Genetic and protein sequence variation of *CBF1-4* in cold hardy wild grapevine germplasm

J. P. Londo[1), 2)] and A. J. Garris[3)]

[1)] United States Department of Agriculture, Agricultural Research Service, Grape Genetics Research Unit, Geneva, USA
[2)] Cornell University, College of Agricultural and Life Sciences, Department of Horticulture, Geneva, USA
[3)] Communications Department, College of Agriculture and Life Science, Cornell University, Ithaca, USA

Summary

Grapevine (*Vitis vinifera*) production is limited by a wide array of abiotic stresses including cold, drought, and salinity stresses. The regulation of stress response genes under these conditions is very complex but a common and essential component of the gene network is the *C-Repeat Binding Factor (CBF)* genes. Four unique *CBF* genes have been described in grapevine to date and gene or protein level variation may help explain differences in stress resistance phenotypes in grapevine. This study was conducted to evaluate the level of genetic diversity found at these genes in different wild grapevine species as well as in a panel of cultivated grapevine varieties. Examination of the genetic variation at the four *CBF* genes in grapevine revealed high sequence variation with many different alleles in wild and cultivated grape. However, predicted protein sequence variation of the different genes revealed a different pattern. *Vitis CBF* genes 1, 2, and 3, which play roles in the stress response to many different stresses, were seen to have high levels of protein sequence variation. In contrast, *VCBF4*, which is induced by cold, was seen to have very little variation across all wild and cultivated grapevine samples. This contrasting pattern between different gene family members suggests an essential role of *VCBF4* in grapevine as the protein sequence is highly conserved between wild North American grapevine species and cultivated varieties.

K e y   w o r d s : C-Repeat Binding Factor, *VCBF4*, wild grapevine, cold stress resistance, genetic variation.

Introduction

CBF genes occupy a central role in the best known example of a cold regulated gene cascade in plants. *CBF* genes have been identified in all plants to date, although copy number and function vary by species. The *CBF* genes were first described in Arabidopsis, where 4 different gene family members have been identified. *AtCBF*1, 2 and 3 are each triggered by exposure to cold conditions, but the timing and tissue expression levels of these genes are unique (Novillo *et al.* 2007). *AtCBF*4 shares the characteristic AP2 binding regions present in *AtCBF*1, 2, and 3 but has been shown to be upregulated by exposure to drought stress and ABA treatments instead of cold (Haake *et al.* 2002). The nomenclature for this gene family is reversed in *Vitis* relative to Arabidopsis. *Vitis(V)CBF* 1, 2, and 3 are up-regulated following exposure to dehydration, salt, and chilling stresses, whereas *VCBF* 4 is up-regulated only after chilling (Xiao *et al.* 2008). *VCBF4* is the most similar *CBF* gene when compared with Arabidopsis and has the highest dissimilarity compared to the other *Vitis CBF* genes. Previous studies which examined the differences in gene expression between *VCBF4* in *V. vinifera* (cold sensitive) and *V. riparia* (cold tolerant) did not detect any differences despite different phenotypes (Xiao *et al.* 2008). However, this study examined the sequences and expression of just one tolerant and one sensitive genotype. A broader survey of genetic variation at the *VCBF* loci may reveal polymorphisms or amino acid substitutions that contribute to differences in cold hardiness. The objective of this study was to quantify the levels of genetic and protein sequence variation for the *CBF* gene family within and between different species of the genus *Vitis*.

Methods

The USDA maintains a living germplasm collection of wild *Vitis* species in Geneva, New York. From this collection, we sampled leaf material for DNA extraction and sequencing of 91 different accessions (Table). Samples included three *V. aestivalis*, five *V. cinerea*, 17 *V. labrusca*, three *V. palmata/rubra*, nine *V. rupestris*, and 41 *V. riparia*. These genotypes represent examples of cold hardy germplasm. Because *V. riparia* is frequently used in grapevine breeding for cold hardiness and is better represented in the collection, *V. riparia* sample numbers exceed other species in this study. Twelve cultivars of *V. vinifera* collected from a local vineyard (Ventosa Vineyards) were also included to represent cold-sensitive genotypes. The survival of *V. vinifera* in the Finger Lakes wine region results from milder environmental conditions due to proximity to large lakes. The geographic region from which accessions were originally collected is included in the Table. DNA was extracted using Qiagen 96-well plant mini kits (Qiagen, Hilden, Germany). Following extraction, PCR reactions were run for all samples and all primer pairs. We amplified and sequenced the entire coding region of each of the *Vitis CBF* genes. Primers for PCR amplification and sequencing for the 4 genes are as follows: ***VCBF1***, Forward (TTTAGCTGGGCATC-

