GENETIC ANALYSIS OF GRAPEVINE LEAFROLL-ASSOCIATED VIRUS 3 POPULATION FROM GALICIA, SPAIN

Journal:	*Plant Pathology*
Manuscript ID:	PP-15-049.R3
Manuscript Type:	Original Article
Date Submitted by the Author:	08-May-2015

Complete List of Authors:	Pesqueira, Ana; Universidad de Santiago de Compostela, Producción Vegetal Cabaleiro, Cristina; Universidad de Santiago de Compostela, Producción Vegetal Velasco, Leonardo; IFAPA, Plant Protection

Topics:	epidemiology
Organisms:	viruses & viroids
Other Keywords:	grapevine leafroll virus, population genetics, diversity, recombination

SCHOLARONE™
Manuscripts

eview

Running head: Genetic analysis of GLRaV-3 from Galicia

Genetic analysis of *Grapevine leafroll-associated virus 3* population from Galicia, Spain

A. M. Pesqueira[a], C. Cabaleiro[a] and L. Velasco[b]*

[a]*Plant Protection, Departamento de Producción Vegetal, Escola Politécnica Superior, Universidad de Santiago de Compostela, 27002 Lugo; and* [b]*Instituto Andaluz de Investigación y Formación Agraria y Pesquera (IFAPA), 29140 Churriana, Málaga, Spain*

*E-mail: leonardo.velasco@juntadeandalucia.es

For Peer Review

Grapevine leafroll-associated virus 3 (GLRaV-3; *Ampelovirus*, *Closteroviridae*) isolates from Galicia in northwestern Spain were selected to characterize their genetic diversity according to different factors (age, origin, location, variety, etc.). The vines belonged either to local white and red varieties autochthonous to Galicia or to varieties from other Spanish regions but widely used in Galicia. These GLRaV-3 isolates came from different vineyards in Galicia located in coastal or inner areas. Multiplex RT-PCR allowed the detection of isolates belonging to groups I, II, III–V and VI. Two genomic regions were studied in the isolates, the HSP70h and the CP, using specific primers that allow the detection of variants from groups I to V. Some possible recombinants could be detected; however, multiple infections with different variants indicated that they were not genuine recombinants. No differences were found in the population structure considering variety or geographical factors. Isolates belonging to four groups were found in the distinct areas surveyed: groups I and II were the most common, followed by groups VI and III, as is the case in the rest of the world. In the same surveys, the presence of insect vectors for GLRaV-3 was investigated and found

lacking in inland areas but present in those with milder climate. Genetic analysis did not support isolation of the GLRaV-3 isolates in Galicia, suggesting that the uncontrolled exchange of infected vines and/or rootstocks has been a major agent of virus spread.

Keywords: Grapevine leafroll virus, population genetics, diversity, recombination

Introduction

For Peer Review

Grapevine leafroll disease (GLD) is spread worldwide in grapevine areas, with *Grapevine leafroll-associated virus 3* (GLRaV-3) as the most important agent for the disease (Maree *et al.*, 2013). Many studies attest the negative influence of GLD on yield and must quality, and thus the importance of the use of healthy certified material both for the scion and the rootstock (for a comprehensive approach, see Martelli, 2014). Galicia, a region located in northwestern Spain, holds a high number of autochthonous grapevine varieties. In addition, there are varieties common both in Galicia and northern Portugal and vineyards planted with well-known Spanish varieties. Grapevine is widely distributed in Galicia, with about 25 000 ha grown mainly in small plots by numerous farmers (244 180 plots). They produce an average of 80 000 tonnes of wine, and represent a key sector in the socio-economy of the region by contributing to the stability of populations in rural areas (http://mediorural.xunta.es/). According to the EU Commission Directive 2005/43/EC (Anon, 2005) on the marketing of grapevine propagation material, the absence of GLRaV-3, among other viruses, must be granted in certified stock. In Galicia, certified virus-free plant material was not available before 2000, more than 20 years after the beginning of the expansion of vineyards based on autochthonous varieties. Therefore, most of the vineyards were planted with 'standard' propagation material that is currently allowed and widely used; in such material the rootstocks are usually certified but for the variety, only identity and visual

inspection for 'harmful organisms' are required; this standard plant material does not have specific identification numbers and cannot be tracked. In addition to the acquisition of standard and certified plants from professional nurseries, many growers choose or need to propagate their own plant material from minor varieties, which are field grafted or sent to the nurseries to be grafted onto certified rootstocks. Thus, the present plant material in Galician vineyards is quite heterogeneous and the origin of the leafroll viruses difficult to determine. Although GLD has been studied in Galicia since the mid-1990s (Cabaleiro & Segura, 1997; Cabaleiro *et al.*, 2008), up to the present, no studies have been carried out to investigate the genetic variation of local GLRaV-3 isolates and its relationship with variety, vine origin or location that could help in understanding the epidemiology of this important disease.

GLRaV-3 and some of its vectors have been distributed worldwide by means of infected plant material and subsequently by imported or local vectors which are mainly mealybugs (Hemiptera: Pseudococcidae), but also scale insects (Hemiptera: Coccidae). In particular the species *Planococcus ficus* and *Pseudococcus longispinus* are the most distributed and efficient vectors (Tsai *et al.*, 2008). In the first surveys carried out in Galicia (1991/2) to determine the prevalence of GLRaV-3, mealybugs were found only in two vineyards close to the coastal areas, but with low populations and no more than two generations per year, probably because accumulated degree-days are lower than in other grapevine-growing regions (Cid *et al.*, 2010). In a second survey, in 2003/4, in six vineyards out of 47, small colonies of *Planococcus* were observed in single plants or small foci; always in vineyards under mild climate and humid conditions close to the sea or at the end of the Miño river. In the inland appellations no mealybugs were detected, although some scale insects occasionally appeared.

In recent years GLRaV-3, as the most distributed leafroll species, has been the subject of intensive analysis of molecular variability, particularly based on the coding genomic

For Peer Review

regions for the heat shock protein (HSP70h) and capsid protein (CP) (Turturo *et al.*, 2005; Jooste *et al.*, 2010; Sharma *et al.*, 2011; Wang *et al.*, 2011). Research carried out in different wine-producing zones has revealed the existence of molecular variants of GLRaV-3 that have been grouped into six main clades or groups (reviewed in Maree *et al.*, 2013), apart from some other divergent variants (Seah *et al.*, 2012; Goszczynski, 2013). Isolates belonging to one group or another have different representation in the vineyards, and the reasons for these differences are not clear, but can be related to differences in natural virus transmission among the variants (Gouveia *et al.*, 2011). Although the presence of recombinant isolates has been described (Turturo *et al.*, 2005), in another later study recombination was not detected (Wang *et al.*, 2011). The reason for this contradiction can be found in the common presence of multiple infections with different variants in the same vine. This study investigated the population structure of GLRaV-3 in Galicia, Spain, and its relationship with host origin and environmental factors in order to provide clues for understanding virus spread in the region.

Material and methods

Vineyards and sample collection

Vines were originally collected from at least 15 vineyards where GLRaV-3 had been detected using serological tests, trying to cover as many varieties as possible and find sources of variation related to the age of vines, plant type and origin (Fig. 1; Table 1). Samples were taken from the established germplasm collections in the University Campus and in an experimental vineyard in Portomarín (Lugo, Spain) and from five conventional vineyards that are representative for the region. In addition, three isolates belonging to grapevine collections were included in the analysis: one (Y-285) from the Institut Nationale de la Recherche Agronomique (INRA collection, Colmar, France) and two isolates (M20 and M34) from

For Peer Review

Neiker (Instituto Vasco de Investigación y Desarrollo) collection (La Rioja, Spain). All vineyards and plants were inspected for the presence of mealybugs and other scale insects and in some of them, spatial analysis of infected plants was performed to know whether or not virus spread had occurred (Pesqueira, 2013). Leafroll symptoms were observed (in several years) for most of the plants (Table 1). DAS-ELISA could detect all the isolates described in this work (Agritest). In total, 98 samples were chosen for GLRaV-3 variant specific multiplex RT-PCR. Among these, 54 representative samples from all the GLRaV-3 positive plants in the two grapevine germplasm collections and a random selection of isolates from the five vineyards were used for GLRaV-3 genetic analysis.

