Molecular Phylogenetics and Evolution 129 (2018) 258–267

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenomics, biogeography, and adaptive radiation of grapes

Zhi-Yao Ma[a], Jun Wen[b], Stefanie M. Ickert-Bond[c], Ze-Long Nie[d], Long-Qing Chen[a], Xiu-Qun Liu[a,*]

[a] Key Laboratory of Horticultural Plant Biology (Ministry of Education), College of Horticulture and Forestry Science, Huazhong Agricultural University, Wuhan 430070, China
[b] Department of Botany, National Museum of Natural History, MRC166, Smithsonian Institution, Washington, DC 20013-7012, USA
[c] UA Museum of the North Herbarium and Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK 99775-6960, USA
[d] Key Laboratory of Plant Resources Conservation and Utilization, College of Biology and Environmental Sciences, Jishou University, Jishou, Hunan, China

ARTICLE INFO

Keywords:
Adaptive radiation
Evolutionary diversification
Genome resequencing
Phylogenetics
Vitis

ABSTRACT

The application of whole-genome resequencing based on next-generation sequencing technologies provides an unprecedented opportunity for researchers to resolve long-standing evolutionary problems. Taxa belonging to the grape genus (*Vitis* L.) represent important genetic resources for the improvement of cultivated grapes. However, it has been challenging to resolve the deep phylogenetic relationships within *Vitis*, limiting the current understanding of the evolutionary history of *Vitis* and preventing the use of valuable wild grape resources. In this study, we obtained whole-genome sequence data from 41 accessions representing most taxa within subgenus *Vitis* and aligned these sequences to the *Vitis vinifera* L. reference genome. We reconstructed deep phylogenetic relationships within subgenus *Vitis* based on 2068 single-copy orthologous genes, which led to a robust topology with bootstrap support values of 100% for almost all branches. Three main clades are recovered within subgenus *Vitis* reflecting their continental distribution through North America, Europe, and East Asia, respectively. Our results suggest that the most possible migration route of the East Asian *Vitis* is from northeastern Asia southward to South Asia and Southeast Asia. The East Asian *Vitis* seems to have experienced adaptive radiation during the Miocene. This study provides novel insights into the diversification history of the grape genus *Vitis*.

1. Introduction

The advent of next-generation sequencing (NGS) technologies has greatly improved our ability to resolve key biological problems at the genomic, transcriptomic, and epigenetic levels (Morozova and Marra, 2008; Reis-Filho, 2009; Davey et al., 2011; Koboldt et al., 2013; Goodwin et al., 2016). NGS platforms can help identify millions of variants in the genome, greatly promoting the development of genomics as a new discipline (Koboldt et al., 2013). The whole-genome resequencing technology based on NGS has been extensively used to resolve complex phylogenetic problems, reconstruct biogeographic histories, and investigate adaptive evolution in fungi, birds, fishes, livestocks, insects, and primates (e.g., Li et al., 2013; Qu et al., 2013; Xu et al., 2014; Zhou et al., 2014, 2016; Cornetti et al., 2015; Gladieux et al., 2015; Chen et al., 2016, 2018; Mei et al., 2017). Phylogenetic relationships among closely related species of flowering plants have also been inferred using whole-genome resequencing. For instance, Xu et al. (2012) used the genomes of three species of *Oryza* L. to reconstruct phylogenetic relationships and identified selected candidate regions in cultivated rice. Yoder et al. (2013) clarified phylogenetic relationships among 29 taxa within *Medicago* L. by using whole-genome resequencing data, while Liu et al. (2017) reconstructed the phylogeny of *Actinidia* among 25 taxa using the same technique.

The grape (*Vitis vinifera* L.) is one of the earliest domesticated and major economic fruit crops in the world (Myles et al., 2011). It is consumed as table grapes or processed into wines, raisins, and non-alcoholic grape juices (Wen et al., 2007; Myles et al., 2011; Wan et al., 2013). The grape genus (*Vitis* L.) and its closely related genera are widely recognized as valuable genetic resources of cultivated grapes (Zecca et al., 2012; Liu et al., 2016). Many wild grapes have been used as important resources to improve resistance of cultivated grapes against abiotic and biotic stresses (Staudt and Kassemeyer, 1995; Wang et al., 1995; Staudt, 1997; Brown et al., 1999; Wan et al., 2007; Zhang et al., 2012; Gerrath et al., 2015).