Table

GRIN			GRIN			GRIN			Abbreviations	
Vitis Species	Accession	Location	*Vitis* Species	Accession	Location	*Vitis* Species	Accession	Location		
Aestivalis	PI_483137	PA	*Riparia*	PI_255189	NY	*Riparia*	PI_588439	MO	PA	Pennsylvania
	PI_483185	NJ		PI_483165	PA		PI_588440	KS	NJ	New Jersey
	PI_588677	MO		PI_483166	NY		PI_588455	MO	MO	Missouri
				PI_483167	VT		PI_588456	KS	IL	Illinois
Cinerea	PI_588134	IL		PI_483168	VT		PI_588457	IL	TX	Texas
	PI_588210	TX		PI_483169	VT		PI_588562	WI	MA	Massachusetts
	PI_588222	MO		PI_483171	NY		PI_588653	IA	NY	New York
	PI_588352	MO		PI_483172	NY		PI_588711	MAN	CT	Connecticut
	PI_588678	MO		PI_483175	NH		PI_594344	WI	VT	Vermont
	PI_597232	IL		PI_483177	NY					NE	Nebraska
				PI_483178	PA	*Rupestris*	PI_588146	MO	IA	Iowa
Labrusca	PI_483130	MA		PI_483182	NJ		PI_588147	AR	MN	Minnesota
	PI_483133	NH		PI_495622	NE		PI_588160	TX	ND	North Dakota
	PI_483145	PA		PI_588054	IA		PI_588224	AR	MT	Montana
	PI_483147	PA		PI_588167	QUE		PI_588225	OK	CO	Colorado
	PI_483150	NH		PI_588258	MN		PI_588231	TX	KS	Kansas
	PI_483151	NH		PI_588259	MAN		PI_588330	MO	WI	Wisconsin
	PI_483156	NH		PI_588260	MAN		PI_588333	MO	AR	Arkansas
	PI_483157	NH		PI_588261	MN		PI_588355	MO	OK	Oklahoma
	PI_483159	NH		PI_588262	MN					NH	New Hampshire
	PI_483162	NH		PI_588271	ND	*Vinifera*	Cabernet Franc		BC	British Columbia (CA)
	PI_483163	NY		PI_588275	ND		Cabernet Sauvignon		MAN	Manitoba (CA)
	PI_588307	NJ		PI_588344	ND		Chardonnay		QUE	Quebec (CA)
	PI_588583	CT		PI_588345	QUE		Lemberger			
	PI_588647	CT		PI_588346	MT		Merlot			
	PI_588675	BC		PI_588347	IL		Noiret			
	PI_588182	PA		PI_588349	ND		Pinot Grigio			
	PI_588145	MO		PI_588350	MT		Pinot Noir			
				PI_588353	ND		Riesling			
Palmata	PI_588155	IL		PI_588435	CO		Sangiovese			
	PI_588152	MO		PI_588437	CO		Syrah			
	PI_588233	IL		PI_588438	IL		Friulano			

CGTGTC; TTCGCGGACAAAATTCTATGCTTC), Reverse (TCCAATAAGCTTGGTGGGTTA; TTATAAGCATAGTTGA; GTCAAACTAATCATCA; CCACCGGAGAAACTTCCAGA); ***VCBF2***, Forward (GTGTCCAATTCGCGGACAAT; CATCATCGTCATCTTC), Reverse (GAATGGGTTCGAGTGGCAAA; CAACTGAAGCCTTCAC; CTTCCTCGACCCTGCTGACA; ATCCTGGACTTCTTAC); ***VCBF3***, (CCCTGACCGCTTCTCGATCT; GCCGCCCTAGCATTCAGAG; CAATTTCCCCGACTCC), Reverse (CTGATGACAGCGGCAAGGAT; GGTTGTGGGAGTTGCATTGG); and ***VCBF4***, Forward (CAGAACCGTGTTTGGGAACC; GGATATGGCTGGGGACGTTT), Reverse (CCACCGGATTTACCTCCCAC, CCGCCTTCTGAATGTCCTTG).

