

# The Tannat genome: Unravelling its unique characteristics

C. Da Silva<sup>1,a</sup>, A.D. Molin<sup>2</sup>, A. Ferrarini<sup>2</sup>, E. Boido<sup>3</sup>, C. Gaggero<sup>4</sup>, M. Delledonne<sup>2</sup>, and F. Carrau<sup>3</sup>

<sup>1</sup> PDU Espacio de Biología Vegetal del Noreste, Centro Universitario de Tacuarembó, Universidad de la República, 45000 Tacuarembó, Uruguay

<sup>2</sup> Centro di Genomica Funzionale, Dipartimento di Biotecnologie, Università degli Studi di Verona, 37134 Verona, Italy

<sup>3</sup> Área Enología y Biotecnología de Fermentación, Facultad de Química, Universidad de la República, 11800 Montevideo, Uruguay

<sup>4</sup> Departamento de Biología Molecular, Instituto de Investigaciones Biológicas Clemente Estable, 11600 Montevideo, Uruguay

**Abstract.** Tannat (*Vitis vinifera*) is the most cultivated grapevine variety in Uruguay for the production of high quality red wines. Its berries have unusually high levels of polyphenolic compounds (anthocyanins and tannins), producing wines with intense purple colour and high antioxidant properties. Remarkably, more than 40% of its tannins are galloylated, which determines a greater antioxidant power. Technologies of massive sequencing allow the characterization of genomic variants between different cultivars. The Tannat genome was sequenced with a 134X coverage using the Illumina technology, and was annotated using transcriptomes (RNA-Seq) of different berry tissues. When comparing the genomes of Tannat with Pinot Noir PN40024 (reference genome) we found an expansion of the gene families related to the biosynthesis of polyphenols. A search based on the recently reported epicatechin galloyl transferase (ECGT) from tea leaves determined five putative genes encoding the ECGT in Tannat. Genetic analysis of one of the transcription factor that regulates the anthocyanin synthesis during berry ripening, VvMYBA1, shows the presence of Gret1 retrotransposon in one of the VvMYBA1 alleles in the Tannat clones analysed. This work makes original contributions about the molecular bases of the biosynthesis of anthocyanins and tannins during the development of the flagship grape of Uruguay.

## 1. Introduction

Tannat (*Vitis vinifera*) is a grapevine cultivar originally from southwestern France that is now mainly cultivated in Uruguay, becoming its flagship. The first Tannat vines were introduced into Uruguay in the 1870s by European immigrants, but since the 1970s, many of the plants have been replaced with new French Tannat commercial clones, allowing Uruguay to produce high-quality red wines [1]. This *V. vinifera* cultivar has unusually high levels of polyphenolic compounds, producing wines with an intense purple colour, mouthfeel structure, aging potential, and remarkable antioxidant properties [2–4].

Tannat grapes are the richest in tannins [5]. The total flavan-3-ol content of Tannat seeds (1946 mg/kg) [5] is 6 times higher than the content reported for Pinot noir (317 mg/kg) [6] under similar experimental conditions. Also Tannat is the grape cultivar with the highest rate of galloylated tannins in seeds, which determines a greater antioxidant power [5]. The mechanism involved in galloylation of flavan-3-ols is still poorly understood. Recently, an epicatechin galloyl transferase (ECGT) has been isolated from tea leaves [7], allowing the study of this process in grapevine.

Its other remarkable characteristic is the high pigment contents, mainly given by the high proportion of anthocyanins. This polyphenolic compounds include malvidin, delphinidin, and petunidin monoglucosides [8]. One of the transcription factors that specifically regulate anthocyanin synthesis during ripening in grape is

VvMYBA1. In white grape varieties, one of the causes found to be responsible of the absence of anthocyanin production is a transposon insertion in the promoter of the VvMYBA1 gene [9, 10].

Polyphenols perform many crucial biological functions in plants [11], including the control of physiological traits of plant development, play a protective role against abiotic stresses [12–14] and also contribute to defense against biotic stresses [15–17].

Polyphenols acquired from the moderate consumption of red wine also provide pharmaceutical and nutritional benefits to humans [18], such as helping to prevent cancer and reducing the inflammation associated with coronary artery disease [19]. The proanthocyanidin content in local red wines in areas such as Gers in France, where Tannat is the main grape used for red wine production, correlate strongly with increased human longevity in this region [20]. It has been shown that the health-promoting effects of galloylated procyanidins are stronger than those of non-galloylated procyanidins [21, 22].