For Peer Review

Total RNA extraction

For the selected samples that were subjected to sequence analysis, 100 mg of plant tissue from wood scrapings of phloem or from petioles from mature leaves were used. RNA was extracted using the RNeasy Plant Mini kit (QIAGEN) according to the manufacturer's recommended protocol. After extraction, RNA was eluted in 50 μL RNase-free water and stored at –70°C until further use. For the GLRaV-3 variant specific analysis in the set of 98 samples, 80–100 mg of plant tissue (leaf petioles) were homogenized using TissueLyser II (QIAGEN) and processed according to the method described by Chooi *et al*. (2013b). The quality and yield of all RNA extractions were determined using the ND-1000 spectrophotometer (NanoDrop Technologies).

Variant-specific GLRaV-3 detection using multiplex RT-PCR

The multiplex RT-PCR assays to discriminate between different GLRaV-3 variants in the same samples, were carried out as described in Chooi *et al.* (2013b). The PCR products were examined on 3–4% TBE agarose gels under UV light after staining with 1× RedSafe solution (Intron Biotechnologies). This assay was performed in order to overcome the limitation of the primers proposed by Turturo *et al.* (2005) to detect variants other than those belonging to groups I to V.

For Peer Review

cDNA synthesis and PCR for amplicon sequence analysis

The amplification reactions were performed in a single-step RT-PCR in a final volume of 50 μL using 200 ng of the RNA extract and the SuperScript III One-Step RT-PCR System kit (Invitrogen) with 0·5 μM of each forward and reverse primer. The primer pair LC1/LC2 was used for the amplification of the HSP70h genomic region, and CP3U/CP3D for the CP region (Turturo *et al.*, 2005). The reactions of reverse transcription and amplification were performed in an Omn-E thermal cycler (Hybaid) under the following conditions: cDNA synthesis at 50°C for 25 min; denaturation at 94°C for 2 min; and 40 cycles of amplification at 94°C for 15 s, 55°C for 30 s and 68°C for 1 min. Synthesis was completed with a final extension at 68°C for 5 min. The resulting amplicons were inspected by electrophoresis in agarose gels after staining with ethidium bromide or RedSafe and visualized under UV light.

Viral sequence determination and analysis

Amplicons were purified using the PureLink Quick Gel Extraction kit (Invitrogen) and sequenced in both directions using an ABI 3730xl Genetic Analyzer (Applied Biosystems) based on the primers with which each amplicon was generated. Sequences were processed

with the software GENEIOUS v. 7.1 (Biomatters). Both genomic regions were selected because they are the most studied and have the most sequences available in the databases. The nucleotide sequences obtained in this study were deposited in GenBank (Table 1). The genetic analysis was restricted to variants belonging to groups I to V given the primers chosen in this study (Turturo *et al.*, 2005).

Phylogenetic trees and diversity analysis of HSP70h and CP genomic regions

A database was built consisting of the sequences obtained in this work (Table 1) and a set of sequences available in GenBank for GLRaV-3 as comparison. The sequences for each genomic region were aligned using the CLUSTALW algorithm (Larkin *et al.*, 2007) in MEGA v. 6.06 (Tamura *et al.*, 2013). The aligned sequences of the HSP70h and CP genomic regions were used to infer maximum likelihood phylogenetic trees with the general time reversible (GTR+G) model and four categories of rates for the HSP70h and CP genes, using 500 bootstrap replicates for assessing the confidence of the tree branches.

Genetic distance analysis between pairs of sequences and genetic diversity for the nucleotide and derived amino acid sequences were carried out in MEGA. The nucleotide genetic diversity (d) and synonymous (dS) and nonsynonymous (dNS) substitutions at codons were calculated using maximum composite likelihood algorithms (MCL) and Pamilo, Bianchi-Li (Li, 1993; Pamilo & Bianchi, 1993) with 1000 bootstrap replicates. Also, the values of the dNS/dS ratio for each genomic region were calculated in order to determine the selection pressures. A value of dNS/dS > 1 indicates a positive selection, while values < 1 indicate negative or purifying selection. Additional parameters were also estimated such as the nucleotide polymorphism (π: average estimate of nucleotide differences between two

For Peer Review

sequences randomly in a population) and the statistic Θ, number of segregating sites (*S*) (Watterson, 1975).

Population analysis of GLRaV-3 isolates

The evolutionary divergence in the Galician GLRaV-3 population was estimated by examining different parameters. In order to identify the amino acids under selective pressure, the differences between synonymous and nonsynonymous substitutions were calculated for each amino acid position in the alignment using the fixed-effects likelihood (FEL) method and the internal fixed-effects likelihood (IFEL; Kosakovsky Pond *et al.*, 2006) available on the server Datamonkey (www.datamonkey.org). The evolutionary analysis to assess the type of selective pressure on one or another genomic region in populations was performed with different phylogenetic tests: Tajima's D neutrality test (Tajima, 1989), based on the analysis of the number of segregating sites and the average number of nucleotide differences; Fu and Li's D* statistics based on the differences between the number of singletons (mutations that appear only once between the sequences) and number of mutations; and finally Fu and Li's F* test based on the differences between the number of singletons and the average number of nucleotide differences between pairs of sequences. Wright's fixation index, which is the ratio of the total genetic variance contained in a subpopulation (subscript S) with respect to the total genetic variance (subscript T), was calculated. F_{ST} values range between 0 and 1, being 0 when two populations are undifferentiated because they have identical distribution of haplotypes, and 1 as fully differentiated populations. It is considered that if $F_{ST} > 0{\cdot}33$ it indicates an infrequent genetic flow. The statistical tests mentioned and the F_{ST} estimates were calculated using DNASP v. 5.0 (Librado & Rozas, 2009). This same program was used to estimate the presence of subdivisions for each genomic region in each group of GLRaV-3

For Peer Review

isolates with the statistics Hudson-Ks*, Z and the statistic tests of genetic differentiation Snn (Hudson, 2000). An Snn value close to 1 indicates high population differentiation and a value of Snn < 0·5 suggests that there has been a random exchange among the different groups compared. The correlation between geographic distance and genetic diversity was investigated by a Mantel test with 10 000 permutations using PASSAGE v. 2.0 software (Rosenberg & Anderson, 2011).

For Peer Review

Recombination event analysis

Potential recombination events were evaluated in the concatenated genomic regions HSP70h and CP using the RDP4 package (Martin *et al.*, 2010), which includes the recombination-detecting algorithms GENECONV, BOOTSCAN, MAXCHI, CHIMAERA, SISCAN, 3SEQ and RDP, using their default parameter values.

Results

Analysis of the genetic diversity of GLRaV-3 isolates from Galicia

The isolates collected were classified according to their origin from autochthonous Galician varieties (minor red varieties and Albariño) or other allochthonous varieties (Mencía and Tempranillo). GLRaV-3 variant analysis using multiplex RT-PCR of the 98 samples revealed the presence of variants belonging to groups I, II, III–V and VI (Table 2). Group II was the most prevalent, representing 84·7% of the isolates, followed by group I (26·5%), group VI (17·3%; isolates NZ1 and NZ2; Table 2) and groups III–V (7·1%). The multiplex RT-PCR approach used does not allow discrimination among groups III to V (Chooi *et al.*, 2013b).

