2016

Toward the Elucidation of Cytoplasmic Diversity in North American Grape Breeding Programs

Jonathan Fresnedo-Ramirez

Qi Sun

Chin-Feng Hwang

Craig A. Ledbetter

David W. Ramming

See next page for additional authors

Authors

Jonathan Fresnedo-Ramirez, Qi Sun, Chin-Feng Hwang, Craig A. Ledbetter, David W. Ramming, Anne Fennell, M. Andrew Walker, James J. Luby, Matthew D. Clark, Jason P. Londo, lance Cadle-Davidson, and Gan-Yuan Zhong

Mol Breeding (2016) 36:116
DOI 10.1007/s11032-016-0538-z

SHORT COMMUNICATION

Toward the elucidation of cytoplasmic diversity in North American grape breeding programs

Jonathan Fresnedo-Ramírez · Qi Sun · Chin-Feng Hwang · Craig A. Ledbetter · David W. Ramming · Anne Y. Fennell · M. Andrew Walker · James J. Luby · Matthew D. Clark · Jason P. Londo · Lance Cadle-Davidson · Gan-Yuan Zhong · Bruce I. Reisch

Received: 4 May 2016 / Accepted: 21 July 2016 / Published online: 1 August 2016

Abstract Plants have an intriguing tripartite genetic system: Nuclear genome × Mitochondria × Plastids and their interactions may impact germplasm breeding. In grapevine, the study of cytoplasmic genomes has been limited, and their role with respect to grapevine germplasm diversity has yet to be elucidated. In the present study, the results of an analysis of the cytoplasmic diversity among 6073 individuals (comprising cultivars, interspecific hybrids and segregating progenies) are presented. Genotyping by sequencing (GBS) was used to elucidate plastid and mitochondrial DNA sequences, and results were analyzed using multivariate techniques. Single nucleotide polymorphism (SNP) effects were annotated in reference to plastid and mitochondrial genome sequences. The cytoplasmic diversity identified was structured according to synthetic domestication groups (wine and raisin/table gr.ape types) and

J. Fresnedo-Ramírez (✉) · Q. Sun
BRC Bioinformatics Facility, Institute of Biotechnology, Cornell University, Ithaca, NY 14853, USA
e-mail: fresnedoramirez.1@osu.edu

Q. Sun
e-mail: qisun@cornell.edu

C.-F. Hwang
State Fruit Experiment Station at Mountain Grove Campus, Darr School of Agriculture, Missouri State University, Springfield, MO 65897, USA
e-mail: ChinFengHwang@missouristate.edu

C. A. Ledbetter · D. W. Ramming
USDA-ARS San Joaquin Valley Agricultural Sciences Center, Parlier, CA 93648, USA
e-mail: Craig.Ledbetter@ars.usda.gov

D. W. Ramming
e-mail: dwramming@yahoo.com

A. Y. Fennell
Department of Plant Science, South Dakota State University, Brookings, SD 57007, USA
e-mail: Anne.Fennell@sdstate.edu

M. A. Walker
Department of Viticulture and Enology, University of California, Davis, CA 95616, USA
e-mail: awalker@ucdavis.edu

J. J. Luby · M. D. Clark
Department of Horticultural Science, University of Minnesota, St. Paul, MN 55108, USA
e-mail: lubyx001@umn.edu

M. D. Clark
e-mail: clark776@umn.edu

J. P. Londo · L. Cadle-Davidson · G.-Y. Zhong
USDA-ARS Grape Genetics Research Unit, Geneva, NY 14456, USA
e-mail: jason.londo@ars.usda.gov

L. Cadle-Davidson
e-mail: lance.cadledavidson@ars.usda.gov

G.-Y. Zhong
e-mail: GanYuan.Zhong@ars.usda.gov

interspecific-hybridization-driven groups with introgression from North American *Vitis* species, identifying five cytoplasmic groups and four major clusters. Fifty-two SNP markers were used to describe the diversity of the germplasm. Ten organelle genes showed distinct SNP annotations and effect predictions, of which six were chloroplast-derived and three were mitochondrial genes, in addition to one mitochondrial SNP affecting a nonannotated open reading frame. The results suggest that the application of GBS will aid in the study of cytoplasmic genomes in grapevine, which will enable further studies on the role of cytoplasmic genomes in grapevine germplasm, and then allow the exploitation of these sources of diversity in breeding.