State-of-the-art sequencing technology combined with the availability of a grapevine reference genome now allows the characterization of genetic variations that affect the properties of wine [23, 24]. However, recent studies have shown that reliance on a single reference genome may underestimate the variability among different genotypes [25–27]. Plant genomes contain core sequences that are common to all individuals, as well as dispensable sequences comprising partially shared and non-shared genes that contribute to intraspecific variation [28]. The presence in specific cultivars of hundreds of genes not

<sup>a</sup> e-mail: [dasilvacece@gmail.com](mailto:dasilvacece@gmail.com)

shared with the reference genome has been demonstrated through the analysis of transcripts assembled *de novo* from RNA sequencing (RNA-Seq) data in both maize [29] and grapevine [30].

Our group investigated the genetic basis of the unique phenotypic characteristics of Tannat berries, which are most notable for their high content of polyphenolic compounds in general; and more specifically, their high galloylated flavan-3-ols and anthocyanins content; by sequencing the Uruguayan Tannat clone UY11 from the vines introduced into Uruguay in the 1870s [31]; and by RNA-Seq analysis annotated Tannat genes. We identified many genes that are not shared with the reference genome and that are involved in the synthesis of phenolic and polyphenolic compounds in a previous article [32], as well as 5 Tannat genes highly homologous to *Camellia sinensis* epicatechin galloyl transferase (ECGT) that we define as putative ECGTs in *V. vinifera*; and finally we characterized the presence of the retrotransposon Gret1 in 7 Tannat clones by PCR amplification, determined that all Tannat clones analysed have Gret1 retrotransposon in one of the VvMYBA1 alleles.

## 2. Materials and methods

### 2.1. Plant material, nucleic acid extraction, library preparation and sequencing

Details of how plant material was obtained, protocols of nucleic acid extractions, and how library preparation and sequencing was done are in [32].

Sampling and DNA extraction protocol for the 7 Tannat clones used in this study can be found in [31]. DNA of PN40024 was provided by Didier Merdinoglu.

### 2.2. Genome assembly and gene annotation

As described in Da Silva et al. [32], the genomic sequences were assembled using the IMR/DENOM v0.3.3 pipeline, a hybrid approach based on iterative realignment to the reference genome and integration of *de novo*-assembled contigs with a reference genome [27] with default parameters and the 12x PN40024 genome as a reference [33].

Different approaches were taken to annotate Tannat genes. As a first annotation approach, PN40024 V1 annotation gene models were translated onto the reconstructed Tannat genome by taking structural variations in the reference genome into account and adjusting the coordinates accordingly.

To reannotate the Tannat genome, we performed an RNA-Seq analysis using a panel of four tissues/developmental stages (whole berry, skins, and seeds). These transcriptomes were first reconstructed with a reference-guided approach using the Cufflinks suite [34–36].

Then, a *de novo* assembly approach was taken in order to reconstruct Tannat transcripts from our RNA-Seq data using the Velvet/Oases assembler [37,38]. We then applied five different filters to identify *de novo*-assembled transcripts missing from the genome assembly (details on the filtering methodology in Da Silva et al. [32]).

*De novo*-assembled transcripts that passed the filters were classified as missing from the genome assembly and compared with different sets of genomic sequences

to ascertain their presence in other genotypes: Pinot Noir PN40024 [33], Pinot Noir ENTAV115 [39]. These datasets were aligned with BLAT v34x12 (Kent, 2002), for Sanger reads, and BWA [41] against the Tannat transcripts. Transcripts covered ( $\geq 80\%$ ) by PN40024 genomic reads were classified as novel (shared with Pinot Noir but not previously annotated), whereas the remainder were classified as varietal. The novel and varietal transcripts were clustered based on similarity to sequences present in NR, VvGI, and the Tannat genome. Final clusters were considered to be genes.

Genes were functionally annotated by integrating the V1 manual annotation [42] and automatic annotations performed with Blast2GO [43].

For details on the methodology and where to find all data related to genome and transcriptomes see Da Silva et al. [32].

### 2.3. Epicatechin Galloyl Transferase (ECGT) homologous search

The search for proteins highly homologous to CsECGT (CN103740740) in Tannat was performed using tBlastn [44] against the Tannat genome, considering ECGT-like proteins the ones with at least 50% identity and more than 90% coverage.

### 2.4. Genetic variants in MYBA1

The presence of the retrotransposon Gret1 was evaluated by PCR amplification on 7 different Tannat clones (UY7, UY9, UY11, UY15, 399, 475, 717) and Pinot Noir clone PN40024. Kobayashi et al. [9] defined the allele MYBA1a as the MYBA1 gene with Gret1 retrotransposon inserted in the promoter region, and the allele MYBA1b as the MYBA1 gene with 3'-LTR of Gret1 in the promoter region. Primers used and PCR reaction specifications were as described by Kobayashi et al. [9]. PCR products were separated by electrophoresis and visualized by ethidium bromide staining on a 1.0% TBE agarose gel. The DNA size marker used was 2-Log DNA Ladder (0.1–10.0 kb, New England BioLabs inc.).