Additionally, amplicons from representative isolates belonging to group VI were sequenced in order to confirm their identity (Table 1). Isolates V5-4.35NZ, F10CAS2NZ, F10CAS15NZ and F10CAS18NZ showed 94·54% nucleotide identity in 110 nt of the HSP70h/HSP90h genomic region with respect to isolate NZ1-B (JX220900). Isolates F10CAS2NZ, F10CAS15NZ and F10CAS18NZ were 100% identical showing only one synonymous nucleotide difference with respect to V5-4.35NZ. Isolates 3A3NZ, F26CAS48NZ, F37CA18NZ, F38CA18NZ and F38CA15NZ showed 92·7% nucleotide identity with respect to isolate NZ2 (JX220899) in 340 nt within the HSP70h region, and were all identical to each other. In addition, 49% of the samples showed infection by GLRaV-3 variants belonging to two different groups, while 3·1% of the samples showed triple infections. Regarding the distribution of the isolates in the five inspected vineyards, evidence for spread of variants within the vineyard was most evident in two of the three coastal ones, Goián-Castañal and Meaño, where NZ1 or NZ2 variants (group VI) seemed to be spreading over a group II variant background (Table 3). In the other vineyards it seems that different GLRaV-3 introductions have appeared with time because of the probable planting of vines with different origins (Table 3). Four of the vines included in this study are clones from centenary non-stock plants and have been kept free of insect vectors in the germplasm collections. The GLRaV-3 isolates present in them belong to either variant group I or II, or to a mixed infection in the case of isolate L13 (Table 1).

For a set of selected isolates, genetic diversity was analysed using nucleotide and amino acid sequences obtained for the two genomic regions of GLRaV-3, corresponding to 413 nt of the coding genomic region for HSP70h and 390 nt corresponding to a part of the CP-encoding region. To estimate the genetic structure of the GLRaV-3 isolates from Galicia the phylogenetic relationships were investigated using the maximum likelihood (ML) method (Fig. 2); the Bayesian inference approach displayed the same topology (not shown). The

For Peer Review

primers chosen in this work restricted these analyses to variants belonging to groups I to V, excluding variants from group VI. Galician isolates could be assigned to only three of the variant groups previously described for the CP region: groups I, II and III. Similarly, for the HSP70h genomic region, sequences were assigned to the same three genetic variant groups. Most of the isolates could be assigned to group II followed by group I, after defining the different groups on the basis of reference sequences. Some isolates were assigned to group I or II disregarding the genomic region considered (Fig. 2). In other cases, the sequence showed ambiguous nucleotide positions, reflecting the presence of amplicons from different isolates in the same sample. Multiplex RT-PCR confirmed in these plants the simultaneous presence of different variants. Isolates GA2 (from an ungrafted centenary plant of Albariño), V-1A2, 1B15 and 2B2 (all of them Albariños) and the isolate MC20 from one Tempranillo plant from La Rioja (Spain), present in the collection, were assigned to group III according to both the HSP70h and CP regions. No isolates were found that could be assigned to groups IV or V.

The highest genetic distance between pairs of isolates for the HSP70h region was 0·155 corresponding to isolates F2CA1 (group II) and 1B15 (group III; Table S1). For the CP region, the highest genetic distance, 0·135, corresponded to isolates MP4.11 (group II) and 1B15 (group III; Table S2). Average diversity values obtained by MLC from the set of isolates collected in Galicia both for the CP and the HSP70h regions, were lower than the average diversity of the set of 185 sequences deposited in GenBank for the CP region and for the set of 96 GenBank sequences corresponding to the HSP70h (Table 4). The average genetic diversity of the Galician GLRaV-3 isolates was found to be 0·058 for the CP region, similar to that described by Turturo *et al.* (2005), who obtained a value of 0·049 for the same genomic region corresponding to a set of isolates from 14 countries. This value was also in

For Peer Review

accordance with the one obtained from the set of GenBank isolates, which was 0·053, and close to the value of 0·063 obtained by Gouveia *et al.* (2011) for traditional Portuguese varieties. For the HSP70h region, the diversity of the isolates studied in this work was 0·053, higher than the diversity value of 0·034 reported previously from a collection of isolates from different parts of the world (Turturo *et al*., 2005). More specifically, for both genomic regions, the average genetic diversity was higher for the autochthonous varieties than for the allochthonous ones. When isolates were classified according to variety, the average pairwise nucleotide diversity π ranged between a minimum of 0·033 for the HSP70h region belonging to the group of the autochthonous red varieties and 0·071for the HSP70h region in the Albariños. The average nucleotide diversity value for the CP region was slightly higher than for HSP70h in the isolates from Galicia, in contrast with the values obtained for the set of GLRaV-3 isolates from the rest of the world present in GenBank, where the values for both genomic regions are very different, and much higher for the CP region.

For Peer Review

Identification of selective pressures in CP and HSP70h regions of GLRaV-3

The number of synonymous substitutions per site (dS) for the Galician isolates was higher, although not significantly, for the CP than for the HSP70h coding region, a similar trend to the one observed for the set of representative GLRaV-3 isolates from the rest of the world (Table 4). Nonsynonymous substitutions (dNS) per site were lower than the synonymous substitutions for both genomic regions. The ratio between nucleotide diversity at nonsynonymous and synonymous positions per site (dNS/dS) provided an estimate of the degree and direction of selective constraints operating in the coding regions. The dNS/dS values were low in all the groups from Galicia but higher in the Albariños because of the presence of group III in that set. The highest values for the Galician isolates were found in the

CP region, although in the same range as those for the HSP70h region, indicating that these genomic regions are under similar selective pressures. Similarly, for the set of sequences from GenBank, the dNS/dS ratios were in the same range for the HSP70h and CP regions. In all cases, the dNS/dS ratio was less than 1, implying that both genomic regions are under predominantly purifying selection.

According to the FEL analysis, eight sites could be identified as being under negative and none under positive pressure for the CP regions ($P < 0{\cdot}05$). Using the FEL at $P < 0{\cdot}1$ allowed the identification of a single position under positive pressure in this CP region. Alternatively, the IFEL at $P < 0{\cdot}1$ allowed identification of a single position under positive pressure in that genomic region and 13 negatively selected sites. The FEL method allowed the identification of 26 positions under negative pressure for the HSP70h region and none under positive pressure ($P < 0{\cdot}1$). For the HSP70h region, no position was found to be under positive pressure using the IFEL method and seven under negative pressure ($P < 0{\cdot}05$). Using the IFEL at $P < 0{\cdot}1$ it was possible to identify a single position under positive pressure and 14 under negative pressure in this genomic region. Population genetic parameters were investigated and neutrality tests were performed using three statistical tests: Tajima's D, Fu and Li's D* and Fu and Li's F* (Tables 4 & 5). In the case of the HSP70h region, the isolates from Galicia showed no significant deviation from neutrality. Similarly, for the CP region none of the three statistics deviated significantly from neutrality. Regarding specific subpopulations, only in the case of the autochthonous reds, results of two of the three statistical tests showed a significant deviation of neutrality in the HSP70h region. Thus the isolates from Galicia are not subjected to new selective processes. In contrast, in the analysis of the set of GenBank worldwide sequences for the HSP70h region, the results from the three

For Peer Review

statistical tests showed a significant deviation from neutrality, suggesting some demographic expansion of GLRaV-3 world population (Table 4).