Sequence trace files were evaluated for quality using the alignment software *Sequencher* (Gene Codes Corporation, Ann Arbor, MI) and polymorphic sites were called manually for all sequences using standard ambiguity codes. Haplotypes were phased using DNaSP 5.0 software (Librado and Rozas 2009). Tests of selection/neutrality (Tajima's D, Fu's F) were performed using Arlequin (Excoffier and Lischer 2010). Coding regions were extracted from the total sequenced lengths of the four genes and translated into protein sequences using DNaSP 5.0. Allelic variation was separated for each species tested to determine both the total number of protein alleles as well as the number of species-specific alleles. In addition to the samples we sequenced, *VCBF* alleles were mined from NCBI-GenBank (http:// www.ncbi.nlm.nih.gov/) to identify novel alleles present in this germplasm. After predicting amino acid changes, the program SNAP (Bromberg and Rost 2007) was used to infer the potential neutrality or non-neutrality of amino acid substitutions. The program SplitsTree (Huson and Bryant 2006) was used to infer and visualize evolutionary relationships among the protein alleles at each locus using a maximum likelihood protein distance measure and the neighbor joining method of tree construction.

Results and Discussion

Results of the sequencing of the four *VCBF* genes revealed interesting patterns of genetic variation both between different species as well as between the different *VCBF* genes. Our original intent was to sequence the promoters of all four genes; however, the extent of heterozygous indels in the promoter regions prevented resolution of these regions. As a result, we sequenced 993 bp of the *VCBF1* locus including the coding sequence (753bp), 774 base pairs of the *VCBF2* coding sequence, 826 bp of the *VCBF3* locus including the coding sequence (717bp), and 885 bp of the *VCBF4* locus including the coding sequence (663bp). Of the 91 different grape accessions, 89 were sequenced for *VCBF1*, 87 for *VCBF3*, and all 91 for both *VCBF2* and *VCBF4*. All sequences have been uploaded to NCBI-GenBank (accession numbers KF582054-582411).

Overall structure of the CBF genes: The *Vitis CBF* genes are members of the *AP2* superfamily of genes. As has been previously reported (XIAO *et al.* 2006, 2008), *VCBF1*, *2*, and *3* are closely related to each other, but distantly related to *VCBF4*. All four *VCBF* proteins possess conserved N-terminal domains for nuclear localization (NLS) and for DNA binding (AP2). They also contain the DSAWRL and A Domains. The C-terminal of the genes is more variable but several conserved hydrophobic clusters (HC) can be identified (WANG *et al.* 2005). In addition to these domains, *VCBF*1, 2 and 3, contain three conserved serine repeat clusters with unknown function (XIAO *et al.* 2006, 2008). Alignment of the four *VCBF* genes is demonstrated in Fig. 1 along with amino acid substitutions observed in this study.

V C B F 1 : Analysis of genetic variation at the *VCBF1* locus revealed 45 SNPs in the coding region (5.98 % variation) with 20 synonymous changes and 25 non-synonymous changes. Twelve of these SNPs were singletons with 33 informative. The Ka/Ks ratio for this region was 0.39, suggesting mild purifying selection. Haplotype diversity ranged from 0.123 in *V. labrusca* to 0.894 in *V. cinerea*. There were a total of 39 different haplotypes at the *VCBF1* locus. While Tajima's D values were not significant, Fu's F (F = -13.25; P = 0.004) was significant and negative at the *VCBF1* locus when considering the entire dataset. A significant negative Tajima's D value was observed for *V. labrusca* (D = -1.97548; P = 0.008) and *V. rupestris* (D = -1.68878; P = 0.039) samples, with a marginal Tajima's D for *V. riparia* (D = -1.38922; P = 0.064) accessions. Negative values for Tajima's D indicate there are many singleton mutations at silent sites and suggest population expansion in these species or purifying selection. The genetic variation at the *VCBF1* locus results in a total of 28 novel protein alleles. Thirty alleles were found when including GenBank published sequences for *V. vinifera* 'Chardonnay' (GI 39578542), *V. riparia* 'Thunder Bay' (GI 39578538), and *V. amurensis* (GI 99029884). Evolutionary relationships between these alleles as determined by protein maximum likelihood distance measures are shown in Fig. 2. Six alleles were shared between species; the three alleles shared between *V. riparia* and *V. rupestris* reflect the close evolu-

tionary history of these two species (ARADHYA *et al.* 2013). Two alleles were shared between *V. vinifera* and *V. riparia*. However, one of these two alleles was found only in *V. riparia* samples (*V. riparia* allele 9, *V. vinifera* allele 7) which have hermaphroditic flowers. Hermaphroditic flowers are a hybrid trait and this evidence suggests that these *V. riparia* are carrying *V. vinifera* alleles. One allele was shared between *V. labrusca* and *V. aestivalis*.