## 3. Results

### 3.1. Genome sequencing, assembly, gene annotation, and identification of Tannat-specific genes

Genomic DNA from the Uruguayan Tannat clone UY11 was used to generate 322,786,617 Illumina reads ( $2 \times 100$ ) representing 134-fold base pair coverage of the grapevine genome. Overall, the Tannat genome was 4.5 Mb (1%) shorter than the PN40024 genome, probably reflecting the limitations of current technologies for the detection of long insertions [27]. Although most of the deletions were located in intergenic regions, 7.29% of the deleted bases were considered as parts of genes in PN40024, and 99 genes were notable for the deletion of  $>50\%$  of the PN40024 sequence length. For details and further information see Da Silva et al. [32].

When PN40024 V1 annotation was transferred onto the Tannat genome as a first annotation approach, we found that from a total of 29,971 known genes, 22,983 were confirmed to lack deleterious mutations or structural alterations and were therefore transferred to the Tannat

**Table 1.** Summary of Genes in *V. vinifera* cv Tannat [32].

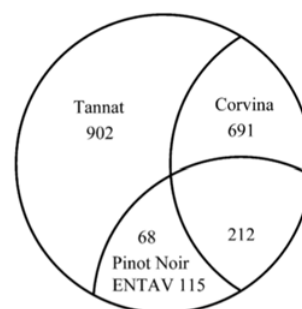
| Classification         | No. of genes |
|------------------------|--------------|
| Known (V1 annotation)  | 28,779       |
| Novel on assembly      | 2,866        |
| Novel outside assembly | 1,162        |
| Varietal               | 1,873        |
| Total                  | 34,680       |

genome annotation. The remaining 6,988 genes (23.3%) were predicted to encode proteins affected by deletions, truncations, or other disruptions. For these genes, it was not possible to transfer the reference annotation in a reliable manner [32].

The RNA-Seq analysis generated 395,863,776  $2 \times 100$  reads. Reference-based assembly of the RNA-Seq data provided experimental support for 16,169 of the 22,983 genes reliably transferred from PN40024 to Tannat. Also allowed us to reannotate 5,796 genes corresponding to loci transferred from the V1 annotation that appeared to be disrupted in the Tannat genome, as well as 2,866 non annotated protein-coding genes with homology to sequences in the National Center for Biotechnology Information (NCBI) non redundant (NR) and/or *Vitis vinifera* Gene Index (VvGI) databases. The features of newly annotated (or reannotated) genes, such as the average mRNA length (1,662 nucleotides) and the average number of exons per gene (6.77), were similar to those of PN40024 and other plant species [32].

We did not initially identify protein-coding genes that were present in the Tannat genome but missing from the PN40024 reference genome, probably because the assembly method used made it difficult to incorporate long insertions into the Tannat genome sequence. We therefore performed a de novo assembly of the RNA-Seq reads to generate a NR set of 114,786 transcripts with an average size of 1,491 nucleotides, which is similar to the corresponding value for the PN40024 annotation (1,331 nucleotides) and the Tannat annotation (1,554 nucleotides). Among the de novo reconstructed transcripts, a large fraction (88%) mapped with high confidence to the reconstructed Tannat genome. After all the filtering process, we found 5,052 high-confidence putative protein-coding transcripts that could not be mapped against the Tannat genome. Among this non mapping transcripts, 4,501 were validated by comparison with raw Tannat genomic reads. After clustering and manual inspection, these transcripts were grouped into 3,035 genes, which were compared with raw PN40024 genomic reads in order to identify genes that were still hidden in the unassembled portion of the PN40024 genome. This comparison revealed that the reference genome assembly lacks 1,162 genes and that Tannat possesses a set of 1,873 genes that are not shared with PN40024 (Table 1). Therefore, the Tannat genome appears to comprise 28,779 genes that are annotated on the reference genome (referred to as known genes), 4,028 genes previously unannotated or not assembled in the reference genome (referred to as novel genes), and 1,873 genes that appear to be unique to Tannat (varietal genes) [32].

To gain more insight into the origins of the 1873 Tannat varietal genes, we compared them with genomic sequences available for Pinot Noir clone ENTAV 115 [39] and with internally produced Illumina genomic reads from the



**Figure 1.** Categorization of the 1873 Genes Not Shared with PN40024. Number of genes found in common among Tannat, Corvina, and Pinot Noir (ENTAV 115) [32].