Analysis of the differentiation between GLRaV-3 subpopulations

To estimate the degree of genetic differentiation, the F_{ST} coefficient was determined, which additionally provides an estimate of the gene flow. The F_{ST} value for the CP region in GLRaV-3 populations from Galicia compared with the world isolates was 0·0890, indicating no genetic differentiation. In the case of the HSP70h region, the F_{ST} value of Galician versus world isolates was 0·1643. Usually, an absolute value of F_{ST} greater than 0·33 suggests a reduced genetic flow. Thus, lower values indicate that there has not been any isolation of GLRaV-3 populations from Galicia with respect to the rest of the world. In addition, the diversity of the population of GLRaV-3 from Galicia is in the same range as for the set of GenBank isolates. Regarding the analysis among Galician GLRaV-3 subpopulations, several different groups were considered: one formed by the autochthonous red varieties, not common and probably coming from a few mother plants (including Brancellao, Castañal, Caíño Tinto/Tinta Femia and Retinto); a group including the autochthonous white variety (Albariño); and finally a group including allochthonous varieties (Mencía and Tempranillo). To test the genetic differentiation between and among subpopulations, three statistical tests based on sequences with high mutation rate and low number of samples, Ks*, Z and Snn, were used (Hudson, 2000; Table 5). The Ks*, Z and Snn tests were not significant for the CP region, which indicates no genetic differentiation between subpopulations of GLRaV-3 except for the autochthonous whites, because of the presence of group III isolates in this subpopulation. A similar result was obtained for the HSP70h region. For both the CP and HSP70h regions the Snn test showed significant differences between some subpopulations,

For Peer Review

but this result was not supported by the other two statistics (Ks* and Z). In addition, when considering the location of the different populations, inland or coastal, no differences were found. Finally, the Mantel test did not show significant correlation between geographic distance among isolates and their respective genetic distances (data not shown).

Recombination events in the GLRaV-3 isolates

To assess the frequency and extent of recombination during GLRaV-3 diversification in Galicia, the two genomic regions were concatenated and recombination events were analysed for the 43 isolates where the sequence for both regions was available (Table 1). Potential recombination events were confirmed by at least five of the methods in isolates C1, L13, D2-A16, MC34, TAC2.9 and F35.BR8. All of them showed as a major parent, corresponding to the HSP70h region, an isolate similar to 69A (group I), and as a minor parent, the CP region, one isolate belonging to group II. However, variant-specific multiplex RT-PCR showed that in those samples at least two different variants were present belonging to groups I and II. Consequently, the amplicons of the HSP70h and CP probably belonged to different variants and not to a recombinant isolate (Table 2). No recombination event was found within the genomic regions, but a possible explanation could be that the sequences analysed in this work were short.

Discussion

In Galicia, where autochthonous and allochthonous varieties are cultivated, two main environments can be distinguished. One is the coast where GLRaV-3 vectors are occasionally found and seem to have been spreading increasingly in the last 10 years, and the other is the

For Peer Review

interior of the region (inland), where environmental conditions do not allow the winter survival of the vectors and consequently they are scarce or not found at all. In this work, an extensive survey was conducted to analyse the structure of the GLRaV-3 populations in Galicia that will help in understanding the epidemiology of the leafroll disease and developing control strategies.

The detection of GLRaV-3 isolates was carried out by combining a multiplex RT-PCR protocol with the partial sequencing of two genomic regions, and allowed the identification in Galicia of viral variants from four of the main six currently recognized phylogenetic groups (groups I to VI; Bester *et al*., 2012a; Chooi *et al*., 2013a,b). The distribution of the GLRaV-3 isolates in the five selected vineyards is differential, reflecting the different vines and the cultivation history at each particular vineyard. Half of the samples showed multiple infections with different GLRaV-3 variants, including a centenary non-stock vine.

The phylogenetic analysis for the HSP70h and CP regions of the Galician GLRaV-3 isolates assigned them to only three of the groups proposed because the primers chosen for these analyses excluded group VI isolates (Bester *et al.*, 2012b; Chooi *et al.*, 2013b). In some cases the isolates were included in different groups, depending upon the genomic region considered, as has been reported previously (Turturo *et al.*, 2005). These contradictions could be explained in terms of the perennial nature, vegetative propagation and grafting of the grapevine that may be exposed to multiple infections; therefore, different isolates in the same plant could have been sequenced for both regions.

Although *in silico* analysis supported recombination among several of the Galician GLRaV-3 isolates, the presence of mixed infections of GLRaV-3 variants in the same plant is the most probable explanation. In all the cases where recombination events were detected *in*

For Peer Review

silico, multiplex RT-PCR showed more than one GLRaV-3 variant in the sample. Recombination was not detected within the HSP70h and CP regions in these isolates. Thus, from these results it cannot be concluded that recombination in the GLRaV-3 isolates from Galicia had occurred. Similarly, no evidence of recombination in a region spanning 428 nt within the CP gene was found in a survey performed in the Napa Valley in California (Sharma *et al.,* 2011) or the 3′ end of the genome (Wang *et al.*, 2011). The recombination events described by Turturo *et al.* (2005) are probably artefacts because of GLRaV-3 mixed infections. Gouveia *et al.* (2011) detected isolates from group IV in traditional Portuguese varieties also grown in Galicia; however, none of the isolates analysed in the present study could be assigned to that group. In addition, no isolate from Galicia could be assigned to group V, apparently a very minor one, in which there are the Chilean isolate CL-817, some Portuguese isolates (Gouveia *et al.*, 2011), and some Chinese isolates recently identified (Farooq *et al.*, 2013). Genetic variants assigned to group I were dominant in a recent sampling in China (Farooq *et al.*, 2013), while in Portugal groups I and II are the most common (Gouveia *et al.*, 2011), and although groups IV and V were present, the same were poorly represented in other wine regions. In another survey in New Zealand, groups I and VI were predominant in germplasm banks and in commercial vineyards (Chooi *et al.*, 2013a). Phylogenetic analysis of the CP region from the set of 185 GenBank isolates showed that group I variants are the most frequent (57·8%), followed by group II (24·3%) and group III (14%), although the representation of GLRaV-3 variants in the databases could be biased as the tools to detect group VI variants were only recently developed (Sharma *et al.*, 2011; Bester *et al.*, 2012b; Chooi *et al.*, 2013a,b) and divergent variants are probably underrepresented (Seah *et al.*, 2012; Goszczynski, 2013). Groups I and II were the most common in Galicia, as apparently is the case in the rest of the world. The reasons for the differential predominance of GLRaV-3 variants, as indicated by Gouveia *et al.* (2011), are

For Peer Review

not clear, but it is feasible to think that the wider distribution of some groups over others may have arisen from the selection of less aggressive variants infecting plant material or differences in the transmission efficiency by insect vectors for the different groups. However, this is in contrast with the results from Blaisdell *et al.* (2012) who found higher transmission efficiency of isolates from group VI with respect to isolates from group I. In the present study the isolate Y-285 from Yemen (INRA collection) belonging to group II was found to be efficiently transmitted to 100% of the recipient plants (authors' unpublished data). In two vineyards (Goián-Castañal and Meaño) where GLRaV-3 appears to be naturally spreading, it was observed that group II isolates were present in all but one of the GLRaV-3 infected plants while group VI isolates were present only in some of the plants, which might suggest that group II isolates are more efficiently transmitted than group VI ones.

In some cases it has been possible to link the GLRaV-3 genetic variant group with the area of cultivation. For example, in Australian vineyards groups II and VI are predominant both in single and multiple infections, with group III being the least represented (Jooste *et al.*, 2012). In the same study it was possible to assign group II variants with a specific production area, as previously noted by Jooste *et al.* (2011), who could not relate GLRaV-3 group variants with symptom severity, damage, origin or variety. No relationship was found between Galician GLRaV-3 isolates and location, age, source or variety, as described for GLRaV-3 (Turturo *et al.*, 2005) or the related species GLRaV-1 (Komínek *et al.*, 2005; Alabi *et al.*, 2011). This lack of correlation has been explained in terms of virus spread through the exchange of plant material, for although an autochthonous variety may be grown almost exclusively in a given location, infected rootstocks may come from anywhere in the world and, being symptomless, may be good leafroll virus carriers.