Of the amino acid substitutions observed at the *VCBF1* locus, SNAP predicted that 12 amino acid substitutions are non-neutral. Of those 12 changes, 8 were singleton substitution alleles. Of these substitutions, 5 are located in important functional regions of CRT binding factors. Substitutions R56G, K72T, and E85K are found in the AP2 binding domain and could affect function (Fig. 1). Substitution R56G results in a positive to nonpolar change, K72T results in a positive to polar change, and E85K codes for a negative to positive charge substitution. Substitutions R56G and K72T are found in a heterozygous state in our samples and thus are unlikely manifest any major difference in abiotic stress phenotypes. Substitution E85K was present in the GenBank *V. vinifera* 'Chardonnay' sample and may impact *VCBF1* function in this genotype. The other two substitutions that hypothetically impact *VCBF1* function are found in the GenBank *V. vinifera* 'Chardonnay' and *V. amurensis* sequences. 'Chardonnay' has a phenylalanine to serine (F202S) substitution in hydrophobic cluster 2 while *V. amurensis* has a glutamate to valine (E220V) substitution in hydrophobic clusters 4 in the C-terminal of the protein that may impact function. However, results from WANG *et al.* (2005) suggest that CBF function in *Arabidopsis* is still possible when mutations occur in a single hydrophobic cluster. Thus any of these substitutions alone may not significantly impact function.

The other four substitutions predicted to be non-neutral are found in at least two different alleles across our samples. Two of these substitutions occur in conserved regions. N106I occurs in *V.* hybrid 'Noiret', *V. vinifera* 'Syrah', and as a heterozygous allele in 'Cabernet Franc'. This substitution occurs in the end of the AP2 domain and may or may not impact function as both asparagine and isoleucine can function in protein binding domains (BETTS
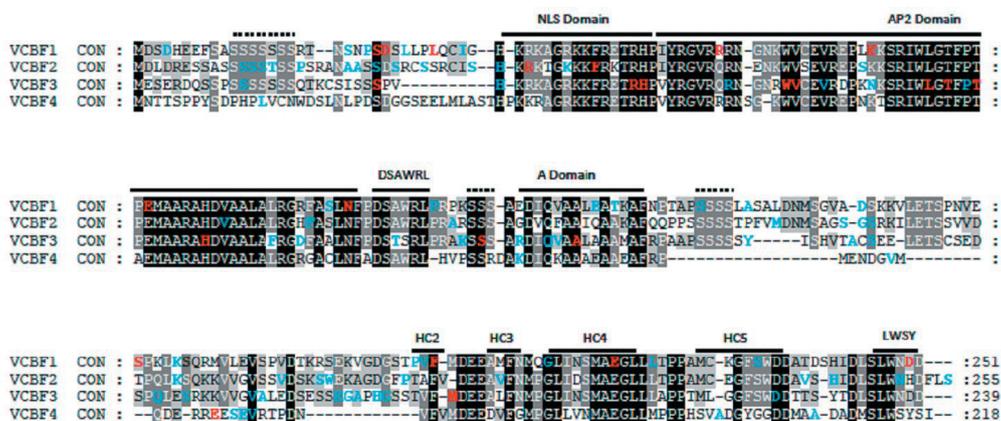
Fig. 1: Multiple Alignment of consensus protein sequence for *VCBF* genes. Conserved Domains highlighted with labeled lines. Dashed lines indicate conserved serine repeats found in *VCBF1*, 2 and 3. Colored residues indicate neutral (blue) and non-neutral (red) amino acid substitutions.
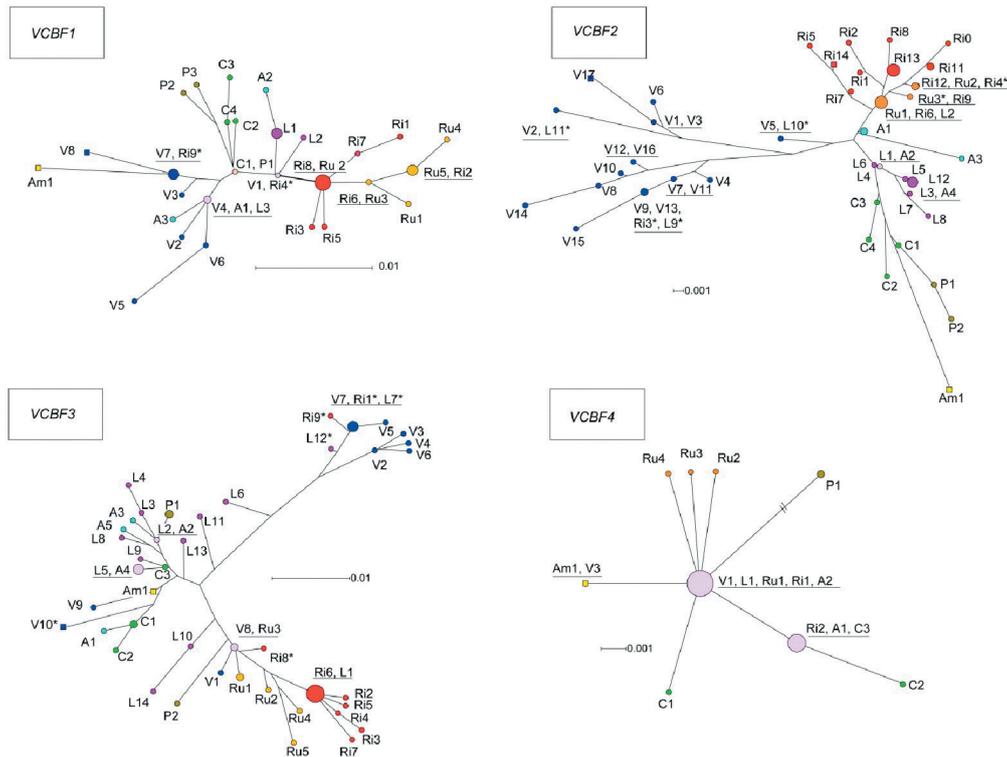
Fig. 2: Neighbor-Joining Network of protein alleles observed at the *VCBF1-4* loci. The sizes of the nodes indicate relative number of accessions with that allele. The smallest circles indicate alleles found as singleton heterozygous alleles. Square nodes denote GenBank samples and asterisk alleles indicate an allele from a wild-hybrid genotype. Nodes which are shared between species are indicated by underline. Labeling of alleles are by species and are as follows: *Amurenesis* (Am), *Aestivalis* (A), *Cinerea* (C), *Labrusca* (L), *Riparia* (Ri), *Rupestris* (Ru), *Palmata* (P), *Vinifera* (V).