Keywords Plastid genome · Mitochondria genome · Cytoplasmic differentiation · Cytoplasmic lineage · Organelle genes · *Vitis* · Grapevine

Plants have two genetic systems that interact within their cells: a nuclear genome and a cytoplasmic component, which includes a mitochondrial and a plastid genome. The interaction of these three genetic complexes contributes to certain features that may play a role in traits of agricultural interest: the nuclear-cytoplasmic male sterility system in several crop species (Eckardt 2006); abiotic stress responses, such as cold tolerance in cucumber (Chung et al. 2007); and some agronomic traits in potato (Sanetomo and Gebhardt 2015). However, with the exception of cotton, studies on cytoplasmic genomes from a breeding perspective have been done primarily with annual or biennial plants. In grapevine, the maternally inherited cytoplasmic genomes, chloroplasts and mitochondria, have been studied (Goremykin et al. 2009), but not with attribution to specific traits. In the case of breeding germplasm, differentiation of full-sib families was performed using RFLP markers within the chloroplast genome (Strefeler et al. 1992), as well as maternal lineages between domesticated and Asian grapevine accessions (Lózsa et al. 2015). More recently, maternal inheritance has been suggested for the type of anthocyanin in the berries (Liang et al. 2009). The cytoplasmic genetic complex of crops is another layer of information that breeders and geneticists want to understand and exploit. The application of next generation sequencing (NGS) technologies in grapevine may enable study of this layer of information and make it useful for genetic research as well as breeding pursuits.

The present analysis included genotyping by sequencing (GBS) data from 6073 individual vines from the breeding programs of Cornell University (2768), South Dakota State University (682), University of Minnesota (716), Missouri State University (177), and USDA-San Joaquin Valley Agricultural Sciences Center (1666). The vines are used as germplasm for research and breeding in their respective programs as well as in the *Vitis*Gen project (www.vitisgen.org) funded by the USDA-Specialty Crops Research Initiative. Those 6073 individuals included biparental families having at least 20 individuals, some of them products of hybridization with various wild *Vitis* species. Also, 64 selections and cultivars used for wine, table grape, and raisin production, as well as germplasm development, were included in the analysis. The genetic background of the germplasm comprised a wide range of diversity, going from pure *Vitis vinifera* L. accessions (such as the progeny from a cross between 'Riesling' and 'Cabernet Sauvignon') to interspecific hybrids [e.g. Illinois-547-1 = *V. cinerea* (Engelm. ex A. Gray) Engelm. ex Millard accession B9 × *V. rupestris* Scheele accession B38] and complex hybrids (e.g. 'Seyval blanc' with background from *V. vinifera*, *V. rupestris*, and *V. aestivalis* Michx.).

GBS data were generated as described by Hyma et al. (2015), taking advantage of 384X multiplexing. In order to identify GBS markers in organelles, the genome sequences reported for chloroplast (Jansen et al. 2006) and mitochondria (Goremykin et al. 2009; Velasco et al. 2007) were included as independent chromosomes (20 and 21, respectively) in the FASTA

B. I. Reisch
Horticulture Section, School of Integrative Plant Science, Cornell University, Geneva, NY 14456, USA
e-mail: bruce.reisch@cornell.edu

Present Address:
D. W. Ramming
Ramming's Specialty Crops, Fresno, CA 93737, USA

Present Address:
J. Fresnedo-Ramírez
Department of Horticulture and Crop Science, The Ohio State University/OARDC, Wooster, OH 44691, USA

file containing the 'Pinot noir'-derived PN40024 (Jaillon et al. 2007) 12X version 2 assembly (Adam-Blondon et al. 2011). This file was processed and integrated in the TASSEL-GBS pipeline (Glaubitz et al. 2014) to call markers in both nuclear and organellar genomes. Subsequently, the markers were filtered using vcftools version 0.1.12a (Danecek et al. 2011) based on locations (organelle chromosomes), sequencing depth (>10), amount of missing data (<1 %) and presence of singletons and doubletons, which were excluded from the final marker set. An analysis based on the fixations index (F_{ST}, Weir and Cockerham 1984) per marker was performed considering biparental populations as items to look for differentiation. Thus, the identified markers were retrieved on an individual basis and used for principal component analyses based on marker covariance, as well as multidimensional scaling analyses using SNPRelate1.2.0 for R 3.2.1 (R Core Development Team 2016). Then, a cluster analysis of the markers with pools of individuals was performed using pvclust 2.0 (Suzuki and Shimodaira 2006) for R, using the Euclidean distance, the Ward's hierarchical clustering method, and the k-means method to distinguish clusters with bootstrapped $p \leq 0.05$ among clusters and $p \geq 0.05$ for individuals within a cluster. The outgroup was an accession of *Vitis acerifolia* Raf. (syn. *Vitis longii* W.R. Prince & Prince). As an additional measure of quality control, 177 vines resulting from reciprocal crosses between 'Cabernet Sauvignon' and 'Norton' (submitted by the Missouri State University breeding program, 19 siblings from the cross ♀ 'Cabernet Sauvignon' × ♂ 'Norton', and 158 from the reciprocal cross ♀ 'Norton' × ♂ 'Cabernet Sauvignon') were used to test maternal inheritance of markers.