For the CP and HSP70h regions, the genetic diversity of the Galician GLRaV-3 isolates was found to be similar to that of GLRaV-3 isolates reported in previous studies

For Peer Review

(Turturo *et al*., 2005; Gouveia *et al*., 2011; Table 3). The degree of selection pressure on the analysed genomic regions was estimated by the ratio between synonymous and nonsynonymous substitutions (dNS/dS) that indicates the degree of variation in the nucleic acid, reflected in amino acid variation. For the GLRaV-3 CP and HSP70h regions, this ratio was < 1 (i.e. lower number of nonsynonymous substitutions than synonymous). This suggests that both genomic regions are under negative or purifying selection. According to these results, even though both genomic regions are under negative pressure, the dNS/dS ratio was slightly higher for the CP region than for the HSP70h one, suggesting increased pressure on the latter, although the differences were not significant. In the aforementioned study in which three regions corresponding to the *RdRp*, the *HSP70h* and the *CP* genes were compared, it was shown that the major selection pressures were found on the *CP*, although the differences were not statistically significant (Turturo *et al.*, 2005). Similarly for the CP region, the degree of polymorphism was reported to be smaller than in other regions of the 3′ end of the genome (Wang *et al.*, 2011). On average, vector-borne viruses have lower genetic diversity in the CP genomic region than those not transmitted by this means. Among the viruses causing GLD, the higher selective pressures for the CP region have been found for GLRaV-1 and GLRaV-3, which are transmitted by more vectors than the other ampeloviruses (reviewed in Maree *et al.*, 2013). In addition to the transmission, another constrain to changes in the amino acid sequence of the CP is its structural function (Alabi *et al.*, 2011).

Sites under negative pressure can be part of the protein functional motifs, while those under positive selection may indicate the changes necessary for the adaptation of the virus to new hosts and/or resistance breaking. No sites subjected to positive selection (at $P < 0{\cdot}05$) were found which together with the evidence of purifying selection indicate that the GLRaV-3 isolates in Galicia have been present in the region for an extended period, as found by Wang *et al.* (2011) in the Napa Valley. It seems that most of the variants of GLRaV-3 are

For Peer Review

present in the main grape-producing regions of the world, both in countries with a long grapevine tradition as well as in production areas which are recent but established with European plant material (Gouveia *et al.*, 2011; Jooste *et al.*, 2011, Sharma *et al.*, 2011). Previous evidence of positive selection in field populations of GLRaV-3 is lacking (Wang *et al.*, 2011), suggesting that the virus is not under new selective pressures (Maree *et al.*, 2013).

Insect vectors for GLRaV-3 were found only in two of the coastal vineyards (both in Meaño area). On the other hand, in one of the vineyards (Goian-Caiño) that included the autochthonous vine Caíño Tinto, the pattern of GLRaV-3 variants in the different samples indicates independent vine introduction. In a previous study in the Galician region of Rías Baixas, in five out of nine vineyards the spatial analysis showed foci of infected plants and gradients towards a border of the plots suggesting vector transmission inside the plots, although no vectors were observed at that time (Cabaleiro & Segura, 1997). It seems that leafroll vectors could be responsible for the spread of the virus only in mild climate areas and it could affect the molecular differentiation of the isolates belonging to Rías Baixas (Cabaleiro *et al.*, 2008; Cid *et al.*, 2010); however, there is a lack of differentiation among the populations of isolates collected in Galicia in zones with some or no activity of vectors. Even if some differentiated populations of GLRaV-3 were present in the autochthonous varieties before the phylloxera, they disappeared after the uncontrolled massive introduction of non-certificated grafted plants. And although vectors are not common, viruses from this infected vegetatively propagated material would reach the autochthonous varieties and mixed infections might be frequent.

For Peer Review

Acknowledgements

The authors wish to acknowledge C. Greif (INRA) and F. J. Legorburu (NEIKER) for sending GLRaV-3 infected reference plants and Drs L. Rubio (IVIA) and E. Cretazzo (IFAPA) for reviewing the manuscript. The technical assistance of M. J. Alcalá is gratefully acknowledged. This work was funded by grants RTA2008-00078-C03 from the Instituto Nacional de Investigaciones Agrarias (INIA), AVA201301.3 co-financed by IFAPA and FEDER and programme 'Consolidación y estructuración de unidades de investigación competitivas (GPC)' from Xunta de Galicia (CN 2012/321). The Diputación de Lugo continued support allows maintenance of the grapevine collection at Portomarín (Lugo).

For Peer Review

References

Alabi OJ, Rwahnih MA, Karthikeyan G *et al.*, 2011. *Grapevine leafroll-associated virus 1* occurs as genetically diverse populations. *Phytopathology* **101**, 1446–56.

Anon, 2005. Commission Directive 2005/43/EC of 23 June 2005 amending the Annexes to Council Directive 68/193/EEC on the marketing of material for the vegetative propagation of the vine. *Official Journal of the European Union* **L 164**, 37–45.

Bester R, Maree HJ, Burger JT, 2012a. Complete nucleotide sequence of a new strain of grapevine leafroll-associated virus 3 in South Africa. *Archives of Virology* **157**, 1815–9.

Bester R, Jooste AEC, Maree HJ, Burger JT, 2012b. Real-time RT-PCR high-resolution melting curve analysis and multiplex RT-PCR to detect and differentiate grapevine leafroll-associated virus 3 variant groups I, II, III and VI. *Virology Journal* **9**, 219.

Blaisdell GK, Zhang S, Daane K, Almeida RPP, 2012. Patterns of virus transmission from hosts with mixed infections. In: *Proceedings of the 17th Congress of the International*

Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG). Davis, CA, USA: University of California, Davis, 178–9.

Cabaleiro C, Segura A, 1997. Field transmission of *Grapevine leafroll associated virus* 3 by *Planococcus citri* Risso. *Plant Disease* **87**, 283–7.

Cabaleiro C, Couceiro C, Pereira S, Cid M, Barrasa M, Segura A, 2008. Spatial analysis of epidemics of *Grapevine leafroll associated virus-3*. *European Journal of Plant Pathology* **121**, 121–30.

Chooi KM, Cohen D, Pearson MN, 2013a. Molecular characterisation of two divergent variants of *Grapevine leafroll-associated virus 3* in New Zealand. *Archives of Virology* **158**, 1597–602.

Chooi KM, Cohen D, Pearson MN, 2013b. Generic and sequence-variant specific molecular assays for the detection of the highly variable *Grapevine leafroll-associated virus 3*. *Journal of Virological Methods* **189**, 20–9.

Cid M, Pereira S, Cabaleiro C, Segura A, 2010. Citrus mealybug (Hemiptera: Pseudococcidae) movement and population dynamics in an arbor-trained vineyard. *Journal of Economic Entomology* **103**, 619–30.

Farooq ABU, Ma Y, Wang Z *et al.*, 2013. Genetic diversity analyses reveal novel recombination events in *Grapevine leafroll-associated virus-3* in China. *Virus Research* **171**, 15–21.

Goszczynski DE, 2013. Brief report of a new highly divergent variant of grapevine leafroll-associated virus 3 (GLRaV-3). *Journal of Phytopathology* **161**, 874–9.

Gouveia P, Santos MT, Eiras-Dias JE, Nolasco G, 2011. Five phylogenetic groups identified

For Peer Review

in the coat protein gene of *Grapevine leafroll-associated virus 3* obtained from Portuguese grapevine varieties. *Archives of Virology* **156**, 413–20.

Hudson RR, 2000. A new statistic for detecting genetic differentiation. *Genetics* **155**, 2011–4.

Jooste AEC, Maree HJ, Bellstedt DU, Goszczynski DE, Pietersen G, Burger JT, 2010. Three *Grapevine leafroll-associated virus 3* (GLRaV-3) variants identified from South African vineyards show high variability in their 5′ UTR. *Archives of Virology* **155**, 1997–2006.

Jooste AEC, Pietersen G, Burger JT, 2011. Distribution of *Grapevine leafroll associated virus-3* variants in South African vineyards. *European Journal of Plant Pathology* **131**, 371–81.