Corvina cultivar (Fig. 1). This revealed 280 Tannat varietal genes shared with the Pinot Noir clone ENTAV 115 that were probably lost by the PN40024 clone during the self-breeding process. Of the remaining genes, 691 were shared with Corvina and 902 were Tannat varietal genes. The presence of only partially overlapping differential sets of dispensable genes in the three cultivars concurs with the recent large-scale analysis of simple sequence repeat data showing that Corvina, Tannat, and Pinot Noir cultivars belong to different phylogenetic clades related by common ancestors [32, 45].

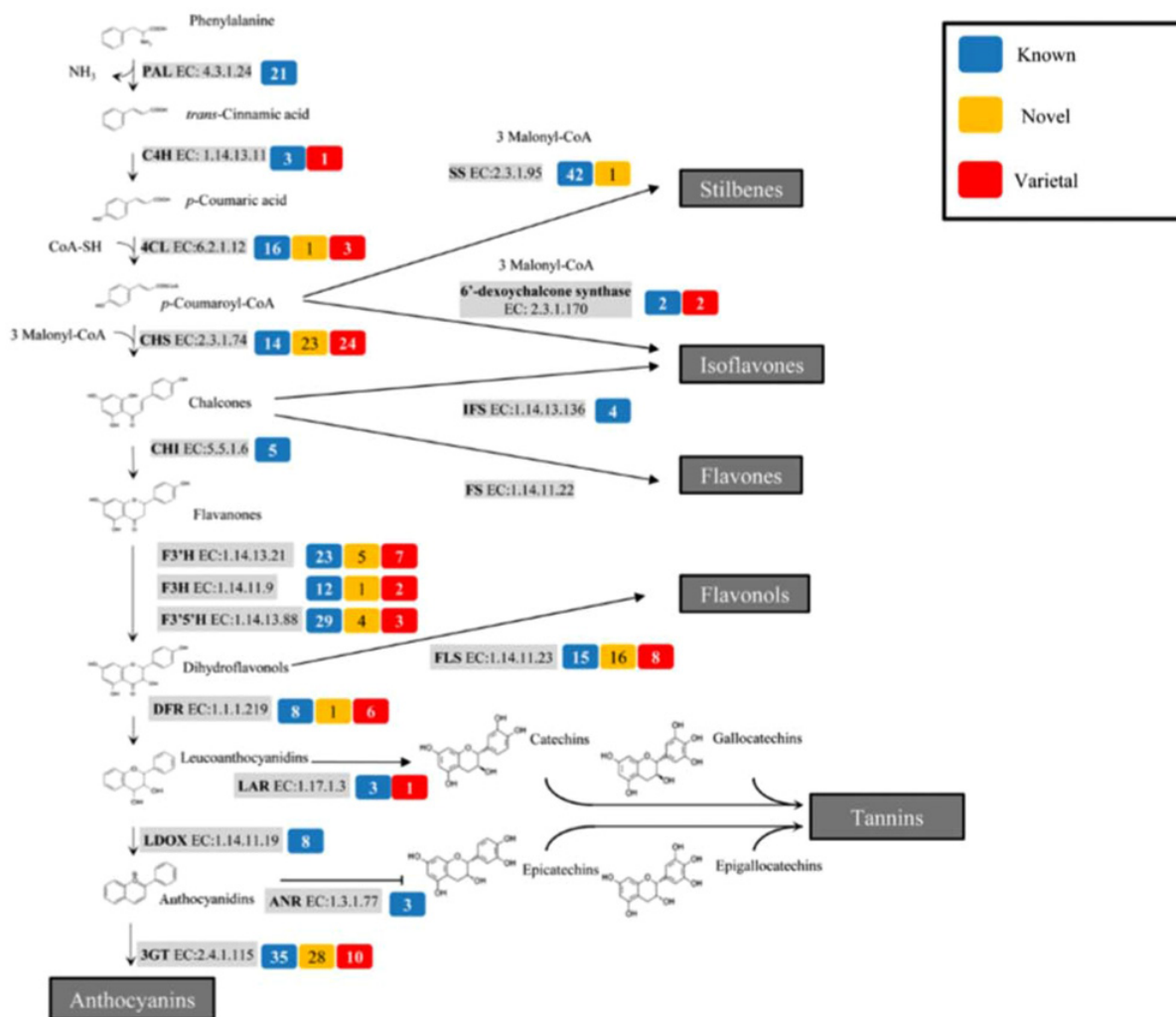
### 3.2. Expansion of gene families related to polyphenol biosynthesis