Finally, the cytoplasmic DNA markers identified in the F_{ST} analysis were stored in a VCF file that subsequently was used as the input file for the SNP annotation and effect prediction according to the current annotations available using SnpEff 4.1 (Cingolani et al. 2012). Given that the annotations for the grapevine chloroplast and mitochondria are not available either in the current genomic annotation of grapevine or in the built-in databases of SnpEff, custom annotation databases were generated for each organelle, following the manual of SnpEff 4.1 and using the existing information available in GenBank (accessions: DQ424856.1 and FM179380.1, for chloroplast and mitochondria, respectively). The putative impact of SNP effects yielded for the functional annotation procedure by SNPs through SnpEff 4.1 is described in the 'Variant annotation in VCF format' document available at the SnpEff website (Cingolani et al. 2015).

The NCBI-GenBank accession for the chloroplast genome sequence refers to a cultivar called 'Maxxa', which is not registered in the *Vitis* International Variety Catalogue (http://www.vivc.de). Therefore, in communication with Jansen et al. (2006), it was clarified that the 'Maxxa' chloroplast genome sequence corresponds to an accession of the cultivar 'Syrah', which was derived from the BAC libraries used in the study of Tomkins et al. (2001) (Christopher Saski, Pers. Comm.). Up to the date of submission of this manuscript, the information in GenBank has not been corrected.

There were 22 SNPs identified in the chloroplast and 30 SNPs identified in the mitochondria sequences. Thus, 52 GBS markers were analyzed to examine the cytoplasmic diversity of breeding progenies in the *Vitis*Gen database. The maternal inheritance of the 52 selected markers was confirmed based on all markers matching the maternal parent in all 177 siblings from the reciprocal cross between 'Cabernet Sauvignon' and 'Norton'. According to the principal component analysis results and considering the 52 GBS cytoplasmic markers, 81.7 % of the variation is captured by the first two principal components (62.4 % by PC1 and 19.3 % by PC2). Considering chloroplast polymorphisms, 34.6 % of the variation was captured by the first two principal components. Considering mitochondrial polymorphisms, 82.4 % was captured through the first two principal components. Given these results, a graphical three-dimensional representation through a multidimensional scaling plot was generated (Fig. 1) in which each axis captured 33.3 % of the variation, which allowed the visualization of cytoplasmic relationships and variability among the breeding accessions. Genotypes studied sorted into four main clusters: cluster A corresponded to genotypes related to *V. rupestris*; cluster B encompassed genotypes related to *V. riparia* and *V. aestivalis*; and clusters C and D included genotypes closer to *V. vinifera*; however, they may each be from distinct cytoplasmic lineages.

The grouping of the breeding progenies according to their cytoplasmic constitution resulted in six cytoplasmic groups (Fig. 1), which are: (1) a group