Jooste AEC, Bester R, Maree HJ, de Koker W, Burger JT, 2012. A survey of red and white cultivars to test an improved detection technique for *Grapevine leafroll-associated virus 3* (GLRaV-3) variants identified in South African vineyards. In: *Proceedings of the 17th Congress of the International Council for the Study of Virus and Virus-like Diseases of the Grapevine (ICVG).* Davis, CA, USA: University of California, Davis, 122–3.

Komínek P, Glasa M, Bryxiová M, 2005. Analysis of the molecular variability of *Grapevine leafroll-associated virus 1* reveals the presence of two distinct virus groups and their mixed occurrence in grapevines. *Virus Genes* **31**, 247–55.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost S, 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* **23**, 1891–901.

For Peer Review

Larkin MA, Blackshields G, Brown NP *et al.*, 2007. CLUSTALW and CLUSTALX version 2. *Bioinformatics* **23**, 2947–8.

Li WH, 1993. Unbiased estimation of the rates of synonymous and non-synonymous substitution. *Journal of Molecular Evolution* **36**, 96–9.

Librado P, Rozas J, 2009. DNASP v. 5, a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–2.

Maree HJ, Almeida RPP, Bester R *et al.*, 2013. Grapevine leafroll-associated virus 3. *Frontiers in Microbiology* **4**, 82.

Martelli GP, 2014. Directory of virus and virus-like diseases of the grapevine and their agents. *Journal of Plant Pathology* **96** (Supplement), 7–8.

Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P, 2010. RDP3, a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462–3.

Pamilo P, Bianchi NO, 1993. Evolution of the *Zfx* and *Zfy* genes: rates and interdependence between the genes. *Molecular Biology and Evolution* **10**, 271–81.

Pesqueira AM, 2013. *Los Virus del Enrollado de la Vid (GLRaV) en Cultivares Tintos de* Vitis vinifera *L. en Galicia. Transmisión y Caracterización Molecular de Aislados Locales de GLRaV-3.* Lugo, Spain: Universidad de Santiago de Compostela, PhD thesis.

Rosenberg MS, Anderson CD, 2011. PASSAGE: Pattern analysis, spatial statistics and geographic exegesis. Version 2. *Methods in Ecology and Evolution* **2**, 229–32.

Seah Y, Sharma AM, Zhang S, Almeida RP, Duffy S, 2012. A divergent variant of Grapevine

For Peer Review

leafroll-associated virus 3 is present in California. *Virology Journal* **9**, 235.

Sharma AM, Wang J, Duffy S *et al.*, 2011. Occurrence of *Grapevine leafroll-associated virus* complex in Napa Valley. *PLoS ONE* **6**, e26227.

Tajima F, 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **12**, 585–95.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S, 2013. MEGA 6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725–9.

Tsai CW, Chau J, Fernandez L, Bosco D, Daane KM, Almeida RPP, 2008. Transmission of grapevine leafroll-associated virus 3 by the vine mealybug (*Planococcus ficus*). *Phytopathology* **98**, 1093–8.

Turturo C, Salderelli P, Yafeng D *et al.*, 2005. Genetic variability and population structure of *Grapevine leafroll-associated virus 3* isolates. *Journal of General Virology* **86**, 217–24.

Wang J, Sharma AM, Duffy S, Almeida RPP, 2011. Genetic diversity in the 3′ terminal 4·7-kb region of *Grapevine leafroll-associated virus 3*. *Phytopathology* **101**, 445–50.

Watterson GA, 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* **7**, 256–76.

For Peer Review

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site.

Table S1 Estimates of evolutionary divergence between sequences.

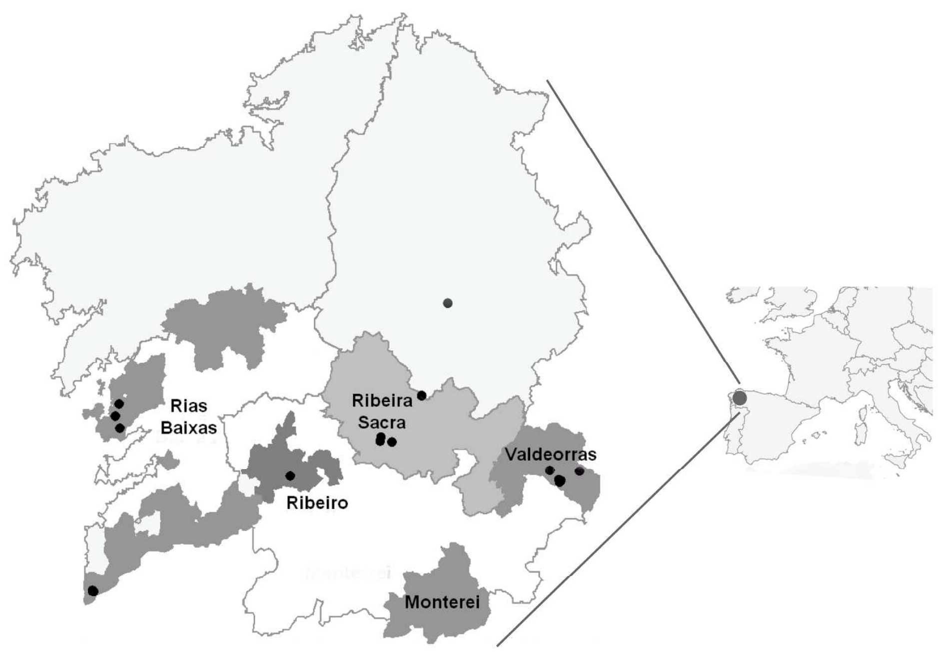
Table S2 Estimates of evolutionary divergence.

Figure legends

Figure 1 Map of the five regions with quality wine production (Denominaciones de Origen Controladas, DOC) for grapevine in Galicia (Spain) including the location of the vineyards surveyed for the characterization of the 98 *Grapevine leafroll-associated virus 3* isolates.

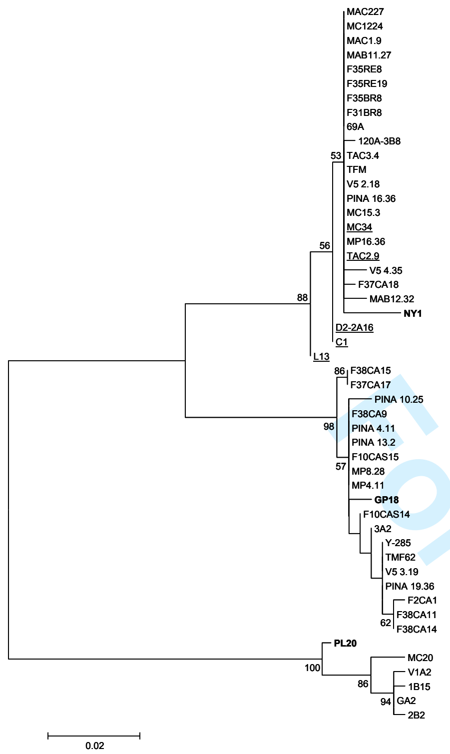
Figure 2 Maximum likelihood phylogenetic tree of *Grapevine leafroll-associated virus 3* isolates based on the HSP70h (a) and CP (b) genomic regions. A set of representative isolates (NY-1, GP-18 and PL-20) from each of the three GLRaV-3 groups was included in the analysis (in bold). The putative recombinant isolates detected are underlined (see main text for discussion). Bootstrap values (1000 replicates) are given at the branch nodes when higher than 50%.