In order to understand the unusual high levels of polyphenols in the berry skin and seed, we checked the Tannat gene set for the presence of new family members in pathways related to polyphenol biosynthesis. We found 148 novel genes and 141 varietal genes corresponding to 23 different enzymes in pathways related to polyphenol biosynthesis. These included genes encoding some of the key enzymes in the phenylpropanoid pathway, such as cinnamate-4-hydroxylase, 4-coumarate:CoA ligase (4CL), and chalcone synthase (CHS). The CHS gene family in particular showed remarkable overrepresentation, with 47 new genes compared with 14 in the current V1 annotation. Gene families encoding key enzymes in the flavonoid pathway were also expanded: Flavonoid 3' hydroxylase (F3'H) was overrepresented, with 12 new genes compared with 23 genes in current annotation; three new flavonone 3 hydroxylase (F3H) genes were added to the 12 known genes; both the dihydroflavonol reductase (DFR) and flavonol synthase (FLS) families were expanded extensively, with the former adding seven new genes to the eight already present in the current annotation and the latter adding 24 members to the current 15. Finally, we identified 38 and 12 new genes similar to anthocyanidin 3-O-glucosyltransferases and 2'-hydroxyisoflavone reductases, respectively, compared with the 35 and 15 genes, respectively, in the current V1 annotation [32]. See Fig. 2.

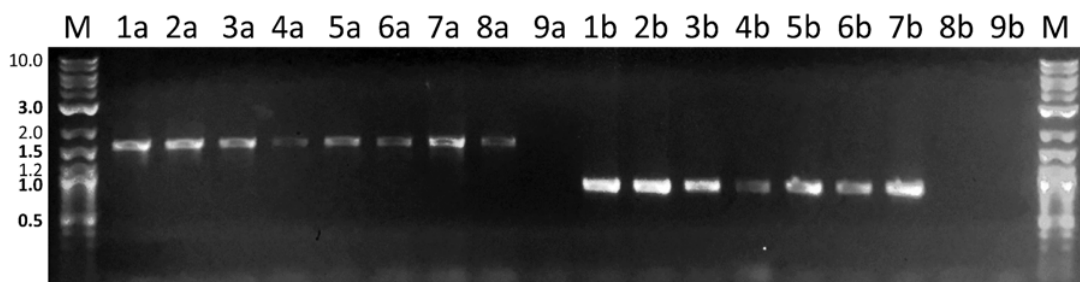
### 3.3. Epicatechin Galloyl transferases-like (ECGT-like) in Tannat

The search for proteins highly homologous to the ECGT from tea (*Camellia sinensis*) (Patent CN103740740A) retrieved 5 proteins in Tannat, encoded by the following





**Figure 2.** Schematic Diagram of Metabolic Pathways Involved in Polyphenol Biosynthesis. For each enzyme, the number of known genes (blue box), novel genes (yellow box), and varietal genes in Tannat (red box) is reported. FS, flavone synthase; IFS, 2-hydroxyisoflavanone synthase; LDOX, leucocyanidin oxygenase; ANR, anthocyanidin reductase [32].



**Figure 3.** PCR analysis of MYBA1a and MYBA1b alleles. Samples named as “a” indicates the MYBA1a allele analysis, while samples named as “b” refer to the MYBA1b allele analysis. DNAs from different Tannat clones were amplified and run in the following lanes: (1) UY7; (2) UY9; (3) UY11; (4) UY15; (5) 399; (6) 475; (7) 717; (8) Pinot Noir clone PN40024; and (9) the non-template control. M indicates the 2-Log DNA Ladder (0.1–10.0 kb, New England BioLabs Inc.) [32].

genes: VIT\_03s0088g00050, VIT\_03s0088g00260, VIT\_03s0091g01240, VIT\_03s0091g01290 and VIT\_11s0052g00790. These genes until now were annotated as serine carboxypeptidase-like (SCPL).

### 3.4. Anthocyanin biosynthesis in Tannat

In order to understand anthocyanin synthesis in Tannat we assessed the presence or absence of the retrotransposon Gret1 in the promoter region of the transcription factor MYBA1. We found that all Tannat clones are

heterozygous for this trait, having Gret1 inserted in one MYBA1 allele (allele MYBA1a), and having only Gret1 3'-LTR on the other allele (allele MYBA1b). The highly homozygous Pinot Noir clone PN40024 is homozygous for the MYBA1a allele (see Fig. 3).

## 4. Discussion

Because currently available sequencing technologies still produce reads that are too short for the accurate assembly of complex genomes (Schatz et al., 2012), we

reconstructed the Tannat genome using a hybrid approach [27] that relied on both the de novo assembly of Illumina genomic reads and iterative mapping against the PN40024 reference genome [33]. The highly contiguous genome assembly is comparable in length to the reference genome (98.9%).