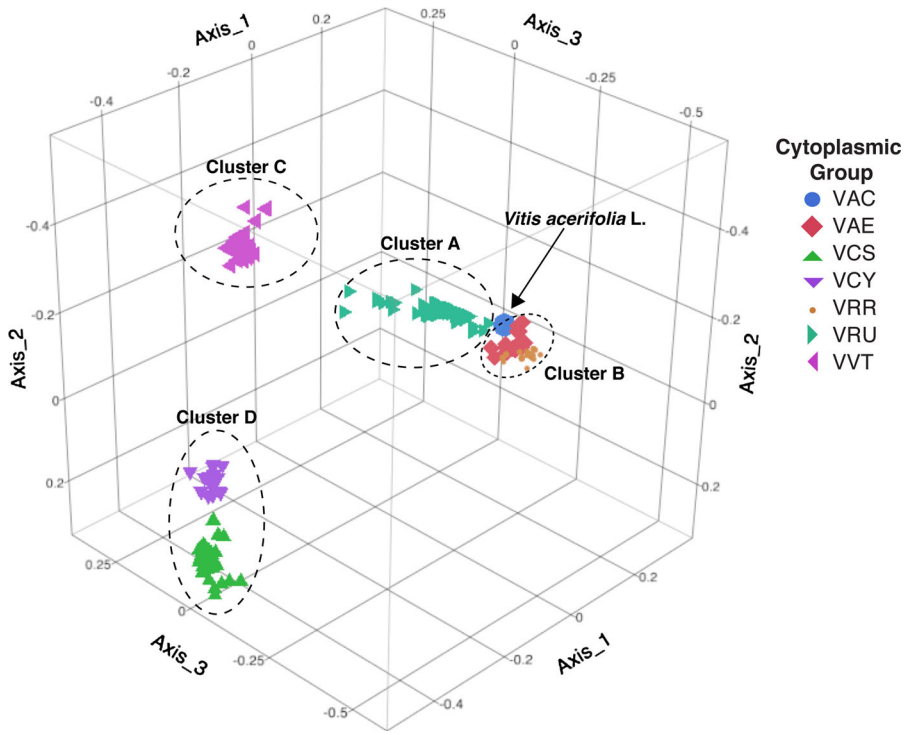

Fig. 1 Tridimensional scaling plot of the 6073 accessions of grapevine analyzed for their cytoplasmic diversity, based on 52 plastid and mitochondrial SNP markers. Four clusters were identified as significantly distinct (encircled by dotted lines), while five cytoplasmic groups plus a distinct outgroup were distinguished: VAC corresponds to the accession of *V. acerifolia*. VAE corresponds to the group of accessions related to *V. aestivalis*, VCS includes accessions of *V. vinifera* related to 'Cabernet Sauvignon', VCY, includes accession of the other lineage of *V. vinifera* in this study, and which are related to 'Chardonnay', VRR includes accessions related to *V. riparia* and *V. rupestris*, VRU includes accessions related to another large lineage of *V. rupestris*, and VVT includes accessions related to *V. vinifera* but mainly individuals bred for table grape and raisin production

related to *V. rupestris* (VRU) composed of 2044 individuals, including known cultivars such as 'Chambourcin' and 'Cayuga White'; (2) a group with *V. vinifera* origin and related to 'Cabernet Sauvignon' (VCS) composed of 1330 individuals, which included known cultivars such as 'Marquette', 'Flame Seedless', 'Regent', 'Riesling' and 'Ruby Cabernet'; (3) a group encompassing table grape and raisin accessions (VVT) composed of 862 individuals, including known cultivars such as 'Scarlet Royal' and 'Jupiter'; (4) a group related to *Vitis riparia* Michx. and *V. rupestris* (VRR) composed of 835 individuals, encompassing cultivars such as 'Frontenac' and 'Valvin Muscat' (and this group may be related to group 1 as an additional lineage); (5) a group related to *V. aestivalis* (VAE) composed of 643 individuals, including accessions related to 'Tamiami' and 'Norton'; and (6) a group with *V. vinifera* (VCY) origin composed of 319 individuals and related to 'Chardonnay', which also included table grape cultivars such as 'Autumn King' and raisin cultivars such as 'Sunpreme'. *V. acerifolia* (VAC) was identified as an isolated individual between groups 1 and 5. Finally, 40 individuals presented ambiguous genotypes, due to missing data or suspicious heterozygosity, particularly for the mitochondrial markers, and were excluded from the study. These individuals included known cultivars such as 'La Crescent', 'Horizon', 'Rubired', and 'Seyval blanc'.

Although it has been proposed that at the genomic level close kinship between *V. vinifera* grapevine genotypes for wine and raisin/table grapes is scarce (Myles et al. 2011), our results suggest that *V. vinifera* germplasm may share a cytoplasmic lineage regardless of domestication events. Thus, in terms of cytoplasmic genes affected by SNPs, the annotation

of the polymorphisms identified in this study shows that up to 10 organelle genes (Table 1, nine annotated genes and one open reading frame) may be affected to distinct degrees, from high impact (stop codon) to low impact and modifiers (a variant that is of low impact and lacks deleterious effects, Cingolani et al. 2015). Six plastid genes were found with SNP effects (*atpA*, *rcpoC2*, *psbD*, *psbB*, *rps12* and *ndhE*) (Table 1). A high-impact effect (i.e. a new stop codon) was noted for the gene *psbB* for Photosystem II P680 chlorophyll A apoprotein (Westhoff et al. 1983), with a transition from G → A occurring at position 77,699 bp. This protein is related to the use of energy from light to extract electrons from water, to generate oxygen and a proton gradient for the subsequent formation of ATP. Interestingly, nine individuals showed this transition: five selections from the Cornell breeding program, two from Missouri State University, and two from University of Minnesota. In the latter case, this transition was noted in the table grape cultivar 'Swenson Red', which descends from *V. riparia* and *V. labrusca* L. However, the maternal background seems to come from *V. labrusca* given that the female grandparent 'Witt' is reported to result from self-pollination of 'Concord', having *V. labrusca* maternal ancestry (Hedrick et al., 1908). Unfortunately, in the set of breeding selections analyzed, germplasm with maternal descent from *V. labrusca* was not considered. Therefore, our results suggest that the addition of *labrusca*-related genotypes will help to further describe cytoplasmic diversity within *Vitis* breeding programs.

The role of chloroplasts in plant response to environmental stimuli is supported by the relevance of this organelle for photosynthesis and metabolism; in addition, evidence supports the role of this organelle for environmental sensing to trigger adaptive responses due to abiotic and biotic stresses, and to control the aging processes (Bouvier et al. 2009). Thus, in crops such as tomato, differential transcription of the gene *psbB* has been linked to responses to salinity–alkalinity stress (Li et al. 2015). However, there are no studies addressing the role of certain chloroplast polymorphisms to adaptive responses in grapevine. In the present study, it is interesting that 'Swenson Red' showed a particular high-impact SNP effect, since this table grape cultivar is considered to be moderately cold-hardy, and the gene *psbB* is a member of a complex gene cluster for land plant chloroplasts (*psbB-psbT-psbH-petB-petD* operon, Barkan 2011), which may be involved in the response to cold stress (Kupsch et al. 2012). Further research is needed, which may involve the manipulation of plastid genes.