For Peer Review

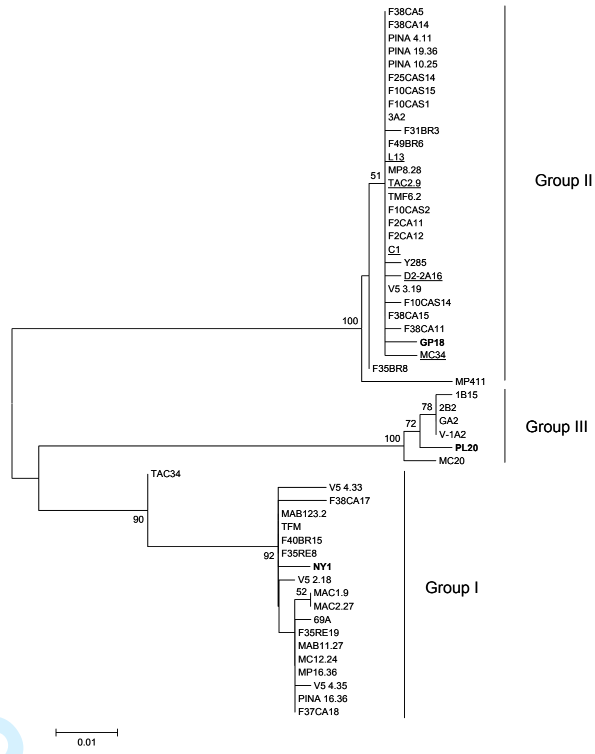


209x157mm (300 x 300 DPI)

(a)



(b)



For Peer Review

Table 1 Characteristics of the *Grapevine leafroll-associated virus 3* (GLRaV-3) isolates selected for this study

Isolate	Cultivar[a]	Vine data[b]	DOC/Loc[c]	YS/TY[d]	CP[e]	HSP70h[e]	Group[f]
1B15	*Albariño*	C,G,J,L	RB/Meaño	–	KP658158	KP658177	III
2B2	*Albariño*	C,G,J,L	RB/Meaño	–	KP658159	KP658178	III
3A2	*Albariño*	C,G,J,L	RB/Meaño	–	KP658160	KP658179	II
3A3NZ	*Albariño*	C,G,J,L	RB/Meaño	–	na	KP658198	VI (NZ2)
69A	*Albariño*	D,I,M	RB/Simes	18/18	KC953891	KC953928	I
120A-3B8	*Albariño*	C,G,J,L	RB/Meaño	–	na	KC953933	I
C1	*Albariño*	D,I,M	RB/Corneda	18/18	KC953883	KC953911	I, II
D2–2A16	*Albariño*	C,G,J,L	RB/Meaño	–	KC953907	KC953926	I, II
GA2	*Albariño*	D,J,L	RB/Galiñanes	–	KC953890	KC953937	II
L13	*Albariño*	D,I,M	RB/Cambados	18/18	KC953885	KC953914	I, II
V-1A2	*Albariño*	C,G,J,L	RB/Meaño	–	KJ476947	KC953929	III
F31BR3	*Brancellao*	A,E,F,I,L	RB/Goián	4/4	KC953900	na	II
F31BR8	*Brancellao*	A,E,F,I,L	RB/Goián	4/4	KC953898	KC953934	I
F35BR8	*Brancellao*	A,E,F,I,L	RB/Goián	3/3	KC953895	KC953935	I, II
F40BR15	*Brancellao*	A,E,F,I,L	RB/Goián	3/3	KC953897	na	I
F49BR6	*Brancellao*	A,E,F,I,L	RB/Goián	4/4	KJ463739	KC953917	II
F2CA1	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KC953894	KC953913	II
F2CA11	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KC953889	na	II
F37CA18	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658176	KP658193	I
F37CA18NZ	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	na	KP658202	VI (NZ2)
F38CA5	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658171	na	II
F38CA9	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	na	KP658188	II
F38CA11	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658172	KP658189	II
F38CA14	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658173	KP658190	II
F38CA15	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658174	KP658191	II
F38CA15NZ	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	na	KP658200	VI (NZ2)
F38CA17	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	KP658175	KP658192	II
F38CA18NZ	*Caiño Tinto*	A,E,F,L	RB/Goián	3/3	na	KP658201	VI (NZ2)
F10CAS1	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	KP658161	na	II
F10CAS2	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	KC953893	na	II

For Peer Review

F10CAS2NZ	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	na	KP658195	VI (NZ1)
F10CAS14	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	KP658162	KP658180	II
F10CAS15	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	KP658163	KP658181	II
F10CAS15NZ	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	na	KP658196	VI (NZ1)
F10CAS18NZ	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	na	KP658194	VI (NZ1)
F25CAS14	*Castañal*	A,B,E,F,I,L	RB/Goián	3/3	KP658164	na	II
F26CAS48NZ	*Castañal*	A,B,E,F,L	RB/Goián	–	na	KP658199	VI (NZ2)
MAB11.27	Mencía	E,F,L	RS/Abertesga	0/2	KC953899	KC953908	I
MAB12.32	Mencía	E,F,L	RS/Abertesga	0/2	KC953906	KC953912	I
MAC1.9	Mencía	E,H,J,L	RS/A Cova	3/5	KC953902	KC953930	I
MAC2.27	Mencía	E,H,J,L	RS/A Cova	4/4	KC953903	KC953920	I
MC14.24	Mencía	E,H,J,,L	RS/Carracedo	1/3	KC953905	KC953921	I
MC15.3	Mencía	E,H,J,L	RS/Carracedo	2/3	na	KC953916	I
MP4.11	Mencía	E,H,L	RS/Portomarín	0/5	KC953882	KC953932	II
MP8.28	Mencía	E,H,J,L	RS/Portomarín	5/5	KC953901	KC953927	II
MP10.25	Mencía	E,H,J,L	RS/Portomarín	5/5	KP658167	KP658184	II
MP13.2	Mencía	E,H,L	RS/Portomarín	0/4	na	KP658183	II
MP16.36	Mencía	E,H,L	RS/Portomarín	0/4	KC953904	KJ463743	I
MP19.36	Mencía	E,H,J,L	RS/Portomarín	3/4	KP658168	KP658185	II
V5-2.18	Mencía	E,F,I,L	V/Seadur	3/3	KJ463740	KC953936	I
V5-3.19	Mencía	E,F,I,L	V/Seadur	3/3	KJ463741	KC953923	II
V5-4.33	Mencía	E,F,I,L	V/Seadur	2/3	KP658166	na	I
V5-4.35	Mencía	E,F,I,L	V/Seadur	3/3	KP658165	KP658182	I
V5-4.35NZ	Mencía	E,F,I,L	V/Seadur	3/3	na	KP658197	VI (NZ1)
Y-285	Raziki	K,J,M	Yemen/INRA	18/18	KC953888	KC953909	II
F35RE8	*Retinto*	A,E,G,I,L	RB/Goián	3/4	KC953886	KC953924	I
F35RE19	*Retinto*	A,E,G,I,L	RB/Goián	3/4	KC953881	KC953931	I
MC20	Tempranillo	K,M	R/NEIKER	–	KC953879	KC953910	III
MC34	Tempranillo	K,M	R/NEIKER	–	KC953884	KC953918	I, II
TAC2.9	Tempranillo	E,G,I,L	RS/A Cova	5/5	KC953892	KC953925	I, II
TAC3.4	Tempranillo	E,G,I,L	RS/A Cova	4/5	KC953880	KC953922	I
TMF6.2	Tempranillo	E,G,I,M	RS/Montefurado	4/4	KC953896	KC953919	II

For Peer Review

TFM	*Tinta Femia*	A,B,I,M	RB/Beluso	18/18	KC953887	KC953915	I

[a]Autochthonous cultivars are indicated in italics.
[b]Other characteristics of the vines: A, minor cultivar; B, plant material from few infected centenary plants; C, healthy stock, GLRaV-3 field infection; D, Centenary and ungrafted plant; E, standard stock plants; F, plants from local nursery (Galicia); G, plants or rootstocks from Spanish nursery (not Galicia); H, plants from foreign nursery (Italy); I, strong leafroll symptoms; J, mild leafroll symptoms; K, non-Galician germplasm stock; L, symptoms and samples from its original location (grafted plants except centenary); M, symptoms and samples from direct plants in USC vineyards (Lugo).
[c]DOC and vine location: R, Rioja; RB, Rías Baixas; RS, Ribeira Sacra; V, Valdeorras.
[d]Years with visual symptoms observed/total number of years inspected.
[e]GenBank accession nos.; na, sequence not available because the amplicon was not obtained or ambiguous nucleotide sequences.
[f]Group to which the isolates were ascribed. When two groups are indicated for the same isolate it reflects mixed infection.

