP)和贝叶斯推断法(BI)的节点支持率(仅标高于50%的节点)。

Note: Numbers above branches are bootstrap support values based on maximum likelihood, maximum parsimony and bayesian inference, respectively (only values larger than 50% were shown).

图 2 利用钻天柳叶绿体全基因组序列与 NCBI 上已知的杨属和柳属叶绿体全基因组序列构建的系统发育树。

Fig. 2 Phylogenetic tree based on all available complete cpDNA sequences from Populus and Salix species in NCBI

序列研究钻天柳有着重要的参考价值。比较后发现,钻天柳叶绿体基因组的基因组成和结构与网上发表的杨属和柳属物种的叶绿体基因组非常相似,说明杨柳科叶绿体基因组的基因组成和结构具有较高的保守性。

利用本研究组装出来的叶绿体基因组全序列与9个杨属和4个柳属物种构建系统发育树,拓扑结果表明,钻天柳并未如传统分类认为的那样单独形成一支,也没有处于杨柳科的基部,而是被聚到柳属分支内部,与黄花柳亚属(*Vetrix*)的红皮柳和簸箕柳(二者为杞柳(*S. integra*)的近缘种)以100%支持率聚为一支。这些系统发育关系与 Chen 等[4]的分析是一致的,他基于 *trn*D-T、*atp*B-*rbc*L、*rbc*L 序列联合分析的邻接树以及严格一致树中发现,旧世界柳亚属的垂柳与北美柳亚属的 *S. interior* 聚于一个分支。柳属分支中的钻天柳与细柱柳(*S. gracilistyla*)(同为杞柳的近缘种)聚于一个分支,因此表明钻天柳属可并入柳属。当前笔者基于叶绿体基因组全序列数据的分析可以更进一步明确钻天柳应被纳入柳属。另外,笔者构建的系统发育树所确定的杨属各物种之间的亲缘关系与 Kersten 等[19]得到的结果略有不同,主要区别在于青杨和胡杨,可能是由于他们是采用 UPGMA 法来构建的系统发育树。

4 结论

本研究通过将钻天柳叶绿体全基因组与从 NCBI 上下载得到的已知的杨柳科植物进行系统发育分析,明确了钻天柳应并入柳属。分子系统学在分类上问题解决效果良好,以往的分子系统学在分类上的应用主要以单个基因或者几个基因来进行系统发育分析。本研究则是通过完整的叶绿体基因来进行系统发育分析,这在物种的分类上有着较大的应用价值。随着测序技术的不断发展,杨柳科的系统学研究也将有着极大的促进,本文也能为以后的杨柳科系统进化研究提供一些参考。

参考文献:

[1] Skvortsov A K. Willows of Russia and Adjacent Countries: Taxonomical and Geographical Revision [M]. Joensuu: University of Joensuu, 1999:1 - 307.

[2] Sohma K. Pollen diversity in Salix (Salicaceae) [J]. Sci Rep Tohoku Univ, IV, 1993, 40(2): 77 - 178.

[3] Azuma T, Kajita T, Yokoyama J, *et al*. Phylogenetic relationships of *Salix* (Salicaceae) based on *rbc*L sequence data [J]. American Journal of Botany, 2000, 87(1): 67 - 75.

[4] Chen J H, Sun H, Wen J, *et al*. Molecular phylogeny of *Salix* L. (Salicaceae) inferred from three chloroplast datasets and its systematic implications [J]. Taxon, 2010, 59(1): 29 - 37.

[5] Hoshikawa T, Kikuchi S, Nagamitsu T, *et al*. Eighteen microsatel-

lite loci in *Salix arbutifolia* (Salicaceae) and cross-species amplification in *Salix* and *Populus* species[J]. Molecular Ecology Resources, 2009, 9(4): 1202 – 1205.

[6] McFadden G I, van Dooren G G. Evolution: red algal genome affirms a common origin of all plastids[J]. Current Biology, 2004, 14(13): R514 – R516.

[7] Odintsova M S, Yurina N P. Chloroplast genomics of land plants and algae[J]. Biotechnological applications of photosynthetic proteins: biochips, biosensors and biodevices, 2006: 57 – 72.