and Russell 2003). The final substitution occurs in the HC6/LWSY domain at the C-terminus of the protein and is an aspartic acid to tyrosine substitution. However, this substitution was conserved in *V. labrusca* accessions and also occurs in *V. aestivalis* and *V. riparia* (Xiao *et al.* 2008), suggesting this change is likely neutral.

V C B F 2 : The *VCBF2* locus had a total of 67 SNPs in the coding region (8.7 % variation) with 31 synonymous changes and 29 non-synonymous changes. Additionally, indel polymorphisms detected at this locus resulted in the deletion of nine nucleotides that code for three serine amino acids in the N-terminal of the protein as well as two insertion polymorphisms of six nucleotides each, coding for a glycine-serine addition and an isoleucine-serine addition. Deletions in the N-terminal were observed in *V. rupestris*, *V. riparia*, *V. cinerea*, and *V. palmata* alleles. It was not detected in *V. vinifera* alleles, *V. aestivalis* and most *V. labrusca* alleles. The insertion polymorphisms were restricted to one *V. labrusca* allele and *V. vinifera* 'Syrah' for the serine-glycine insert, and exclusive to *V. vinifera* alleles for the isoleucine-serine insert. The Ka/Ks ratio for this region was 0.302 again indicating slight purifying selection at this locus. No evidence for selection was detected with the Tajima's D test for the entire dataset or for any individual species. However, significant Fu's F values were observed for the entire dataset (F = -24.01; P = 0.001), *V. riparia* (F = -12.05; P = 0.003) and *V. rupestris* (F = -5.08; P = 0.001) populations. This result suggests that the number of alleles present exceeds the number of expected alleles, either due

to population expansion or genetic hitchhiking (Fu 1997). Haplotype diversity at the *VCBF2* locus was high within species (0.894-0.938) except for *V. palmata* (0.333).