For Peer Review

Table 2 *Grapevine leafroll-associated virus 3* (GLRaV-3) variants identified for each phylogenetic group from the positive samples in Galicia, including isolates from the collections and from the five vineyards using the multiplex RT-PCR protocol

	Phylogenetic group				
Type of infection	Group I	Group II	NZ1[a]	NZ2[a]	Groups III–V[b]
Single	13	61	0	3	5
Double	11	18	6	6	1
Triple	1	2	2	0	1
Total	26	83	8	9	7

[a]Isolates belonging to Group VI.

[b]Isolates that could not be ascribed to groups I or II according to Chooi *et al.* (2013b).

For Peer Review

Table 3 Distribution of GLRaV-3 variants in the five vineyards studied according to the multiplex RT-PCR

Vineyard/DOC[a]	Variants	Isolates
Gaitán-Castañal/RB (Coast)	Group II	10
	Groups II + VI (NZ1)	7
	Group VI (NZ2+ NZ1)	1
	Total	*18*
Meaño/RB (Coast)	Group II	8
	Groups II + VI (NZ2)	2
	Total	10
Seadur/V (Inland)	Group I	2
	Group II	10
	Group VI (NZ1)	1
	Groups I + VI (NZ1)	1
	Total	*14*
Portomarin/RS (Inland)	Group I	1
	Group II	13
	Groups I + II	1
	Total	*15*
Goián-Caíño/RB (Coast)	Group II	10
	Groups I + II	3
	Group VI (NZ2)	4
	Group II + VI (NZ2)	1
	Groups I + VI (NZ2)	1
	Total	*19*

[a]Vineyards and DOCs: RB, Rías Baixas; RS, Ribeira Sacra; V, Valdeorras.

For Peer Review

Table 4 Genetic diversity, population genetic parameters and neutrality tests[a] of *Grapevine leafroll-associated virus 3* according to origin and genomic region

Gene	Population		*n*	d ± SE	dS ± SE	dNS ± SE	dNS/dS	*S*	p_s	Θ	π	Fu & Li's *D**	Fu & Li's *F**	*D*
HSP70h	Location[b]	World	96	0·030±0·004	0·075±0·012	0·010±0·003	0·1333	118	0·2857	0·0556	0·0273	–2·8343*	–2·8698*	–1·7893*
		Galician	44	0·053±0·009	0·139±0·025	0·016±0·005	0·1151	74	0·1791	0·0411	0·0483	0·8901	0·8966	0·5077
	Origin[c]	Autochthonous	24	0·065±0·016	0·169±0·028	0·020±0·006	0·1183	69	0·1670	0·0447	0·0594	1·3506	1·5080	1·1405
		Allochtonous	20	0·036±0·008	0·100±0·021	0·009±0·004	0·0900	33	0·0799	0·0225	0·0339	0·6468	0·6468	0·6468
	Variety	Albariño (white)	10	0·079±0·020	0·196±0·037	0·028±0·008	0·1429	66	0·1598	0·0564	0·0706	0·6468	0·6468	0·6468
		Autochthonous reds	14	0·039±0·009	0·103±0·022	0·010±0·005	0·0971	32	0·0774	0·0243	0·0366	1·1681	1·1681*	2·1689*
		Allochtonous reds	20	0·036±0·007	0·100±0·023	0·009±0·005	0·0900	33	0·0799	0·0225	0·0339	0·6468	0·6468	0·6468
CP	Location	World	185	0·053±0·012	0·201±0·029	0·021±0·005	0·1045	162	0·4164	0·0718	0·0609	–0·8669	–1·1085	–0·9948
		Galician	48	0·058±0·010	0·159±0·030	0·019±0·006	0·1195	67	0·1717	0·0387	0·0530	0·4970	0·8837	1·1628
	Origin	Autochthonous	29	0·057±0·010	0·163±0·031	0·019±0·006	0·1166	61	0·1564	0·0398	0·0522	1·0568	1·2630	1·0971
		Allochtonous	19	0·039±0·014	0·140±0·031	0·017±0·06	0·1214	45	0·1153	0·0330	0·0486	0·5742	1·1287	1·9143
	Variety	Albariño (white)	9	0·022±0·023	0·208±0·047	0·026±0·007	0·1250	57	0·1461	0·0537	0·0669	0·4101	0·6605	1·1437
		Autochthonous reds	20	0·043±0·008	0·119±0·027	0·013±0·005	0·1092	39	0·1000	0·0282	0·0389	0·9550	1·2727	1·4038
		Allochtonous reds	19	0·039±0·010	0·140±0·031	0·017±0·006	0·1214	45	0·1153	0·0330	0·0486	0·5852	1·1031	1·8626

[a]*n*, number of isolates; d, average substitution number per site (MLC); dS, synonymous substitutions per site (Pamilo, Bianchi & Li); dNS, non-synonymous substitutions per site (Pamilo, Bianchi & Li); dNS/dS, rate between non-synonymous and synonymous substitutions; SE, standard error with 1000 bootstrap replicates; *S*, number of segregating sites; p_s, *S*/*n*; Θ, ps/a1, population mutation rate; π, nucleotide diversity; *D*, Tajima's test statistic.

*, significant at $P < 0{\cdot}05$ that rejects the null hypothesis of selective neutrality.

[b]World, corresponding sequences available in GenBank (as of December 2014); Galician, isolates collected in Galicia from both autochthonous and foreign cultivated vines.

[c]Autochthonous varieties, including Albariño, Castañal, Brancellao, Caiño Tinto (Tinta Femia) and Retinto; allochtonous varieties includes non-autochthonous red varieties authorized in Galician DOCs (Mencía and Tempranillo); minor reds: Castañal, Brancellao, Caiño Tinto (Tinta Femia), Retinto.

For Peer Review

Table 5 Genetic differentiation between different subpopulations of GLRaV-3

Gene	Test	Subpopulation: Coast vs inland	Autochthonous vs allochtonous	Autochthonous reds vs whites	Autochthonous reds vs allochtonous	Autochthonous reds vs whites and allochtonous	Whites vs allochtonous
HSP70h	*P*-value of Ks*	0·2070 ns	0·2370 ns	0·0190 *	0·3810 ns	0·2270 ns	0·0170 *
	P-value of Z	0·1370 ns	0·1800 ns	0·0200 *	0·2800 ns	0·2320 ns	0·0070 **
	P-value of Snn	0·0270 *	0·0000 ***	0·0010 **	0·0170 *	0·0050 **	0·0000 ***
	F_{ST}[a]	0·02972	0·02032	0·23979	-0·00374	0·05029	0·19674
CP	*P*-value of Ks*	0·0910 ns	0·0710 ns	0·0180 *	0·0560 ns	0·02248	0·0200 *
	P-value of Z	0·1340 ns	0·0750 ns	0·0040 **	0·0170 *	0·0260 *	0·0310 *
	P-value of Snn	0·1350 ns	0·1010 ns	0·0030 **	0·1560 ns	0·1740 ns	0·0020 **
	F_{ST}	0·06645	0·08275	0·16873	0·12259	0·06588	0·1872

ns, not significant; *, $P < 0{\cdot}05$; **, $P < 0{\cdot}01$; ***, $P < 0{\cdot}001$. When $P < 0{\cdot}05$ it was considered as significantly rejecting the null hypothesis that there is no genetic differentiation between two subpopulations.

[a]F_{ST} is Wright's fixation index and allows the quantification of the genetic differentiation in natural populations. An absolute value of $F_{ST} > 0{\cdot}33$ suggests infrequent gene flow.

For Peer Review