[8] Jansen R K, Raubeson L A, Boore J L, *et al*. Methods for obtaining and analyzing whole chloroplast genome sequences[M]//Methods in enzymology. Academic Press, 2005, 395: 348 – 384.

[9] Wolfe K H, Li W H, Sharp P M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs[J]. Proceedings of the National Academy of Sciences, 1987, 84(24): 9054 – 9058.

[10] Ravi V, Khurana J P, Tyagi A K, *et al*. An update on chloroplast genomes[J]. Plant Systematics and Evolution, 2008, 271(1 – 2): 101 – 122.

[11] Chen D, Zhang X, Kang H, *et al*. Phylogeography of *Quercus* variabilis based on chloroplast DNA sequence in East Asia: multiple glacial refugia and mainland-migrated island populations[J]. PLoS One, 2012, 7(10): e47268.

[12] Li R, Li Y, Kristiansen K, *et al*. SOAP: short oligonucleotide alignment program[J]. Bioinformatics, 2008, 24(5): 713 – 714.

[13] Lu G, Moriyama E N. Vector NTI, a balanced all-in-one sequence analysis suite[J]. Briefings in Bioinformatics, 2004, 5(4): 378 – 388.

[14] Wyman S K, Jansen R K, Boore J L. Automatic annotation of organellar genomes with DOGMA[J]. Bioinformatics, 2004, 20(17): 3252 – 3255.

[15] Lohse M, Drechsel O, Kahlau S, *et al*. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets[J]. Nucleic Acids Research, 2013, 41(W1): W575 – W581.

[16] Frazer K A, Pachter L, Poliakov A, *et al*. VISTA: computational tools for comparative genomics[J]. Nucleic Acids Research, 2004, 32(suppl_2): W273 – W279.

[17] Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets[J]. Molecular Biology and Evolution, 2016, 33(7): 1870 – 1874.

[18] Ronquist F, Huelsenbeck J, Teslenko M. Draft MrBayes version 3.2 manual: tutorials and model summaries[M/OL]. Distributed with the software from http://brahms. biology. rochester. edu/software. html, 2011.

[19] Kersten B, Rampant P F, Mader M, *et al*. Genome sequences of Populus tremula chloroplast and mitochondrion: implications for holistic poplar breeding[J]. PloS One, 2016, 11(1): e0147209.

Phylogenetic Position of *Chosenia arbutifolia* in the Salicaceae Inferred from Whole Chloroplast Genome

FENG Chu-hang[1,2], *HE Cai-yun*[1,2], *WANG Ying*[2], *ZENG Yan-fei*[1,2], *ZHANG Jian-guo*[1,2]

(1. Key Laboratory of Tree Breeding and Cultivation, National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China; 2. BGI-Shenzhen, Shenzhen 518083, Guangdong, China)

Abstract: [**Objective**] To resolve the controversy over the phylogenetic position of *Chosenia arbutifolia* in Salicaceae. [**Method**] The whole chloroplast genome sequences of *C. arbutifolia* was determined by next-generation sequencing, and the phylogenetic position of *C. arbutifolia* was investigated by comparing its sequences with all available complete chloroplast genome sequences from the genera *Populus* and *Salix*. [**Result**] The total genome was 155, 661 bp, consisting of two single-copy regions separated by a pair of inverted repeats (IRs) of 27, 455 bp. The large single-copy (LSC) and small single-copy (SSC) regions spanned 84, 536 bp and 16, 215 bp, respectively. The total GC content of the chloroplast genome was 36.68% and 113 unique genes were annotated, including 79 protein coding genes, 30 tRNA genes, and four rRNA genes. Twenty genes were duplicated in the inverted repeat regions, 14 genes contained one intron, and three genes (*rps*12, *clpP*, and *ycf*3) contained two introns. [**Conclusion**] A phylogenetic tree constructed from all available complete chloroplast genome sequences from the genera *Populus* and *Salix* based on maximum likelihood, maximum parsimony and Bayesian inference strongly supports the merging of *C. arbutifolia* into the genus Salix. This study would supply an important basis for the genetic study as well as conservation of *C. arbutifolia*.

Keywords: *Chosenia arbutifolia*; chloroplast genome; next-generation sequencing; Salicaceae; *Salix*

(责任编辑:张　研)