The genus *Vitis* is typically divided into two subgenera: subgenus *Muscadinia* Planch. ($2n = 40$) and subgenus *Vitis* ($2n = 38$), which are distinguished based on the morphological, anatomical, cytological, and molecular evidences (Moore, 1991; Mullins et al., 1992; Soejima and Wen, 2006; Tröndle et al., 2010; Zecca et al., 2012; Aradhya et al., 2013; Wan et al., 2013; Wen et al., 2013; Liu et al., 2016; Moore and

* Corresponding author.
E-mail address: liu_xiuqun@sina.com (X.-Q. Liu).

https://doi.org/10.1016/j.ympev.2018.08.021
Received 2 April 2018; Received in revised form 30 August 2018; Accepted 30 August 2018
Available online 06 September 2018

Wen, 2016). The phylogeny of subgenus *Vitis* has intrigued researchers for decades. However, resolving deep phylogenetic relationships of subgenus *Vitis* has long been considered a significant challenge, due to phenotypic plasticity, incomplete lineage sorting, hybridization, and introgression between closely related species, among others (Comeaux et al., 1987; Péros et al., 2011; Aradhya et al., 2013; Liu et al., 2016; Ma et al., 2016; Moore and Wen, 2016).

The subgenus *Vitis* includes ca. 70 species mainly distributed in the temperate regions of the Northern Hemisphere, showing an intercontinental disjunct distribution between North America and Eurasia (Chen et al., 2007; Wen et al., 2007, 2018b; Aradhya et al., 2008; Liu et al., 2016). However, there is still much debate over species delimitation and the infra-generic classification within subgenus *Vitis* (Aradhya et al., 2013; Wen et al., 2018a). Fifty-nine species of this subgenus were classified into 11 series by Galet (1988) based on morphology, ecology, and biogeography. Moore (1991) performed a comprehensive taxonomic revision of genus *Vitis* from North America and recognized 12 species and nine varieties within subgenus *Vitis* that were classified into five series. One of the major centers of grapevine diversity lies in East Asia, with the distribution of East Asian species of subgenus *Vitis* being centered in China (Li et al., 1996). However, the species number and classification of subgenus *Vitis* in China is in debate due to overlapping patterns of morphological variation (Li et al., 1996; Li, 1998; Wang et al., 2000; Kong, 2004; Wan et al., 2008). The Chinese species of subgenus *Vitis* were classified into 13 sections by Niu and He (1996), five sections and four series by Wang et al. (2000), and eight sections and five subsections according to Liu et al. (2011).

Several recent studies have reconstructed the phylogeny of subgenus *Vitis* based on molecular data (Di Gaspero et al., 2000; Pelsy, 2007; Aradhya et al., 2008, 2013; Tröndle et al., 2010; Péros et al., 2011; Zecca et al., 2012; Miller et al., 2013; Wan et al., 2013; Liu et al., 2016; Wen et al., 2018a). Although simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) are not adequate to resolve deep relationships within subgenus *Vitis,* data derived from these markers contributed to establishing a phylogenetic framework for the subgenus (Aradhya et al., 2008, 2013). Furthermore, phylogenies based on a few sequences did not confidently resolve deep phylogenetic relationships within the subgenus *Vitis* (Tröndle et al., 2010; Péros et al., 2011; Zecca et al., 2012; Wan et al., 2013; Liu et al., 2016). Additional studies based on SNP arrays corroborated the monophyly of the North American subgenus *Vitis*, but did not provide sufficient characters to resolve deep phylogenetic relationships within subgenus *Vitis* (Miller et al., 2013). Relationships among North American grapes where explored using complete chloroplast genome sequences, supporting the monophyly of the North American subgenus *Vitis* (Wen et al., 2018a). Although earlier studies hypothesized a divergence of subgenus *Vitis* between North America and Eurasia, the existing data only provides weak phylogenetic signal within the Eurasian clade, limiting our understanding of the evolutionary history of *Vitis* and preventing the exploration of valuable wild grape resources.