Seventy-seven different haplotypes were observed for *VCBF2* and these haplotypes give rise to 46 different protein alleles, with the greatest number of alleles found within *V. vinifera* and *V. labrusca*. In addition to the 89 sequenced samples, we also compared protein level variation from published sequences for *V. riparia* 'Thunder bay' (GI 39578544), *V. amurensis* (GI 99029886), and *V. vinifera* 'Chardonnay' (39578550). Evolutionary relationships between these alleles as determined by protein maximum likelihood distance measures are shown in Fig. 2 and closely mirror relationships observed at the *VCBF1* locus.

Alleles at the *VCBF2* locus were most frequently found in a heterozygous state with different alleles identified in every *V. vinifera* except 'Merlot' and 'Friuliano'. Eight alleles were shared between species; *V. labrusca* and *V. aestivalis* (2), *V. labrusca* and *V. vinifera* (3), and *V. riparia* and *V. rupestris* (3). Shared alleles between *V. labrusca* and *V. vinifera* were observed only in *V. labrusca* samples with suspected hybrid backgrounds indicated by the presence of hermaphroditic flowers. Of the three protein alleles observed in *V. rupestris*, all were shared with *V. riparia*.

There were a total of 45 potential amino acid substitutions observed for *VCBF2* with 5 substitutions occurring as multiple changes. SNAP results suggest that only two of the 45 substitutions are non-neutral changes and both changes

occur in the NLS domain, potentially impacting function of *VCBF2*. The SNP at site 46, F46L, was observed only in the GenBank accession for *V. amurensis* and the properties of both phenylalanine and leucine indicate this substitution is unlikely to change function (BETTS and RUSSEL 2003). The second predicted SNP site, R39L is present in 5 different *V. riparia* alleles including one shared allele with *V. rupestris* and may impact NLS domain function in these individuals.

V C B F 3 :   The highest level of genetic variation was observed at the *VCBF3* locus where a total of 64 SNPs were found in the coding region (8.9 % variation). Of these SNPs, 19 were synonymous changes while 38 were non-synonymous, resulting in a Ka/Ks ration of 0.636. This ratio is higher than that observed at *VCBF1* and *VCBF2*, suggesting less purifying selection and a more neutral mutation rate than observed at these loci. There was no evidence of selection at the level of the entire dataset for Tajima's D, but significant negative values for Fu's F (F = -13.58; P = 0.018) were observed. Additionally, significant negative values for Tajima's D (D = -1.67573; P = 0.015) were observed for *V. riparia*, suggesting purifying selection or population expansion in this species subset. Variation at the *VCBF3* locus resulted in 58 different haplotypes with haplotype diversity within species ranging from 0.533 (*V. palmata*) to 1 (*V. aestivalis*).

In addition to the 87 samples we sequenced for *VCBF3*, published sequences for *V. riparia* 'Thunderbay' (GI 39578546), *V. amurensis* (GI 99029888), *V. hybrid* 'Kyoho' (GI 254952726), and *V. vinifera* 'Chardonnay' (GI 39578548) were compared. Variation at this locus results in 42 different protein alleles differing by 39 substitutions, with the vast majority of these alleles observed as singletons and in a heterozygous state, with greatest allelic diversity present in *V. labrusca*. Five alleles were shared between *V. labrusca* and *V. aestivalis* (2), *V. labrusca* and *V. riparia* (1), *V. labrusca*, *V. riparia*, and *V. vinifera* (1), and between *V. vinifera* and *V. rupestris* (1). Evolutionary relationships between these alleles as determined by protein maximum likelihood distance measures are shown in Fig. 2. This tree again closely mirrors results observed at *VCBF1* and *VCBF2,* with *V. riparia* and *V. rupestris* closely related and the other North American species well-differentiated from *V. vinifera*.

SNAP analysis suggest that 13 of the 39 substitutions are non-neutral. Eight of these are found as singletons. Substitutions R42P, L72R, T74S, are found in the heterozygous state in *V. riparia* and *V. rupestris* accessions and each of these substitutions are located in the NLS and AP2 domains of the protein and may impact function. Additionally, T74S is a threonine to serine substitution that occurs near the putative binding sites for *V*CBF-DNA binding (OKAMURO *et al.* 1997). Two other singleton substitutions (R62L, T77A) are found in a GenBank accession of hybrid grape, *V. labrusca* x *V. vinifera* 'Kyoho' (GI: 254952726). Both substitutions occur in the AP2 domain and the R62L substitution occurs at a highly conserved putative DNA binding site, potentially impacting function of *VCBF3* in this accession. Arginine is effective in binding with phosphates and likely enhances DNA binding ability. Thus sub-