Three mitochondrial genes (*nad1*, *nad2* and *nad5*) and one mitochondrial open reading frame (*orf104*) were found with SNP effects (Table 1). The great majority of effects (33 out of 60) were located in intron regions, which could affect more than one gene. Some of these SNP effects may affect splicing, giving rise to multiple protein isoforms. Only one variant, located at 581,741 bp, showed either moderate or low impact for *orf104* and *nad5*, respectively. When the change was a transversion (from A → C), it had a moderate impact in the leading strand on *orf104*, which should be related to proteins reported in tobacco and maize, but is still an uncharacterized protein (Goremykin et al. 2009). When the change was a transversion (from A → T), it had a low impact in the lagging strand on *nad5*, which in its gene ontology is related to ATP synthesis. Interestingly, 884 individuals, mainly from the group VCY showed the transversion affecting *orf104,* while the transversion affecting *nad5* was found in the great majority of the individuals, independent of the cluster identified previously.

Our study provides a first approach to the study of mitochondrial genome diversity within breeding selections of *Vitis*. Nevertheless, the germplasm diversity captured exclusively by mitochondrial polymorphisms is remarkable, considering that 82.4 % of the variation was captured by the first two principal components, and the characterization and use of this diversity may follow several paths. The evolution and colonization of species may be studied in wild *Vitis* populations, as has been done for pines (Wu et al. 1998). For purposes of germplasm conservation and breeding, mitochondrial markers may allow a more robust determination of the maternal backgrounds of not only cultivars (Arroyo-García et al. 2002) but also interspecific hybrid accessions within *Vitis* germplasm collections, especially when parentage is uncertain or untraceable. With the availability of GBS data on germplasm accessions and the development of analytical tools such as AmpSeq (Yang et al. 2016), the elucidation of such cytoplasmic lineages seems attainable, as well as the accurate determination of cytoplasmic haplotypes in individual, family and breeding program bases. The grapevine mitochondrial genome is one of the largest in dicots (Goremykin

Table 1 Cytoplasmic polymorphisms identified in this study, as well as SNP effects on plastid and mitochondrial genes

Locus[a,b]	Reference[c]	Alternative(s)[d]	Effect(s)[e]	Impact(s)[f]	Gene(s)[g]	Gene(s) names
S20_11990	G	T, A	Missense variant	Moderate	*atpA*	ATP synthase CF1 alpha chain
S20_11992	T	C, G	Missense variant, Synonymous variant	Moderate, Low	*atpA*	
S20_12036	T	G, C	Missense variant	Moderate	*atpA*	
S20_12037	C	G, T	Missense variant	Moderate, Low	*atpA*	
S20_12038	T	G, C	Missense variant	Moderate, Low	*atpA*	
S20_12044	G	C, A	Missense variant	Moderate	*atpA*	
S20_18436	A	T	Synonymous variant	Low	*rpoC2*	RNA polymerase beta
S20_36731	C	A, G	Missense variant	Moderate	*psbD*	Photosystem II protein D2
S20_36751	T	G, C	Synonymous variant	Low	*psbD*	
S20_36759	T	C, G	Missense variant	Moderate	*psbD*	
S20_36762	A	G	Missense variant	Moderate	*psbD*	
S20_36770	T	G	Missense variant	Moderate	*psbD*	
S20_36773	T	C	Missense variant	Moderate	*psbD*	
S20_36777	A	G	Missense variant	Moderate	*psbD*	
S20_36784	T	C, G	Synonymous variant	Low	*psbD*	
S20_77694	T	G, C	Synonymous variant, Intron variant	Low, Modifier	*psbB*, *rps12*	Photosystem II P680 chlorophyll A apoprotein, Ribosomal protein S12
S20_77699	G	A, T	Stop gained, Missense variant	High, Moderate	*psbB*, *rps12*	
S20_77708	T	G, C	Missense variant, Intron variant	Moderate, Modifier	*psbB*, *rps12*	
S20_77918	G	A, T	Missense variant, Intron variant	Moderate, Modifier	*psbB*, *rps12*	
S20_77925	A	G, T	Missense variant, Synonymous variant	Moderate, Low	*psbB*, *rps12*	
S20_123934	T	C, A	Missense variant, Intron variant	Moderate, Modifier	*ndhE*, *rps12*	NADH dehydrogenase subunit 4L, Ribosomal protein S12, Ribosomal protein S12
S20_123960	A	G, C	Missense variant, Synonymous variant	Moderate, Low	*ndhE*, *rps12*	
S21_129983	C	A, G	Intron variant	Modifier	*nad5*	NADH dehydrogenase subunit 5
S21_130007	G	A	Intron variant	Modifier	*nad5*	
S21_130037	G	A, T	Intron variant	Modifier	*nad5*	
S21_130340	A	T, C	Intron variant	Modifier	*nad5*	
S21_151790	A	C	Intron variant	Modifier	*nad5*	
S21_151813	A	C, G	Intron variant	Modifier	*nad5*	
S21_151815	A	C	Intron variant	Modifier	*nad5*	
S21_156980	A	T, C	Intron variant	Modifier	*nad5*	
S21_156981	A	T, G	Intron variant	Modifier	*nad5*	
S21_156982	G	T, C	Intron variant	Modifier	*nad5*	
S21_170740	A	C	Intron variant	Modifier	*nad5*, *nad1*	NADH dehydrogenase subunits 5 and 1