The de novo assembly of the transcriptome from berry tissues sampled between flowering and veraison identified 1873 genes that we considered cultivar-specific because they are not shared with the reference genome. Although genes that are not conserved in a certain species are often assumed to be dispensable or redundant for development or survival, there is now evidence to indicate that such genes may have a strong impact on phenotype [46]; therefore, we named them varietal genes [32].

Tannat berries are notable for the production and accumulation of polyphenols, particularly anthocyanidins, in berry skins at maturity and tannins in both seeds and to a lesser extent in berry skins. The biosynthesis of these substances involves a large set of enzymes that convert Phe into diverse phenylpropanoids and flavonoids [16] and whose expression is tightly regulated during berry development [47, 48]. The production and accumulation of tannins in the seeds and skins occurs mainly during the early phases of berry development [5, 49]. We identified 141 varietal genes encoding 19 enzymes involved in the production of polyphenols. Some of them act in the early steps of the pathway and provide precursors for all the derivative branches, such as cinnamate-4-hydroxylase and 4CL [16]. These precursors are directed toward the flavonoid and isoflavonoid pathways by the enzymes CHS and chalcone isomerase, respectively. CHS catalyzes the condensation of 4-coumaroyl-CoA with three C2 units from malonyl-CoA to produce the C15 skeleton naringenin chalcone, which can be converted into flavones, isoflavones, flavonols, anthocyanidins, and tannins. We identified 24 varietal CHS genes in the Tannat genome, and this expansion compared with the reference genome may explain the higher polyphenol content of Tannat berries, since gene amplification is a well-documented mechanism to increase expression levels [50, 51]. Specifically, gene families representing key enzymes in the flavonoid pathway showed some degree of expansion, as F3H, F3'H, and flavonoid 3'5'-hydroxylase (F3'5'H). DFR, which catalyses a key step in the flavonoid pathway common to anthocyanin and tannin biosynthesis and is completely devoted to tannin biosynthesis before veraison; was represented by six varietal genes, and 3GT, the enzyme that catalyses a key step in the anthocyanin biosynthesis, was represented by ten varietal genes [32].

Taken together, these data indicate that the higher production of tannins and polyphenols in Tannat berries may be associated with the expansion of gene families encoding relevant enzymes in the varietal component of the Tannat genome. It is also worth noting that the gene family expansion we observed was not directly proportional to the existing gene family size; for example, some families with numerous members such as 4CL (16 members in the current annotation) showed a modest expansion (one novel and three varietal genes), whereas some small families like 6'-deoxychalcone synthase (two members in the current annotation) doubled in size [32].

In contrast with the expansion of several gene families encoding enzymes participating in polyphenol biosynthesis, transcription factor families involved in the regulation of these pathways showed no expansion of gene members. One of the transcription factors known to regulate anthocyanin synthesis in grape berries is VvMYBA1, which has been found to be one of the causes of the absence of anthocyanin production when the retrotransposon Gret1 is inserted in its promoter region [9, 10]. This retrotransposon insertion has been evaluated in many *V. vinifera*s cultivars, but not Tannat, and found that all red grape varieties analysed have Gret 1 in one of the VvMYBA1 alleles. We characterized seven Tannat clone for the presence-absence of Gret1, and found its presence present in one of the alleles of VvMYBA1 but not the other in all the Tannat clones analysed, as is the case with all red grape varieties. Surprisingly, PN40024 was found to have the Gret1 transposon in both copies of VvMYBA1, a characteristic of white varieties. This is probably due to the high level of introgression of PN0024, a highly homozygous line that is unable to give fruit.

Since Tannat is the grapevine cultivar showing the highest galloylation level in seeds [5] and galloylated compounds have proven to be more active than their non-galloylated counterparts [22], we decided to search for a Tannat putative galloyl transferase based on homology to the one reported for tea leaves. This way we identify five highly homologous proteins in Tannat, which are encoded in the genes VIT\_03s0088g00050, VIT\_03s0088g00260, VIT\_03s0091g01240, VIT\_03s0091g01290 and VIT\_11s0052g00790. The ECGT reported for tea is expressed in leaves and its probable ortholog in grape is VIT\_03s0088g00260 due to its high similarity. In the literature two of these genes (VIT\_03s0088g00260 and VIT\_03s0091g01290) were already reported as related to proanthocyanidin pathways [52–54].

As recent studies have shown that parts of the genome that are not shared among all genotypes of a species can contain functional genes, the observed amplification of genes encoding key enzymes in the polyphenol biosynthesis pathway in a cultivar characterized by very high levels of polyphenolic compounds suggests that the dispensable genome of grapevine contains many genes that can contribute to the establishment of intervarietal differences in phenotype.