stitution of this site to leucine may make this site less reactive and less able to bind to DNA to initiate transcription (BETTS and RUSSELL 2003). One SNP, S114Y, is found in *V. palmata* and is a serine to tyrosine substitution that occurs in the serine repeat domain between the DSAWRL and A domains. It is not known if these serine repeat domains have important function in the *VCBF* genes, and thus it is not clear if this substation would alter protein function. The final two non-neutral substitutions are a tryptophan to tyrosine substitution (W57Y) in the AP2 domain and a premature stop codon at AA58, adjacent to this mutation. These substitutions are found in a homozygous state in *V. vinifera* 'Lemberger'. One of the two alleles observed for 'Lemberger' also carries other shared SNPs identified in this gene elsewhere in the sample dataset. This result suggests that some event may have occurred in 'Lemberger' since its selection from other cultivated varieties as the premature stop codon and tryptophan to tyrosine substitution is found in both 'Lemberger' alleles. Perhaps inactivation from transposon movement or some other double stranded damage has occurred. As a result, 'Lemberger' is not expected to have a functional *VCBF3* gene. Based on function of this gene in other *Vitis* studies, some reduction in abiotic stress pathway activation may occur, though the resolution of *VCBF* action in *Vitis* does not allow us to predict what processes might be affected (XIAO *et al.* 2006). Also, functional redundancy with *VCBF1* and *VCBF2* proteins may allow this mutation to go unselected.

The remaining five substitutions (S26F, H43P, H85Q, A123G, and M192T) are shared between species or between accessions within different haplotype groups within species. All of these substitutions occur in the heterozygous state within the respective *Vitis* accessions. Substitution S26F occurs in a non-conserved region and thus may not impact function. This substitution is also found only with the *V. aestivalis* samples of our study. The remaining four substitutions occur in the AP2 Domain (2), A domain (1), and HC2 domains (1) and have the potential to affect function in these accessions.

V C B F 4 :   Despite comparable levels of genetic variation at the *VCBF4* locus, 48 SNPs (7.2 % variation) 38 polymorphisms were synonymous changes while only 10 were found to be non-synonymous. Thus the Ka/Ks ratio was lowest at the VCBF4 locus at 0.081, suggesting strong purifying selection at this locus. Haplotype diversity remained high (0.922) with 54 haplotypes inferred. Tests of neutrality/ selection indicated a significant negative Tajima's D (D = -1.471; P = 0.047) and significant Fu's F (F = -25.54; P < 0.0001) when tested across the entire dataset. Additionally, a significant Fu's F (F = -8.491; P = 0.001) was observed for *V. riparia* accessions. Published sequences for *V. riparia* 'Thunderbay' (GI 57903606), *V. amurensis* (GI 324103759), *V. hybrid* 'Tamnara' (GI 163914221), and *V. vinifera* 'Chardonnay' (GI 92918850) and 'Muscat of Hamburg' (GI 410519343) were available for comparison. Protein allele variation at the *VCBF4* locus was drastically lower than in the other three CBF genes.

Nine protein alleles were detected in our population with a tenth allele obtained from GenBank samples. One allele was shared between *V. vinifera*, *V. riparia*, *V. rupes-*

tris, *V. labrusca*, and *V. aestivalis* and was the most common allele in the study (53 % of samples). A second allele was shared between *V. riparia*, *V. cinerea*, and *V. aestivalis* (36 % of samples). The Genbank allele was shared between *V. vinifera* 'Muscat of Hamburg' and *V. amurensis* and was not observed in our germplasm survey. As a result of the greatly reduced diversity in protein alleles, evolutionary relationships are very different from those at the other three *VCBF* genes (Fig. 2) with essentially no species resolution.

Of the eleven substitutions observed at this locus, two were predicted to be non-neutral. The first substitution was observed in *V. vinifera* 'Merlot' in a heterozygous state. 'Merlot' carries the most common allele as well as a duplicate allele with a hypothetical premature stop codon located between the DSAWRL domain and the A domain. This mutation could result in a protein that is still nuclear localized, but due to the stop codon would lack the N-terminal translational activation domains required for function. The second substitution E159G is found as a singleton in *V. rupestris* and is located in a non-conserved region of the gene, likely not impacting function.