Table 1 continued

Locus[a,b]	Reference[c]	Alternative(s)[d]	Effect(s)[e]	Impact(s)[f]	Gene(s)[g]	Gene(s) names
S21_258198	T	A, G	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	NADH dehydrogenase subunits 5, 1 and 2
S21_258213	C	A, G	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_258214	G	T, A	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_317016	T	G, A	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_327855	A	C	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_342889	T	A, C	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_355591	A	C	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_363763	A	T	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_367092	C	T, G	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_368888	G	T, C	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_427497	G	A	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_495021	T	A, C	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_495022	T	A	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_495023	C	T, G	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_495024	T	A, G	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_518882	G	T	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_571261	C	T	Intron variant	Modifier	*nad5*, *nad1*, *nad2*	
S21_581741	A	C, T	Missense variant, Synonymous variant, Synonymous variant	Moderate, Low	*orf104*, *nad5*, *nad1*	Protein YP_002608359.1 similar to orf103a in Nicotiana and orf105-c in Zea, NADH dehydrogenase subunits 5 and 1
S21_616454	C	A, T	Intron variant	Modifier	*nad1*, *nad2*	NADH dehydrogenase subunits 1 and 2

[a] Locus names starting with S20 denote plastid SNPs while locus names starting with S21 denote mitochondrial SNPs

[b] Numbers after the underscore denote the physical position on the organelle genome sequence

[c] Reference denotes the nucleotide allele in the reference

[d] Alternative denotes the nucleotide allele(s) identified as polymorphisms; when more than one alternative was identified these are separated by commas (the same pattern is follow in subsequent columns)

[e] Effects denotes the type of effect on the sequence given by the presence of the SNP

[f] Impact denotes the impact of the effect on the sequence as described by Cingolani et al. (2015)

[g] Gene name abbreviation

et al. 2009; Mower et al. 2012); however, the mechanisms of recombination are not yet clear (Marechal and Brisson 2010). So, this source of variation is mostly unexplored, although methods for its study have been developed (Picardi et al. 2010).

Even though empirical experience suggests that grapevine breeders are actively selecting for traits that may be influenced by cytoplasmic genomes during genetic improvement and cultivar development, the actual extent is unknown. For instance, one may expect that certain berry skin hues might be chosen over others due to a particular pattern of oxidation of certain types of anthocyanin (Janeiro and Brett 2007). These pigments might be influenced by certain cytoplasmic features leading a maternal inheritance, as suggested by Liang et al. (2009). However, this hypothesis has yet to be experimentally addressed. Similarly, breeders are sometimes using wild species as female parents, since the practice suggests that by this route abiotic stress resistance may be improved, implying that organelles may have special relevance. Certainly, analyses of the inheritance of such complex traits with an emphasis on maternal inheritance have not been accomplished, and experiments to address these kinds of questions would require efforts and funding that are not usually available within grapevine breeding programs budgets. Evidence from other species suggests that some cytoplasmic-influenced traits have experienced active anthropogenic selection. For instance, in *Arabidopsis thaliana* L., the generation of mitochondrial-derived reactive oxygen species; transition of mitochondria morphology; depolarization of mitochondrial membrane potential; and modulation of mitochondrial respiratory genes are involved in certain signals for the salicylic acid pathway (Nie et al. 2015). This metabolic pathway is pivotal in plant defense responses and then in traits of high breeding relevance, which are readily identified in either commercial germplasm, or among accessions deposited in repositories.

On the other hand, several currently neglected research topics in *Vitis* are now accessible with the use of NGS techniques to better understand grapevine biology. For instance, nothing is known in *Vitis* about the nuclear integrants of organelle DNA (*norgs*, Rousseau-Gueutin et al. 2011) whether from plastids (*nupts*, Timmis et al. 2004) or mitochondria (*numts*, Lopez et al. 1994), and their relevance to grapevine biology, genetics, agroecology, and breeding. In the present analysis, we used sequences of plastid and mitochondria as independent autosomes with the purpose of discovering markers through the alignment with those sequences. Some of those alignments may also come from certain nuclear sequences representing actual *norgs*, since Goremykin et al. (2009) have described horizontal gene transfer in grapevine. The identification and characterization of these norgs are possible (Rousseau-Gueutin et al. 2011) and attainable using techniques of NGS amplicon sequencing.