We thank Didier Merdinoglu for providing us with DNA of PN40024 and the Institut National de la Recherche Agronomique for providing access to PN40024 raw Illumina genomic reads within the framework of GrapeReSeq, PLANT-KBBE2008. This work was supported by Fondazione Cariverona (Completamento e Attività del Centro di Genomica Funzionale Vegetale), a Joint Project 2012 between Biomolecular Research Genomics and the University of Verona, Ministero delle Politiche Agricole Alimentari e Forestali (Valorizzazione dei Principali Vitigni Autoctoni Italiani e dei loro Terroir-Vigneto), and Regione Veneto (Valorizzazione della Tipicità dei Vitigni Autoctoni e dei Vini Veneti-Valvive), and benefited from the networking activities within the European-funded COST ACTION FA1106. We also thank the support of UdelaR (projects: CSIC Group 656 to FC and CSIC INI2015 to CDS), ANII of Uruguay for supporting the travels and PhD fellowship of C.D.S. and PEDECIBA, Uruguay.

## References

- [1] F.M. Carrau, J. Wine Res. **8**, 179 (1997)
- [2] F. Carrau, E. Boido, C. Gaggero, K. Medina, E. Disegna, E. Dellacassa, *Multidisciplinary Approaches on Food Science and Nutrition for the XXI Century*, edited by Rosana Filip (Kerala, India, 2011)
- [3] C. Echeverry, F. Blasina, F. Arredondo, M. Ferreira, J.A. Abin Carriquiry, L. Vasquez, A.A. Aspillaga, M.S. Diez, F. Leighton, F. Dajas, J. Agric. Food Chem. **52**, 7395 (2004)
- [4] N.Q. Khan, B. Patel, S.S. Kang, S.K. Dhariwal, F. Husain, E.G. Wood, M.R. Potheary, R. Corder, Tetrahedron **71**, 3059 (2015)
- [5] E. Boido, M. García-Marino, E. Dellacassa, F. Carrau, J.C. Rivas-Gonzalo, M.T. Escribano-Bailón, Aust. J. Grape Wine Res. **17**, 383 (2011)
- [6] F. Mattivi, R. Guzzon, U. Vrhovsek, M. Stefanini, R. Velasco, J. Agric. Food Chem. **54**, 7692 (2006)
- [7] Y. Liu, L. Gao, L. Liu, Q. Yang, Z. Lu, Z. Nie, Y. Wang, T. Xia, J. Biol. Chem. **287**, 44406 (2012)
- [8] C. Alcalde-Eon, E. Boido, F. Carrau, E. Dellacassa, J.C. Rivas, Gonzalo, Am. J. Enol. Vitic. **57**, 449 (2006)
- [9] S. Kobayashi, N. Goto-Yamamoto, H. Hirochika, Science **304**, 982 (2004)
- [10] A.R. Walker, E. Lee, J. Bogs, D.A. McDavid, M.R. Thomas, S.P. Robinson, Plant J. **49**, 772 (2007)
- [11] T.M. Kutchan, Curr. Opin. Plant Biol. **8**, 292 (2005)
- [12] D. Treutter, Environ. Chem. Lett. **4**, 147 (2006)
- [13] S. Quideau, D. Deffieux, C. Douat-Casassus, L. Pouységú, Angewandte Chemie **50**, 586 (2011)
- [14] L.A. Weston, U. Mathesius, J. Chem. Ecol. **39**, 283 (2013)
- [15] A. Bhattacharya, P. Sood, V. Citovsky, Mol. Plant Pathol. **11**, 705 (2010)
- [16] R.A. Dixon, L. Achnine, P. Kota, C.J. Liu, M.S.S. Reddy, L. Wang, Mol. Plant Pathol. **3**, 371 (2002)
- [17] J. Chong, A. Poutaraud, P. Huguency, Plant Sci. **177**, 143 (2009)
- [18] J.K. Lin, M.S. Weng, *The Science of Flavonoids*, edited by E. Grotewold (Springer, New York, 2006)
- [19] N. Khan, V.M. Adhami, H. Mukhtar, Endocr. Relat. Cancer **17**, R39 (2010)
- [20] R. Corder, W. Mullen, N.Q. Khan, S.C. Marks, E.G. Wood, M.J. Carrier, A. Crozier, Nature **444**, 566 (2006)
- [21] D. Lizarraga, C. Lozano, J.J. Briedé, J.H. van Delft, S. Touriño, J.J. Centelles, J.L. Torres, M. Cascante, FEBS Journal **274**, 4802 (2007)
- [22] H.S. Kim, M.J. Quon, J.A. Kim, Redox Biol. **2**, 187 (2014)