Conclusion

In conclusion, variation was high at each of the *VCBF* loci in this study. Of the four genes, protein allele diversity was very high for *VCBF1*, 2, and 3 but very limited for *VCBF4*. This result, combined with the lower Ka/Ks ratios for *VCBF4* indicates that selection acts on *VCBF4* to reduce changes in protein structure, even across species boundaries. Neighbor joining trees of *VCBF1*, 2 and 3 all resolve previous evolutionary relationships between these wild species (ARADHYA *et al.* 2013), with only *VCBF4* having a minimal tree dominated by two common alleles. This result is perhaps not unexpected given the role of *VCBF4* in cold response and the survival of these germplasm accessions in northern New York. At the same time, amino acid changes in this gene cannot explain differences in cold tolerance between species as all *V. vinifera VCBF4* protein alleles were also shared with wild North American grapevines. This result suggests that differences in amino acid sequence at this locus are important and tightly conserved between species and that mutations in this gene have been eliminated by natural selection. As gene expression differences do also not appear to contribute to the difference in phenotype, genes downstream of the *VCBF4* regulon are the more likely candidates underlying variation in *CBF*-regulated cold tolerance. Future studies designed to examine the expression and function of the alleles identified with potentially non-neutral function should be conducted.

Additionally, studies designed to examine the gene variation of southern wild grapevine species may further elucidate the evolutionary forces taking place at the *VCBF* loci in *Vitis*.

References

ARADHYA. M.; WANG, Y.; WALKER, M. A.; PRINS, B. H.; KOEHMSTEDT, A. M.; VELASCO, D.; GERRATH, J. M.; DANGL, G. S.; PREECE, J. E.; 2013: Genetic diversity, structure, and patterns of differentiation in the genus Vitis. Plant System. Evol. **299**, 317-330.

BETTS, M. J.; RUSSELL, R. B.; 2003: Amino acid properties and consequences of substitutions. In: R. R. BARNES, I. C. GRAY (Eds): Bioinformatics for geneticists. John Wiley and Sons, Ltd.

BROMBERG, Y.; ROST, B.; 2007: SNAP: predict effect of non-synonymous polymorphisms on function. Nucl. Acids Res. **35**, 3823-3835.

EXCOFFIER, L.; LISCHER, H. E. L.; 2010: Arlequin suite ver 3.5: A new series of programs to perform population genetics analysis under Linux and Windows. Mol. Ecol. Res. **10**, 564-567.

FU, Y. X.; 1997: Statistical tests of neutrality of mutations against population growth, hitch-hiking, and background selection. Genetics **147**, 915-925.

HAAKE, V.; COOK, D.; RIECHMANN, J. L.; PINEDA, O.; THOMASHOW, M. F.; ZHANG, J. Z.; 2002: Transcription factor CBF4 is a regulator of drought adaptation in Arabidopsis. Plant Physiol. **130**, 639-648.

HUSON, D. H.; BRYANT, D.; 2006: Application of Phylogenetic Networks in Evolutionary Studies. Mol. Biol. Evol. **23**, 254-267

LIBRADO, P.; ROZAS, J.; 2009: DNaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics **25**, 1451-1452.

NOVILLO, F.; MEDINA, J.; SALINAS, J.; 2007: Arabidopsis CBF1 and CBF3 have a different function than CBF2 in cold acclimation and define different gene classes in the CBF regulon. PNAS **104**, 21002-21007.

OKAMURO, J. K.; CASTER, B.; VILLARROEL, R.; VAN MONTAGU, M.; JOFUKU, K. D.; 1997: The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in Arabidopsis. PNAS **94**, 7076-7081.

XIAO, H.; SIDDIQUA, M.; BRAYBROOK, S.; NASSUTH, A.; 2006: Three grape CBF/DREB1 genes respond to low temperature, drought and abscisic acid. Plant, Cell Environ. **29**, 1410-1421.

XIAO, H.; TATTERSALL, E. A. R.; SIDDIQUA, M. K.; CRAMER, G. R.; NASSUTH, A.; 2008: CBF4 is a unique member of the CBF transcription factor family of *Vitis vinifera* and *Vitis riparia*. Plant, Cell Environ. **31**, 1-10.

WANG, Z.; TRIEZENBERG, S. J.; THOMASHOW, M. F.; STOCKINGER, E. J.; 2005: Multiple hydrophobic motifs in Arabidopsis CBF1 COOH-terminus provide functional redundancy in trans-activation. Plant Mol. Biol. **58**, 543-559.