Another pending topic is the role of heteroplasmy [coexistence in a cell of distinct mitochondrial genomes—and possibly plastid genomes (Woloszynka 2010)] in the diversity of mitochondria in grapevine. Plant mitochondria, unlike animal mitochondria, can recombine and are a very active genomic system (Sloan 2013). Phenomena such as horizontal gene transfer (Goremykin et al. 2009); coexistence of alternative mitotypes; and segregation of these alternative mitotypes during plant development (Woloszynka 2010) enrich the genetic makeup of plant species. Segregation of mitotypes at distinct developmental stages and within specific organs may lead to chimerism, which is an important source of phenotypic variation in grapevine, and often not well understood. The relevance of mitochondrial and plastid diversity also has implications to describe the divergence and diversity of wild relatives of crops, as seen for carrot (Mandel et al. 2012; Mandel and McCauley 2015). In the case of grapevine, the description and understanding of such diversity could inform North American breeding programs given the active use of wild species and complex interspecific hybrids. In this study, we tried to avoid the "noise" possibly produced by heteroplasmy by not accounting for singletons and doubletons, although these features may be indicative of heteroplasmy and need further attention.

In this study, we characterized cytoplasmic diversity in a large sampling of North American and *V. vinifera* genotypes. Cytoplasmic diversity was structured according to synthetic domestication groups (wine and raisin/table grape types) and interspecific-hybridization groups (breeding selections with introgression from related species). Although resources are available for the study of cytoplasmic diversity, such resources may be greatly improved with the application of NGS. In addition, further studies of the role of cytoplasmic genomes in specific traits relevant to

grapevine improvement are desirable (i.e. reciprocal crosses). Consideration of the cytoplasmic genome sequences for any grapevine genome sequence assembly is desirable to continue the further exploration of the role of cytoplasmic genomes in viticulture.

Acknowledgments The authors acknowledge the U.S. Department of Agriculture, National Institute of Food and Agriculture, Specialty Crop Research Initiative (Award no. 2011-51181-30635), and the National Grape and Wine Initiative for funding for the *Vitis*Gen project (http://www.vitisgen.org/). We also thank Shanshan Yang for discussion and suggestions.

Compliance with ethical standards

Conflict of interest The authors declare that there are no conflicts of interest.

References

Adam-Blondon A, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R (2011) Genome sequence initiatives. In: Adam-Blondon A-Fo, Martínez-Zapater JM, Kole C (eds) Genetics, genomics and breeding of crop plants. Science Publishers, Enfield, pp 211–234

Arroyo-García R, Lefort F, de Andrés MT, Ibañez J, Borrego J, Jouve N, Cabello F, Martínez-Zapater JM (2002) Chloroplast microsatellite polymorphisms in Vitis species. Genome 45(6):1142–1149

Barkan A (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol 155(4):1520–1532

Bouvier F, Mialoundama AS, Camara B (2009) A sentinel role for plastids. In: Sandelius A, Aronsson H (eds) The chloroplast, vol 13. Plant cell monographs. Springer Berlin, pp 267–292

Chung S-M, Gordon VS, Staub JE (2007) Sequencing cucumber (Cucumis sativus L.) chloroplast genomes identifies differences between chilling-tolerant and -susceptible cucumber lines. Genome 50(2):215–225

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu XY, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w(1118); iso-2; iso-3. Fly 6(2):80–92

Cingolani P, Cunningham F, McLaren W, Wang K (2015) Variant annotations in VCF format version 1.0. Available on-line http://snpeff.sourceforge.net/VCFannotationformat_v1.0.pdf

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158

Eckardt NA (2006) Cytoplasmic male sterility and fertility restoration. Plant Cell 18(3):515–517

Glaubitz JC, Casstevens TM, Lu F, Harriman J, Elshire RJ, Sun Q, Buckler ES (2014) TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9(2):e90346

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of Vitis vinifera and the issue of rampant horizontal gene transfer. Mol Biol Evol 26(1):99–110

Hedrick UP, Booth NO, Dorsey MJ, Taylor OM, Wellington R (1908) The grapes of New York, Report of the New York Agricultural Experiment Station for the year 1907. JB Lyon Company, Albany

Hyma KE, Barba P, Wang M, Londo JP, Acharya CB, Mitchell SE, Sun Q, Reisch B, Cadle-Davidson L (2015) Heterozygous Mapping Strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case study in grapevine. PLoS ONE 10(8):e0134880

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian Public Consortium for Grapevine Genome C (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449(7161):463–467