Some studies have also attempted to reconstruct the biogeographic history of subgenus *Vitis*. While Péros et al. (2011) suggested that subgenus *Vitis* had an Asian origin, Wan et al. (2013) supported a North American origin for the subgenus. Liu et al. (2016) provided additional support for a North American origin and further suggested that the subgenus originated during the Eocene and subsequently migrated to Eurasia via the North Atlantic land bridges (NALB) or intercontinental long distance dispersal (LDD). Despite the data available to date, several questions remain about the biogeographic history of subgenus *Vitis* as well as about its evolutionary adaptation to novel environments after their arrival in Eurasia. A well-resolved phylogeny is key to understanding the evolutionary history and patterns of species diversification of the subgenus.

Here, we perform whole-genome resequencing of 40 taxa in the genus *Vitis* using the genome of *V. vinifera* L. as reference (Jaillon et al., 2007). By comparing genomes across taxa, we reconstructed their phylogenetic relationships and explored the biogeographic history of the subgenus *Vitis* as a whole.

2. Materials and methods

2.1. Plant material and library preparation

A total of 39 accessions of subgenus *Vitis* [including two accessions of *V. davidii* (Rom. Caill.) Föex] and two accessions of subgenus *Muscadinia* were sampled from North America, Europe, and East Asia for genome resequencing (Supplementary Table A.1, Fig. A.1). Voucher specimens were deposited at the Key Laboratory of Horticultural Plant Biology (Ministry of Education), College of Horticulture and Forestry Science, Huazhong Agricultural University (China) and the US National Herbarium (US). The genomic DNA was extracted from leaves or seeds following a modified CTAB (cetyltrimethylammonium bromide) extraction procedure (Lodhi et al., 1994) and used as input material for the DNA sample preparation. Sequencing libraries were generated using Truseq Nano DNA HT Sample Preparation Kit (Illumina, USA) following the manufacturer's protocols. Index codes were added to attribute sequences to each sample. Briefly, each DNA sample was fragmented by sonication, then DNA fragments were end polished, A-tailed, and ligated with the full-length adaptor for Illumina sequencing with further PCR amplification. Finally, PCR products were purified (AMPure XP system) and libraries were analyzed for size distribution by Agilent2100 Bioanalyzer and quantified using real-time PCR.

2.2. Genome resequencing and quality control

Whole genomes of 41 accessions were sequenced on the Illumina Hiseq2500 platform. In total, we sequenced 472.53 Gb raw data with approximately 20× coverage per sample. Low quality paired-end reads were removed (reads with ≥10% unidentified nucleotides (N); > 10 nt aligned to the adaptor, allowing ≤10% mismatches; > 50% bases having phred quality < 5), as well as putative PCR duplicates generated in the library construction process, which mainly resulted from base-calling duplicates and adaptor contamination.

2.3. Read mapping

High quality paired-end short-insert reads were mapped to the reference *V. vinifera* genome (Jaillon et al., 2007) using the Burrows-Wheeler Aligner (BWA; Li and Durbin, 2009) software with the command 'mem -t 4 -k 32 –M'. BAM alignment files were then generated using the SAMtools package (Li et al., 2009). Alignment results were subsequently improved with the following steps: (i) alignment reads were filtered with mismatches ≤5 and mapping quality = 0; (ii) potential PCR duplication was removed (e.g., when multiple read pairs showed identical external coordinates, we only retained the pair with the highest mapping quality); (iii) reads around the indels were realigned by first identifying regions for realignment where at least one read contained an indel with a cluster of mismatching bases around it.

2.4. SNP calling

To identify high-credibility variation in the 41 accessions, we first used a Bayesian approach using the SAMtools package 0.1.19 (Li et al., 2009) to perform variant calling using the 'mpileup' program with the parameters '-C -D -S -m 2 -F 0.002 -d 1000' (Li et al., 2009). For downstream analysis, variants were strictly filtered by requiring a minimum coverage of 10 and a maximum coverage of 1000, a minimum RMS mapping quality of 20, and no gaps within a 3 bp window. Meanwhile, base quality scores were recalibrated using the Haplotype Caller based method of local de novo assembler and HMM likelihood function implemented in Genome Analysis Toolkit 3.1 (McKenna et al., 2010). This analysis provided empirically accurate