- [23] S. Myles, J.M. Chia, B. Hurwitz, C. Simon, G.Y. Zhong, E. Buckler, D. Ware, PLoS ONE **5**, e8219 (2010)
- [24] Y. Zhang, L. Mao, H. Wang, C. Brocker, X. Yin, V. Vasiliou, Z. Fei, X. Wang, PLoS ONE **7**, e32153 (2012)
- [25] N.M. Springer, et al., PLoS Genet. **5**, e1000734 (2009)
- [26] R.A. Swanson-Wagner, S.R. Eichten, S. Kumari, P. Tiffin, J.C. Stein, D. Ware, N.M. Genome, Research (2010)
- [27] X. Gan, et al., Nature **477**, 419 (2011)
- [28] M. Morgante, E. De Paoli, S. Radovic, Curr. Opin. Plant Biol. **10**, 149 (2007)
- [29] C.N. Hansey, B. Vaillancourt, R.S. Sekhon, N. de Leon, S.M. Kaeppler, C.R. Buell, PLoS ONE **7**, e33071 (2012)
- [30] L. Venturini, et al., BMC Genomics **14**, 41 (2013)
- [31] A. Gonzalez Techera, S. Jubany, I. Ponce de Leon, E. Boido, E. Dellacassa, F.M. Carrau, P. Hinrichsen, C. Gaggero, Vitis **43**, 179 (2004)
- [32] C. Da Silva, G. Zamperin, A. Ferrarini, A. Minio, A. Dal Molin, L. Venturini, G. Buson, P. Tononi, C. Avanzato, E. Zago, E. Boido, E. Dellacassa, C. Gaggero, M. Pezzotti, F. Carrau, M. Delledonne, Plant Cell **25**, 4777 (2013)
- [33] O. Jaillon, et al., Nature **449**, 463 (2007)
- [34] C. Trapnell, L. Pachter, S.L. Salzberg, Bioinformatics **25**, 1105 (2009)
- [35] C. Trapnell, B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold, L. Pachter, Nat. Biotechnol. **28**, 511 (2010)
- [36] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D.R. Kelley, H. Pimentel, S.L. Salzberg, J.L. Rinn, L. Pachter, Nat. Protoc. **7**, 562 (2012)
- [37] D.R. Zerbino, E. Birney, Genome Res. **18**, 821 (2008)
- [38] M.H. Schulz, D.R. Zerbino, M. Vingron, E. Birney, Bioinformatics **28**, 1086 (2012)
- [39] R. Velasco, et al., PLoS ONE **2**, e1326 (2007)
- [40] W.J. Kent, Genome Res. **12**, 656 (2002)
- [41] H. Li, R. Durbin, Bioinformatics **26**, 589 (2010)
- [42] J. Grimplet, J. Van Hemert, P. Carbonell-Bejerano, J. Díaz-Riquelme, J. Dickerson, A. Fennell, M. Pezzotti, J.M. Martínez-Zapater, BMC Res. Notes **5**, 213 (2012)
- [43] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Bioinformatics **21**, 3674 (2005)
- [44] S.F. Altschul, W. Gish, W. Miller, E.W. Myers D.J. Lipman, J. Mol. Biol. **215**, 403 (1990)
- [45] G. Cipriani, et al., Theor. Appl. Genet. **121**, 1569 (2010)
- [46] S. Chen, B.H. Krinsky, M. Long, Nat. Rev. Genet. **14**, 645 (2013)
- [47] P.K. Boss, C. Davies, S.P. Robinson, Plant Mol. Biol. **32**, 565 (1996)
- [48] J. Bogs, F.W. Jaffé, A.M. Takos, A.R. Walker, S.P. Robinson, Plant Physiol. **143**, 1347 (2007)
- [49] C. Conde, P. Silva, N. Fontes, A.C.P. Dias, R.M. Tavares, M.J. Sousa, A. Agasse, S. Delrot, H. Gerós, Food **1**, 1 (2007).
- [50] J.R. Pollack, C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, C.F. Williams, S.S. Jeffrey, D. Botstein, P.O. Brown, Nat. Genet. **23**, 41 (1999)
- [51] Y. Wang, X. Wang, A.H. Paterson, Ann. N. Y. Acad. Sci. **1256**, 1 (2012)
- [52] N. Terrier, L. Torregrosa, A. Ageorges, S. Vialet, C. Verriès, V. Cheynier, C. Romieu, Plant Physiol. **149**, 1028 (2009)
- [53] G. Carrier, Y.F. Huang, L. Le Cunff, A. Fournier-Level, S. Vialet, J.M. Souquet, V. Cheynier, N. Terrier, P. This, Plant Physiol. Biochem. **72**, 87 (2013)
- [54] K. Koyama, M. Numata, I. Nakajima, N. Goto-Yamamoto, H. Matsumura, N. Tanaka, J. Exp. Bot. **65**, 4433 (2014)