Janeiro P, Brett AMO (2007) Redox behavior of anthocyanins present in Vitis vinifera L. Electroanal 19(17):1779–1786

Jansen RK, Kaittanis C, Saski C, Lee SB, Tomkins J, Alverson AJ, Daniell H (2006) Phylogenetic analyses of Vitis (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and phylogenetic methods on resolving relationships among rosids. BMC Evol Biol 6:32. doi:10.1186/1471-2148-6-32

Kupsch C, Ruwe H, Gusewski S, Tillich M, Small I, Schmitz-Linneweber C (2012) Arabidopsis chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. Plant Cell 24(10):4266–4280

Li J, Hu L, Zhang L, Pan X, Hu X (2015) Exogenous spermidine is enhancing tomato tolerance to salinity–alkalinity stress by regulating chloroplast antioxidant system and chlorophyll metabolism. BMC Plant Biol 15(1):1–17

Liang Z, Yang C, Yang J, Wu B, Wang L, Cheng J, Li S (2009) Inheritance of anthocyanins in berries of Vitis vinifera grapes. Euphytica 167(1):113–125

Lopez JV, Yuhki N, Masuda R, Modi W, Obrien SJ (1994) Numt, a recent transfer and tandem amplification of mitochondrial-DNA to the nuclear genome of the domestic cat. J Mol Evol 39(2):174–190

Lózsa R, Xia N, Deak T, Bisztray GD (2015) Chloroplast diversity indicates two independent maternal lineages in cultivated grapevine (Vitis vinifera L. subsp vinifera). Genet Resour Crop Evol 62(3):419–429. doi:10.1007/s10722-014-0169-3

Mandel JR, McCauley DE (2015) Pervasive mitochondrial sequence heteroplasmy in natural populations of wild carrot, Daucus carota spp. carota L. PLoS ONE 10(8):e0136303

Mandel JR, McAssey EV, Roland KM, McCauley DE (2012) Mitochondrial gene diversity associated with the atp9 stop codon in natural populations of wild carrot (Daucus carota ssp. carota). J Hered 103(3):418–425

Marechal A, Brisson N (2010) Recombination and the maintenance of plant organelle genome stability. New Phytol 186(2):299–317

Mower J, Sloan D, Alverson A (2012) Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ (eds) Plant genome diversity, vol 1. Springer, Vienna, pp 123–144

Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES (2011) Genetic structure and domestication history of the grape. Proc Natl Acad Sci U S A 108(9):3530–3535

Nie S, Yue H, Zhou J, Xing D (2015) Mitochondrial-derived reactive oxygen species play a vital role in the salicylic acid signaling pathway in Arabidopsis thaliana. PLoS ONE 10(3):e0119853

Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G (2010) Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. Nucleic Acids Res 38(14):4755–4767

R development core team (2016) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rousseau-Gueutin M, Ayliffe MA, Timmis JN (2011) Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol 157(4):2181–2193

Sanetomo R, Gebhardt C (2015) Cytoplasmic genome types of European potatoes and their effects on complex agronomic traits. BMC Plant Biol 15:162

Sloan DB (2013) One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol 200(4):978–985

Strefeler MS, Weeden NF, Reisch BI (1992) Inheritance of chloroplast DNA in two full-sib Vitis populations. Vitis 31(4):183–187

Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics 22(12):1540–1542

Timmis JN, Ayliffe MA, Huang CY, Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5(2):123–135

Tomkins JP, Peterson DG, Yang TJ, Main D, Ablett EF, Henry RJ, Lee LS, Holton TA, Waters D, Wing RA (2001) Grape (Vitis vinifera L.) BAC library construction, preliminary STC analysis, and identification of clones associated with flavonoid and stilbene biosynthesis. Am J Enol Viticult 52(4):287–291

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Dematte L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R (2007) A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE 2(12):e1326

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358–1370

Westhoff P, Alt J, Herrmann RG (1983) Localization of the genes for the two chlorophyll a-conjugated polypeptides (mol. wt. 51 and 44 kd) of the photosystem II reaction center on the spinach plastid chromosome. EMBO J 2(12):2229–2237

Woloszynska M (2010) Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. J Exp Bot 61(3):657–671

Wu J, Krutovskii KV, Strauss SH (1998) Abundant mitochondrial genome diversity, population differentiation and convergent evolution in pines. Genetics 150(4):1605–1614

Yang S, Fresnedo-Ramírez J, Wang M, Cote L, Schweitzer P, Barba P, Takacs EM, Clark MD, Luby JJ, Manns DC, Sacks GL, Mansfield AK, Londo JP, Fennell AY, Gadoury D, Reisch BI, Cadle-Davidson LE, Sun Q (2016) A next-generation marker genotyping platform (AmpSeq) in heterozygous crops: a case study for marker assisted selection in grapevine. Horticult Res 3:16002